

# PERSONALITY AND COGNITION IN ECONOMIC DECISION MAKING

EDITED BY: Aurora García-Gallego, Manuel I. Ibáñez and Nikolaos Georgantzis  
PUBLISHED IN: Frontiers in Psychology





## **Frontiers Copyright Statement**

© Copyright 2007-2017 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

**ISSN 1664-8714**

**ISBN 978-2-88945-236-1**

**DOI 10.3389/978-2-88945-236-1**

## **About Frontiers**

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## **Frontiers Journal Series**

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## **Dedication to Quality**

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## **What are Frontiers Research Topics?**

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# PERSONALITY AND COGNITION IN ECONOMIC DECISION MAKING

Topic Editors:

**Aurora García-Gallego**, Laboratorio de Economía Experimental and Economics Department, Universitat Jaume I, Spain

**Manuel I. Ibáñez**, Universitat Jaume I, Spain

**Nikolaos Georgantzis**, University of Reading, United Kingdom & Universitat Jaume I, Spain

Psychologists studying cognitive processes and personality have increasingly benefited from the wealth of theory, methodology, and decision making paradigms used in economics and game theory. Similarly, for the economists, personality traits and basic cognitive processes offer a set of coherent explanatory constructs in economic behavior. Given the debate on preference invariance and behavioral consistency across contexts and domains, the papers in this topic shed light on the existence and effect of stable sets of idiosyncratic features on economic decision-making.

While the effects of personality and cognition on economic decisions remain under-explored, the papers contributed in this topic offer more than a stimulus for further research. The general message could be that personality and cognitive processes offer the stable idiosyncratic ground on which individual decisions are made.

**Citation:** García-Gallego, A., Ibáñez, M. I., Georgantzis, N., eds. (2017). Personality and Cognition in Economic Decision Making. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-236-1

# Table of Contents

## Chapter 1: Editorial

- 05 Editorial: Personality and Cognition in Economic Decision Making**

Aurora García-Gallego, Manuel I. Ibáñez and Nikolaos Georgantzis

## Chapter 2: Personality and Economic Decisions

- 08 Take the Money and Run: Psychopathic Behavior in the Trust Game**

Manuel I. Ibáñez, Gerardo Sabater-Grande, Iván Barreda-Tarazona, Laura Mezquita, Sandra López-Ovejero, Helena Villa, Pandelis Perakakis, Generós Ortet, Aurora García-Gallego and Nikolaos Georgantzis

- 23 The Personality Trait of Intolerance to Uncertainty Affects Behavior in a Novel Computer-Based Conditioned Place Preference Task**

Milen L. Radell, Catherine E. Myers, Kevin D. Beck, Ahmed A. Moustafa and Michael Todd Allen

- 37 Prudence, Emotional State, Personality, and Cognitive Ability**

Adriana Breaban, Gijs van de Kuilen and Charles N. Noussair

- 47 Does the Dark Triad of Personality Predict Corrupt Intention? The Mediating Role of Belief in Good Luck**

Huanhuan Zhao, Heyun Zhang and Yan Xu

- 63 Angels and Demons: Using Behavioral Types in a Real-Effort Moral Dilemma to Identify Expert Traits**

Hernán D. Bejarano, Ellen P. Green and Stephen J. Rassenti

- 79 Motivational Hierarchy in the Chinese Brain: Primacy of the Individual Self, Relational Self, or Collective Self?**

Xiangru Zhu, Haiyan Wu, Suyong Yang and Ruolei Gu

- 86 Prosocial Personality Traits Differentially Predict Egalitarianism, Generosity, and Reciprocity in Economic Games**

Kun Zhao, Eamonn Ferguson and Luke D. Smillie

## Chapter 3: Cognition and Economic Decisions

- 104 Cognitive Reflection, Decision Biases, and Response Times**

Carlos Alós-Ferrer, Michele Garagnani and Sabine Hügelschäfer

- 125 Creativity and Cognitive Skills among Millennials: Thinking Too Much and Creating Too Little**

Brice Cornet, Antonio M. Espín and Roberto Hernán-González

- 134 Gender Differences in Performance Predictions: Evidence from the Cognitive Reflection Test**

Patrick Ring, Levent Neyse, Tamas David-Barett and Ulrich Schmidt

- 141 Reference Point Heterogeneity**  
Ayse Terzi, Kees Koedijk, Charles N. Noussair and Rachel Pownall
- 151 Self-identified Obese People Request Less Money: A Field Experiment**  
Antonios Proestakis and Pablo Brañas-Garza
- Chapter 4: Personality, Cognition and Institutions**
- 161 Individual Characteristics vs. Experience: An Experimental Study on Cooperation in Prisoner's Dilemma**  
Iván Barreda-Tarazona, Ainhoa Jaramillo-Gutiérrez, Marina Pavan and Gerardo Sabater-Grande
- 174 At Least I Tried: The Relationship between Regulatory Focus and Regret Following Action vs. Inaction**  
Adi Itzkin, Dina Van Dijk and Ofer H. Azar
- 190 Don't Always Prefer My Chosen Objects: Low Level of Trait Autonomy and Autonomy Deprivation Decreases Mere Choice Effect**  
Zhe Shang, Tuoxin Tao and Lei Wang
- 200 Fluid Intelligence and Cognitive Reflection in a Strategic Environment: Evidence from Dominance-Solvable Games**  
Nobuyuki Hanaki, Nicolas Jacquemet, Stéphane Luchini and Adam Zylbersztein
- 212 Moderating Effects of Social Value Orientation on the Effect of Social Influence in Prosocial Decisions**  
Zhenyu Wei, Zhiying Zhao and Yong Zheng
- 221 Pay What You Want! A Pilot Study on Neural Correlates of Voluntary Payments for Music**  
Simon Waskow, Sebastian Markett, Christian Montag, Bernd Weber, Peter Trautner, Volkmar Kramarz and Martin Reuter
- 231 Commentary: Fairness is intuitive**  
Kristian O. R. Myrseth and Conny E. Wollbrant



# Editorial: Personality and Cognition in Economic Decision Making

**Aurora García-Gallego<sup>1</sup>, Manuel I. Ibáñez<sup>2</sup> and Nikolaos Georgantzis<sup>1,3\*</sup>**

<sup>1</sup> Laboratorio de Economía Experimental and Economics Department, Universitat Jaume I, Castellón, Spain, <sup>2</sup> Department of Basic and Clinical Psychology, Universitat Jaume I, Castellón, Spain, <sup>3</sup> School of Agriculture Policy and Development, University of Reading, Reading, United Kingdom

**Keywords:** personality, cognition, decision making

## Editorial on the Research Topic

### Personality and Cognition in Economic Decision Making

Recently, psychologists studying cognitive processes and personality have increasingly benefitted from the wealth of theory, methodology, and decision making paradigms used in economics and game theory. Similarly, for the economists, personality traits and basic cognitive processes offer a set of coherent explanatory constructs in economic behavior. Given the debate on preference invariance and behavioral consistency across contexts and domains, the papers in this topic shed light on the existence and effect of stable sets of idiosyncratic features on economic decision-making.

In Waskow et al., PWYW decisions are studied while acquiring fMRI data. Participants buy music either under a traditional “fixed-price” (FP) condition or under a PWYW mechanism. The data replicate previous results on the general feasibility of the PWYW mechanism. In the FP-condition, neural activity in frontal areas during decision-making correlates positively with the participants’ willingness to pay. No such relationship was observed under PWYW in any neural structure. Stronger activity of the lingual gyrus was observed during PWYW.

In Proestakis and Brañas-Garza, the authors deal with the degree to which obese people adjust their own behavior as a result of anticipated discrimination. Consistent with the System Justification Theory, the study finds that self-identified obese individuals request lower amounts of money. Self-perceived but not externally reported excessive weight captures the self-weight bias not only for obese but also for non-obese individuals. This self-weight bias, yielding lower salary requests, enhances discriminatory behavior against individuals who feel, but may not actually be, obese and consequently exacerbates the wage gap.

Corgnet et al. studies whether the push for recruiting diligent millennials using criteria such as cognitive reflection can ultimately hamper the recruitment of creative workers. A positive effect is observed of fluid intelligence on originality and elaboration measures of divergent creative thinking. Furthermore, the U-shape relationship between cognitive reflection and fluency and flexibility measures of divergent creative thinking is inverted. This suggests that thinking too much may hinder important dimensions of creative thinking. Diligent and creative workers may thus be rare.

In Zhu et al., event-related potentials were recorded to evaluate brain responses when gambling for individual self, a close friend (relational self), or a class (collective self). When outcome feedback was positive, gambling for the individual “self” evoked a larger reward positivity compared with gambling for a friend or for the class, while there is no difference between the latter two conditions. When outcome feedback was negative, no significant effect was found between conditions. These findings provide direct electrophysiological evidence that the individual self is at the top of the three-tier hierarchy of the motivational system in the collectivist brain.

## OPEN ACCESS

### Edited and reviewed by:

Anat Bardi,

Royal Holloway, University of London,  
United Kingdom

### \*Correspondence:

Nikolaos Georgantzis  
n.georgantzis@reading.ac.uk

### Specialty section:

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

**Received:** 06 April 2017

**Accepted:** 09 May 2017

**Published:** 23 May 2017

### Citation:

García-Gallego A, Ibáñez MI and  
Georgantzis N (2017) Editorial:  
Personality and Cognition in Economic  
Decision Making.  
*Front. Psychol.* 8:848.  
doi: 10.3389/fpsyg.2017.00848

In Shang et al., it is observed that the choice effect is a robust phenomenon in which even “mere choice,” not including actual actions could intensify the preference for the self-chosen over other-chosen objects. Two studies examine the hypothesis. The results showed that the mere choice effect measured by Implicit Association Test (IAT) significantly decreased for participants with lower levels of trait autonomy (Study 1) and when participants were primed to experience autonomy deprivation (Study 2).

In Radell et al., a novel computer-based CPP task is developed in which participants guide an avatar to enter into a room with frequent (i.e., rich) and less frequent (i.e., poor) rewards. Low IU individuals enter into both rooms at about the same rate, while high IU individuals enter into the previously rich room first. The latter’s attraction to rewards is consistent with previously observed behavior in opioid-addicted individuals. Thus, high IU may lead to a cognitive bias favoring increased vulnerability to addiction.

In Itzkin et al., the participants received six decision scenarios, in which they were asked to evaluate regret following action and inaction. Individual regulatory focus was measured by two scales. Promotion-focused individuals attributed less regret than prevention-focused individuals to action decisions. Regret following inaction was not affected by regulatory focus. In addition, a trigger for change decreases regret following action. Orthodox people tend to attribute more regret to an action decision. Thus, both the situation and a decision maker’s orientation affects regret after action and inaction.

In Ring et al., the performance predictions in the 7-item Cognitive Reflection Test (CRT) is studied. After completing the test, subjects predicted their own, other participants’, men’s, and women’s, correct answers. Men scored higher on the CRT than women and both men and women were too optimistic about their own performance. However, men think they perform significantly better than other men and do so significantly more than women. The equality between women’s predictions about their own performance and their female peers cannot be rejected.

In Alós-Ferrer et al., novel evidence is presented on response times and personality traits in standard questions from the decision-making literature where responses are relatively slow (medians around half a minute or above). All questions create a conflict between an intuitive process and more deliberative thinking. For CRT questions, the differences in response times are as predicted by dual-process theories, with alignment and heuristic variants leading to faster responses and neutral questions to slower responses than the original, conflict questions. For decision biases (where responses are slower), evidence is mixed.

In Hanaki et al., the authors study the relationships between the key facets of dominance solvability and two cognitive skills, cognitive reflection, and fluid intelligence. Dominance and one-step iterated dominance are both predicted by one’s fluid intelligence rather than cognitive reflection. Individual cognitive skills, however, only explain a small fraction of the observed failure of dominance solvability. The accuracy of theoretical predictions on strategic decision making thus not only depends on individual cognitive characteristics, but also,

perhaps more importantly, on the decision making environment itself.

Terzi et al. investigates the capacity of four potential reference points—(1) population average payoff, (2) announced expected payoff of peers in similar situations, (3) a historical average of earnings in the same task, and (4) an announced anticipated individual payoff—to organize decisions in a risky decision making task. The population average payoff is the modal reference point, followed by experimenter’s stated expectation of individual earnings, followed by average earnings of other participants. A sizeable share of individuals show multiple reference points. The reference point is not affected by a shock to her income.

In Myrseth and Wollbrant, the association between “intuitive” and “fast” (Cappelen et al., 2015) is discussed. The commentary argues that such an association requires “fast” to rule out “deliberative,” which would need information beyond relative response speed. The precise cut-off time for deliberative decisions may be difficult to establish (see e.g., Schneider and Shiffrin, 1977; Posner and Rothbart, 1998), thus, an individual offered a few seconds, may still have sufficient time to reflect consciously. Thus, “faster” responses ought not to be taken as “intuitive” *prima facie*.

In Breaban et al., an experiment is run to consider the emotional correlates of prudent decision making. Subjects were presented with lotteries, while their emotional responses were recorded with facial recognition software. They had to make binary choices between risky lotteries that distinguish prudent from imprudent individuals. They also perform tasks designed to assess their cognitive ability and a number of personality characteristics. It is found that a more negative emotional states correlate with greater prudence. Higher cognitive ability and less conscientiousness are also associated with greater prudence.

In Bejarano et al., independently reported measures of subjects’ cognitive capabilities, preferences, and sociodemographic characteristics relate to behavior in a real-effort moral dilemma. Rather than simple correlation, clustering subjects into groups based on behavior in the real-effort task reveals important systematic differences across groups. However, the results indicate a need for a more comprehensive theory explaining how combinations of different individual characteristics impact behavior.

In Barreda-Tarazona et al., four different groups of subjects are created based on subjects’ scores in altruism and reasoning ability. Subjects play both one-shot (random changing pairs) and repeated (fixed partners) prisoner’s dilemma (PD) games. Incentivised beliefs regarding cooperation are elicited, showing that high altruism leads to optimism about others’ cooperation and higher cooperation in the first repetitions of PD. Contrary to the one-shot PD, high reasoning ability increases the probability of cooperation.

In Wei et al., individual differences are combined with social influence, revealing the effect of social value orientation (SVO) and social influence on prosocial behavior in trust and dictator experiments. In the trust game, prosocials were less likely than proselves to conform to other members’ behavior, when the majority of group members distrusted the trustee. In the dictator game, prosocial subjects were influenced more by others’

generous choices than their selfish choices, even if the latter benefitted them. The results indicate that the effect of social influence appears to depend on individuals' SVO.

In Zhao et al., two studies examine individual differences in two forms of prosociality—generosity and reciprocity—with respect to two major models of personality, the Big Five and the HEXACO. Both generosity and positive reciprocity determine social preferences. Men were more generous when this was costless and women were more egalitarian overall. HEXACO honesty-humility predicted dictator, but not generosity allocations, while irritability and anger predicted lower generosity, but not dictator allocations. Politeness of Big Five agreeableness was uniquely and broadly associated with prosociality across all games.

Zhao et al. examines the association between the Dark Triad of personality (i.e., Machiavellianism, narcissism, and psychopathy) and corruption. The positive relation between the Dark Triad and bribe-offering or bribe-taking intention was mediated by the belief in good luck. Therefore, belief in good luck may be one of the reasons explaining why people with Dark Triad are more likely to engage in corruption regardless of the potential outcomes.

## REFERENCES

- Cappelen, A. W., Nielsen, U. H., Tungodden, B., Tyran, J.-R., and Wengström, E. (2015). Fairness is intuitive. *Exp. Econ.* 19, 727–740. doi: 10.1007/s10683-015-9463-y
- Posner, M. I., and Rothbart, M. K. (1998). Attention, self-regulation and consciousness. *Philos. Trans. R. Soc. B* 353, 1915–1927. doi: 10.1098/rstb.1998.0344
- Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. detection, search, and attention. *Psychol. Rev.* 84, 1–66. doi: 10.1037/0033-295X.84.1.1

Ibáñez et al. studies the association among different sources of individual differences such as personality, cognitive ability, and risk attitudes with trust and reciprocity in an incentivized binary trust game. Trust associates to positive urgency and emotionality and, specifically, to the extraversion's warmth facet. Participants scoring high in psychopathy exhibit increased electrodermal activity and reduced evoked heart rate deceleration when asked to decide whether or not to trust. Abstract reasoning and low disagreeable disinhibition favor reciprocity, while lack of reciprocity relates with a psychopathic, highly disinhibited, and impulsive personality.

While the effects of personality and cognition on economic decisions remain underexplored, the papers contributed in this topic offer more than a stimulus for further research. The general message could be that personality and cognitive processes offer the stable idiosyncratic ground on which individual decisions are made.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 García-Gallego, Ibáñez and Georgantzis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Take the Money and Run: Psychopathic Behavior in the Trust Game

**Manuel I. Ibáñez<sup>1,2†</sup>, Gerardo Sabater-Grande<sup>3†</sup>, Iván Barreda-Tarrazona<sup>3</sup>, Laura Mezquita<sup>1</sup>, Sandra López-Ovejero<sup>3</sup>, Helena Villa<sup>1</sup>, Pandelis Perakakis<sup>3,4</sup>, Generós Ortet<sup>1,2</sup>, Aurora García-Gallego<sup>3</sup> and Nikolaos Georgantzís<sup>3,5\*</sup>**

<sup>1</sup> Department of Basic and Clinical Psychology, Universitat Jaume I, Castelló, Spain, <sup>2</sup> Centre for Biomedical Research Network on Mental Health, Instituto de Salud Carlos III, Madrid, Spain, <sup>3</sup> Laboratory of Experimental Economics and Economics Department, Universitat Jaume I, Castellón, Spain, <sup>4</sup> Centro de Investigación Mente, Cerebro y Comportamiento, Universidad de Granada, Granada, Spain, <sup>5</sup> School of Agriculture, Policy and Development, University of Reading, Reading, UK

## OPEN ACCESS

### Edited by:

Kimberly J. Saudino,  
Boston University, USA

### Reviewed by:

Renata Mellinda Heilman,  
Babeş-Bolyai University, Romania  
Peter R. Blake,  
Boston University, USA

### \*Correspondence:

Nikolaos Georgantzís  
n.georgantzis@reading.ac.uk

<sup>†</sup>These authors have contributed equally to this work.

### Specialty section:

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

**Received:** 07 July 2016

**Accepted:** 10 November 2016

**Published:** 28 November 2016

### Citation:

Ibáñez MI, Sabater-Grande G, Barreda-Tarrazona I, Mezquita L, López-Ovejero S, Villa H, Perakakis P, Ortet G, García-Gallego A and Georgantzís N (2016) Take the Money and Run: Psychopathic Behavior in the Trust Game. *Front. Psychol.* 7:1866.  
doi: 10.3389/fpsyg.2016.01866

We study the association among different sources of individual differences such as personality, cognitive ability and risk attitudes with trust and reciprocate behavior in an incentivized experimental binary trust game in a sample of 220 (138 females) undergraduate students. The game involves two players, player 1 (P1) and player 2 (P2). In the first stage, P1 decides whether to trust and let P2 decide, or to secure an egalitarian payoff for both players. If P1 trusts P2, the latter can choose between a symmetric payoff that is double than the secure alternative discarded by P1, and an asymmetric payoff in which P2 earns more than in any other case but makes P1 worse off. Before the main experiment, we obtained participants' scores for Abstract Reasoning (AR), risk attitudes, basic personality characteristics, and specific traits such as psychopathy and impulsivity. During the main experiment, we measured Heart Rate (HR) and ElectroDermal Activity (EDA) variation to account for emotional arousal caused by the decision and feedback processes. Our main findings indicate that, on one hand, P1 trust behavior associates to *positive emotionality* and, specifically, to the extraversion's *warmth* facet. In addition, the impulsivity facet of *positive urgency* also favors trust behavior. No relation to trusting behavior was found for either other major personality aspects or risk attitudes. The physiological results show that participants scoring high in psychopathy exhibit increased EDA and reduced evoked HR deceleration at the moment in which they are asked to decide whether or not to trust. Regarding P2, we find that AR ability and mainly low *disagreeable disinhibition* favor reciprocal behavior. Specifically, lack of reciprocity significantly relates with a psychopathic, highly disinhibited and impulsive personality. Thus, the present study suggests that personality characteristics would play a significant role in different behaviors underlying cooperation, with extraversion/*positive emotionality* being more relevant for initiating cooperation, and low *disagreeable disinhibition* for maintaining it.

**Keywords:** behavioral economics, psychopathy, personality, experiment, trust game, risk attitudes

## INTRODUCTION

Cooperation between strangers is an essential characteristic of human societies that differentiates us from other animal species (Fehr and Fischbacher, 2003). Central processes for understanding such cooperation are trust and reciprocity (Nowak, 2006; Walker and Ostrom, 2009; Balliet and van Lange, 2013). In accordance to the centrality of these behaviors for important social, economic and political outcomes, they have become a relevant topic in classic disciplines, such as anthropology, sociology, evolutionary biology, psychology or economics, and in new emerging interdisciplinary fields, such as neuroeconomics (Loewenstein et al., 2008) and behavioral economics (Kahneman, 2003; Camerer et al., 2011). One of the most powerful tools for the development of these fields has been the use of economic games (Evans and Krueger, 2009; King-Casas and Chiu, 2012; Sharp, 2012). Economic games are multiplayer decision-making tasks originally developed within mathematical theory to analyze strategic decision-making among economic agents. Later, they have been extensively used as well-controlled, flexible, and replicable behavioral paradigms to model social interactions such as cooperation, trust, altruism, reciprocity, or retaliation, making them ideal for bridging the gap between theory and naturalistic data (Zhao and Smillie, 2015).

One experimental economic game frequently used for the study of cooperative behavior is the Trust Game<sup>1</sup> (TG), originally developed by Berg et al. (1995) to measure trust, and to show the importance of positive reciprocity in cooperation. Positive reciprocity is defined as the costly behavior of a second mover (trustee) that reward a kind behavior of the first mover (trustor) (Falk and Fischbacher, 2006), whereas trust in this game would be defined as a voluntary transfer of own money to another subject, with future reciprocation expected but not guaranteed (Gunnthorsdottir et al., 2002). The amount sent by the trustor is multiplied by some factor and received by the trustee, who in turn chooses to send all, some, or none of the received money back to the sender. Although the mathematically computed subgame perfect equilibrium solution of the TG predicts no transfer and no return, there are two main results systematically found: trustors tend to invest positive amounts and trustees to reciprocate to some extent (Johnson and Mislin, 2011).

Importantly, there are individual differences in these behaviors, i.e., people differ quantitatively in the extent of investment of the trustor and the reciprocation of the trustee. Interestingly, a significant portion of these individual differences are attributed to genetic factors, with heritability estimates ranging from 10 to 32% for trust behavior, and from 17 to 32% for trustworthiness, depending on the sample, Swedish or U.S., and the model, ACE or AE, (Cesarini et al., 2008). Personality is also under relevant genetic influences (Vukasovic and Bratko, 2015), and the potential role of personality at the basis of these behaviors has been widely acknowledged (Borghans et al., 2008; Ferguson et al., 2011; Heckman, 2011; Zhao and Smillie, 2015). Thus, the main objective of the present study is to explain (part of) these individual differences by means of personality

characteristics. Our major strength and novelty is that we try to explore this association systematically: we assess personality dimensions of the two more relevant personality models of the last decades, the Big Three and the Big Five and explore the role of more specific personality traits. Specifically, we focus on two aspects that could be relevant for collaborative behaviors, not previously examined in the TG: subclinical psychopathy and impulsivity. Examining the personality domains and traits associated to trust and reciprocity will help explaining relevant basic processes underlying cooperative behavior.

Among the most influential personality models in the last decades, those of Eysenck (1992) and McCrae and Costa (2008) are especially relevant for cooperative behavior. In an attempt to link psychological disorders to normal personality, Eysenck (1992) proposed three basic dimensions or facets: Extraversion (E), Neuroticism (N), and Psychoticism (P). P is conceived as normal personality dimension of vulnerability to antisocial behavior and psychopathy, whereas low P would be characterized by traits as empathy, socialization and cooperativeness (Eysenck, 1992). In the other hand, the most widely used and integrative model of personality nowadays is the Five-Factor Model (FFM) (John et al., 2008). This model encompasses five personality dimensions: E, N, Openness to Experience (O), Agreeableness (A) and Conscientiousness (C) (McCrae and Costa, 2008). These domains include specific facets of A and E, such as trust, altruism, straightforwardness, tender-mindedness or warmth that could be especially relevant in interpersonal relationships and for trust and reciprocity (Evans and Revelle, 2008). Consequently, it would be expected that the personality characteristics more relevant in interpersonal behavior, such as A and E would facilitate cooperative behavior, whereas the opposite, exploiting other people and parasitic behavior, would be predicted by low A, P and psychopathic-like characteristics.

Only a few studies have investigated the role of personality domains and their effects on trust and reciprocity in the TG, and no study has explored the role of theoretically relevant specific traits such as impulsivity or psychopathic-like personality. In relation to broad personality dimensions, we deal first with investment behavior of P1, i.e., trust. Although the results are relatively heterogeneous, they tend to show that those personality domains more related to interpersonal behavior, i.e., E and A, were the more consistent personality correlates of trust. Evans and Revelle (2008) found that A was associated with investing, mediated by the trait of trust. Using a strategic version of the TG, Becker et al. (2012) showed a significant correlation between the amount sent as a first mover and A, O and low C. Similarly, Müller and Schwieren (2012) found that the amount sent by the investor correlates significantly with low C and low N, and also significantly positive with A. Swope et al. (2008) found that E was associated with more sending in the TG. Accordingly, Ben-Ner and Halldorsson (2010) found that E and low C were strong predictors of the amount sent to a partner by investors. Also, Haring et al. (2013), using a humanoid robot as a trustee, found that the more extravert a person was, the higher the amount sent in the TG. It is interesting to note that some of these studies found a certain effect of low C on trust behavior (Ben-Ner and Halldorsson, 2010; Becker et al., 2012), probably reflecting the

<sup>1</sup>Also called the ‘investment game.’

role of low deliberation and impulsivity in the decision to trust or not to trust (Müller and Schwieren, 2012).

With regard to trustee behavior, studies seem to suggest a moderate and consistent role of A on reciprocity and, conversely, of low agreeableness-related traits on exploitation behavior. Thus, Ben-Ner and Halldorsson (2010) found that the only personality domain associated to the proportion of received money that is actually sent back was A. Also, Becker et al. (2012) obtained that reciprocity was significantly correlated with A and O. Similarly, Thielmann and Hilbig (2015a) have found that Honesty-Humility, a domain strongly related to the A (Gaughan et al., 2012), predicts trustee returns in three experiments on different variations of the TG. Last, Lönnqvist et al. (2012) found that those participants being both high on N and low on A transferred back much less than did other participants when receiving low investments.

Lönnqvist et al. (2012) have highlighted the fact that the joint presence of high N and low A is (together with low C) a combination typical of Borderline Personality Disorder (BPD) patients (Saulsman and Page, 2004; Samuel and Widiger, 2008). Accordingly, it has been shown that persons with BPD present a striking deficit in trust and reciprocation. When compared with the healthy controls, BPD patients tend to: (a) transfer a smaller amount of monetary units in a TG when acting as investor (Unoka et al., 2009); and (b) send lower returns when acting as a trustee (King-Casas et al., 2008). It is important to note that a core characteristic of BPD patients is impulsivity, mainly the urgency facets (Whiteside et al., 2005), supporting the above mentioned idea that disinhibition/impulsivity traits may play a role in the TG decisions.

But probably the personality disorder more strongly associated to non-cooperative behavior is Psychopathy. Psychopathy is characterized by traits such as social manipulation, exploitation, egocentrism, irresponsibility, deceitfulness, superficial charm, lack of remorse and shallow affect (Miller et al., 2001), and a central characteristic from an evolutionary perspective would be the success of psychopaths in exploiting social emotions of trust and cooperativeness (Mealey, 1995). In terms of the Five Factor Model, psychopathic characteristics may be understood as the extreme end of a continuum along normal personality functioning, and would be strongly represented in (low) A and (low) C domains (Miller et al., 2001; Miller and Lynam, 2003; Gaughan et al., 2012), with the interpersonal affective components (primary psychopathy) more closely related to low A, and the impulsivity and social deviance features (secondary psychopathy) more closely related to low C (Miller et al., 2008).

Surprisingly, the role of psychopathic characteristics has not been explored yet in the TG, although, studies in other economic games seem to indicate that psychopaths, both clinical and sub-clinical, have a tendency to behave in a non-cooperative way. Mokros et al. (2008) found that criminal psychopaths, compared with healthy participants, were markedly more prone to competitive behavior, as well as to non-adherence to the principles of fairness, as evidenced by greater accumulated reward and exploitation of partners Prisoner's Dilemma Game (PDG). Similarly, primary-psychopath participants were both

less generous to social partners in a dictator game and more likely to reject ungenerous offers in an ultimatum game (Koenigs et al., 2010). Montañés-Rada et al. (2003) found that patients with Antisocial Personality Disorder, a personality disorder strongly related to psychopathy (Widiger and Mullins-Sweatt, 2009), showed more non-cooperative behavior both in the presence and in the absence of a non-cooperative opponent using various modifications of the PDG played against a simulated opponent.

A similar tendency has been observed in non-clinical individuals scoring high on different psychopathy scales. Rilling et al. (2007) found that dyads of high-psychopathy individuals were more likely to lead to mutual defection (non-cooperation) relative to low-psychopathy dyads. In addition, they found a high correlation between non-cooperative behavior and psychopathy scores among the male participants of their sample. Curry et al. (2011), using simultaneous one-shot discrete, continuous and sequential PDG, found that undergraduates with higher scores in Machiavellian Egocentricity PPI subscale, a marker for psychopathy (Benning et al., 2003), cooperated less in simultaneous PDG and were less likely to initiate or reciprocate cooperation in sequential PD games. Gillespie et al. (2013) examine the effects of primary (selfish, uncaring) and secondary (impulsive, irresponsible) psychopathic personality traits on the responses of undergraduate participants to the in-group and the out-group (defined in terms of affiliation to a UK University) in dictator and ultimatum games. They found significant differences in game proposals to members of the in-group and the out-group, between low and high scoring participants on secondary psychopathic traits. Using a PDG with a computerized opponent, Johnston et al. (2014) found that participants with low levels of psychopathic traits exhibited increased social cooperation in the context of affective feedback, and that poor cooperation was uniquely predicted by high levels of psychopathic traits. Taken together, these findings seem to confirm that non-cooperative social actions are the norm among high-psychopathy individuals in social-dilemma, mainly ultimatum and PDG (King-Casas and Chiu, 2012).

Another source of individual differences that could also contribute to cooperative behavior could be general intelligence. Previous research has reported evidence of a positive correlation between intelligence and self-reported trust (e.g., Sturgis et al., 2010; Hooghe et al., 2012; Carl and Billari, 2014). Regarding trust behavior in economic games, a meta-analysis of 36 studies that used a repeated PDG and school-level average SAT and ACT scores as proxies for the intelligence, showed that students cooperate 5–8% more often for every 100-point increase in the school's average SAT score (Jones, 2008). Similarly, Burks et al. (2009) using a one-shot sequential PDG in a sample of truck driving students found that subjects with greater intelligence more accurately forecast others' behavior and differentiate their behavior more strongly, depending on the first-mover's choice. Additionally, players with higher cognitive abilities reciprocated cooperation in the second round of this PDG significantly more than low intelligent subjects. Specifically, in a series of incentivized trust games, Corgnet et al. (2015) showed that cognitive ability is positively correlated to trust but not with trustworthy behavior. Thus, individuals' cognitive

ability/intelligence has been associated with cooperative play in economic games.

Pro-social behavior may also be related to individuals' risk attitudes. In fact, Luhmann (1988) and Coleman (1990) describe trust from the viewpoint of standard economics as a subclass of situation involving risk. However, Fehr (2009) states that strong neurobiological as well as behavioral evidence indicates that this view is untenable. Accordingly, behavioral studies have consistently failed in finding any relationship between risk aversion and trust behavior in the investment game (e.g., Bohnet and Zeckhauser, 2004; Bohnet et al., 2008; Houser et al., 2010).

Last, the attentional resources and emotional consequences of decision making in the TG are also interesting to study. For example, the conflict between individual and interpersonal considerations may induce different emotional reactions. Also, the attention of a subject in anticipation of the monetary and emotional consequences associated with decision making in the TG could be the result of interaction between the context and a decision maker's personality. Lorber (2004) investigates the relations of HR and EDA with psychopathy through a meta-analysis of 95 studies. Low resting and task EDA were positively associated with psychopathy, indicating impaired emotional regulation (Casey et al., 2013). Moreover, EDA reactivity was negatively associated with psychopathy. Contrary to the aforementioned relation between EDA and psychopathy, the latter was not associated with HR. In contrast, the relation between cardiac reactivity and psychopathy is less clear (Lorber, 2004; Casey et al., 2013). Here, we investigate these two physiological variables to probe the level of emotional and attentional engagement during the crucial trust decision.

To sum up, collaborative and altruistic behavior is central in human societies. A sequence of trust and reciprocity is usually assumed to be the small-group paradigm equivalent of a society in which citizens trust each other and deserve to be trusted, thus avoiding wasteful use of. An ideal experimental paradigm to examine these behaviors is the TG. In the discrete form adopted here, TG can be seen as a sequential social dilemma type of situation. If P1, who is the first mover, chooses not to trust P2, an egalitarian outcome emerges. Otherwise, if P1 trusts P2, the latter chooses between an egalitarian outcome, which is Pareto-superior to the one discarded by P1 and an unequal one which is favorable to P2 and unfavorable to P1.

The major strength and novelty of this study is that it systematically explores the association between behavior in the trust game with personality and cognitive abilities. To this end, a broad set of personality domains and specific personality facets are assessed. Specifically, we focused on two personality aspects that could be potentially relevant for collaborative behaviors: impulsivity and psychopathy.

No previous studies have directly explored the relationship between psychopathy and behavior in the TG. Considering the non-cooperative, exploitative and parasitic life-style of psychopaths, it is expected that their tendency to benefit from others' effort and trust would manifest in no reciprocating behavior. Indeed, the TG would be paradigmatic for assessing exploitative and other predatory-related behaviors closely related

to psychopathy traits, since one central issue for exploitation is *exploitability*, that is, the observable signs linked with the likelihood of being victimized (Buss and Duntley, 2008). Accordingly, P2's decision would represent an ideal context for expression of psychopathy-like behavior because P1 is in total *exploitability* by P2, who can benefit from P1's trust without receiving any negative consequences.

Conversely, agreeable and extraverted individuals tend to show more pro-social behavior, to cooperate more, to trust in other people, even strangers, and to respond in a positive way in front of kind and altruistic behaviors. Thus, A and E constitute the personality pillars of interpersonal relations, with A covering the quality of social interaction and E favoring the quantity of social interaction. Accordingly, one main hypothesis is that trust in the TG would be mainly associated to E and A, whereas reciprocity would be mainly associated to A. Conversely, psychopathy scores would be mainly associated to non-reciprocity and, in a lesser extent, to lack of trust.

Another underexplored area of personality effects on economic games is impulsivity. Impulsivity is a multifaceted construct of emotional-driven facets (positive and negative urgency), cognitive and behavioral features, (lack of both deliberation and perseverance), and sensitivity to reward (sensation seeking) (Whiteside and Lynam, 2001; Cyders et al., 2007). Because of lack of precedents, our hypotheses are general and speculative. In view of the reviewed literature, we hypothesize that impulsivity facets link to positive reinforcement would favor the more rewarding options in the TG, that is, to trust for P1, and to no-reciprocate for P2. Last, we hypothesize a positively relation of trust with cognitive ability and no association with risk-aversion.

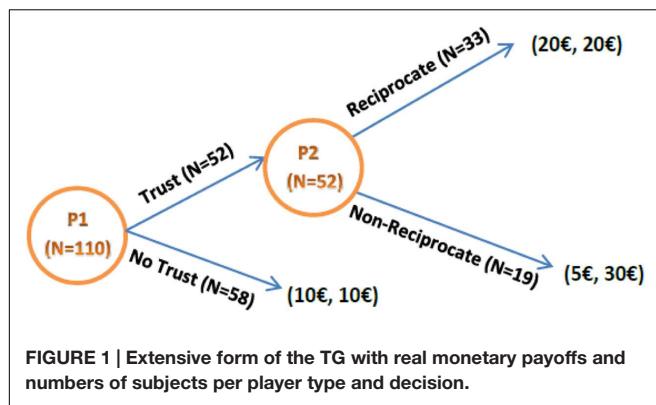
## MATERIALS AND METHODS

### Participants and Procedure

The experiment was run on two different dates. On the first date, 220 (138 females) undergraduate participants were recruited at the Individual Differences and Psychopathology (IDAP) Lab of the Universitat Jaume I. They signed a consent form for the entire experiment which they were informed that would take place on two dates and in two different labs. Then, they were asked to answer different socio-demographic and personality questionnaires.

On a second date, the same subjects were invited to the Laboratorio de Economía Experimental (LEE) of the same university to play a TG with real monetary incentives.<sup>2</sup> We divided the sample in P1 or Trustor ( $N = 110$ , 71 females) and P2 or Trustee ( $N = 110$ , 67 females) players (see **Figure 1**). This part of the experiment was carried out in 28 sessions of eight subjects each (forming four random and anonymous

<sup>2</sup>The data reported here are part of a larger study on personality traits and behavior in a series of games like PD, UG, Dictator and risky choice tasks. Payment was contingent on performance in one of all the economic games, chosen randomly at the end of the session, in order to avoid wealth accumulation effects and portfolio or hedging strategies. To avoid order effects, subjects were faced to the aforementioned contexts in randomized orders.



**FIGURE 1 | Extensive form of the TG with real monetary payoffs and numbers of subjects per player type and decision.**

pairs per session), using specific software prepared in Java by the IT team at the LEE<sup>3</sup>. The size of groups was dictated by the equipment available in the LEE for measuring Skin Conductance Responses (SCR) and HR variations. Continuous EDA and electrocardiographic (ECG) data were recorded during the entire experimental session using a BIOPAC MP150 system and four TEL100C telemetry modules (BIOPAC systems, Inc.). For EDA acquisition, two Ag/AgCl electrodes filled with isotonic gel were placed on each subject's distal phalanges of the middle and the index fingers of the non-dominant hand. The skin conductance signal was sampled at 125 Hz and low-pass filtered offline at 0.5 Hz using a Butterworth digital filter. SCR were automatically detected and their amplitudes were quantified using a custom version of the Matlab EDA toolbox.<sup>4</sup> False SCRs were removed after visual inspection of the entire signal. SCRs were associated to a specific decision if their onset appeared at least 1.0 s after subjects were informed about their possible choices and before the moment of the decision. Only responses above 0.02 microSiemens ( $\mu$ S) were considered as valid.

For ECG acquisition two FLAT active electrodes (Ag/AgCl) were arranged at a modified lead I configuration (i.e., right and left wrists). The ECG signal was sampled at 1000 Hz and filtered offline using a band-pass 0.5 – 50 Hz filter. R-wave detection and artifact correction were performed with the ECGLab Matlab software (Carvalho et al., 2002). We used the KARDIA Matlab software (Perakakis et al., 2010) and custom Matlab scripts (Matlab 2013a, Mathworks Inc.) to analyze the heart period signal during the experimental session. To assess the Phasic Cardiac Responses (PCRs) to a single decision moment, we first calculated the weighted average heart period for a time window of 2 s following the presentation of the decision screen, using the fractional counting procedure described in Dinh et al. (1999). We subsequently subtracted the weighted average heart period calculated for a window 0.5 s before cue onset, in order to express heart period changes as differential values from baseline activity.

<sup>3</sup>Software available upon request from A. Conde (alconvi@gmail.com) and J. V. Guinot (jose.guinot@gmail.com), JOOMALIA-Doing3D. The protocol used for the timing and communication between this software and the one used to measure the ECG is explained in detail in Perakakis et al. (2013).

<sup>4</sup>Freely available at: <https://github.com/mateusjoffily/EDA>.

## Measures

### Personality Measures

We used two broad personality models that include impulsivity and psychopathic-related dimensions, i.e., Eysenck's three factor model and McCrae and Costa's Five Factor Model, and a more specific test of both impulsivity and psychopathic traits. Importantly, these traits have been closely related to the aforementioned broad personality models (Whiteside and Lynam, 2001; Miller et al., 2008).

The Spanish NEO-PI-R (Costa and McCrae, 1999) is a 240-item self-report measure for quantifying 30 specific traits or facets that define the five personality factors or domains: N, E, O, A, and C. Items are responded to on 5-point Likert scales ranging from 0 (strongly disagree) to 4 (strongly agree). The specific facets for A were: Trust, Straightforwardness, Altruism, Compliance, Modesty and Tendermindedness. For C: Competence, Order, Dutifulness, Achievement striving, Self-discipline and Deliberation. For E: Warmth, Gregariousness, Assertiveness, Activity, Excitement seeking and Positive emotion. For N: Anxiety, Hostility, Depression, Self-Consciousness, Impulsiveness and Vulnerability. Last, for O: Fantasy, Esthetics, Feelings, Actions, Ideas and Values.

The Spanish Short version of the Eysenck Personality Questionnaire-Revised (EPQ-RS; Ortet et al., 2001) assesses Eysenck's broad dimensions of P, E, and N. Each scale consists of 12 items and the response alternatives are yes/no.

The Spanish version of the Levenson's Self-Reported Psychopathy Scale (LSRP; Lynam et al., 1999) is a 26-item four-point scale that ranges from 1 (strongly disagree) to 4 (strongly agree). It includes two related scales: the LSRP Primary or Factor 1 scale is associated to an antagonistic interpersonal style characteristic of psychopaths (i.e., low A, grandiosity, selfishness, callousness, manipulativeness), whereas LSRP Secondary or Factor 2 scale is more strongly related to disinhibition and negative emotionality (i.e., anger-hostility, urgency, lack of persistence and rashness; Miller et al., 2008; Lynam et al., 2011).

The UPPS-P Impulsive Behavior Scale (Verdejo-García et al., 2010) is a multidimensional inventory that assesses 5 personality pathways contributing to impulsive behavior: negative urgency, positive urgency, lack of perseverance, lack of premeditation, and sensation seeking. The scale is composed of 59 items with a four-point scale that ranges from 1 (strongly agree) to 4 (strongly disagree).

The AR scale of the Differential Aptitude Test (DAT-5, Bennett et al., 2005). This scale consists in a non-verbal AR test. Each item includes four abstract figures following a given rule, and the participant must choose one of five possible alternatives. The score is the total number of correct responses. One advantage of this test is that it is quite fast to implement: it is comprised of 40 multiple-choice items and has a 20 min time limit. AR would be considered a marker of fluid intelligence (Colom et al., 2007), the component of intelligence most related to general intelligence or g factor (McGrew, 2009).

### Risk Attitude Elicitation

We use two different incentive compatible elicitation procedures: the widely used method by Holt and Laury (2002) (Risk aversion

HL) and the Sabater-Grande and Georgantzis (2002) lottery-panels (SGG).<sup>5</sup>

Following the HL procedure, subjects are presented with a list of 10 pairwise choices between a safe (S) and a risky (R) lottery, each one of which involves a good and a bad outcome. Then, the difference between the good and the bad outcome in S is smaller than that in R. The list of lottery pairs is created by varying the probability of occurrence of the good outcome from  $p = 0.1$  to  $p = 1$  in steps of 0.1. A subject's risk aversion is an increasing function of the number of choices in which he or she has chosen the safe option. Given the monotonicity implied by the design, the actual switching point from S to R is used as the measure of a subject's risk aversion.

In the lottery panel test, SGG, subjects are faced with eight sub-tasks called panels 1, 2, 3... 8. Panels 1-4 involve only gains, while 5-8 involve mixed gambles. Each panel corresponds to a lottery defined as the probability  $p$  of winning a prize X€, else nothing in panels 1-4 (else a fixed loss of 1€ in panels 5-8). In all panels, the winning probability is varied from  $p = 0.1$  to  $p = 1$  in steps of 0.1. Prizes are designed so that, within a panel, the expected value of lotteries linearly increase in the probability of not winning by a constant  $t$  over a fixed gain of 1€ in panels 1-4 and 0€ in panels 5-8. Then,  $t$  represents an incentive for subjects to choose riskier choices. This parameter is increased from panel 1 to 4 and from 5 to 8. Thus, intuitively, a subject should be expected to make riskier choices when moving from panel 1 to 4 and from 5 to 8.<sup>6</sup>

In order to estimate the participants' score in SGG risk attitudes, an exploratory factor analysis with principal axes factor analysis and varimax rotation was performed. According to eigenvalue and parallel analysis, two factors emerged: Factor 1 (Risk aversion F1), comprising Panels 5-8 (with factor loadings from 0.70 to 0.87); and a relatively independent (Factor correlation = 0.20) Factor 2 (Risk aversion F2) comprising Panels 1-4 (with factor loadings from 0.73 to 0.83). These two factors explained 65.5% of the variance.

## Trust Game

The TG has been implemented in the lab in different versions: framed as a continuous investment game (Costa-Gomes et al., 2014), discrete with multiple choices (Berg et al., 1995) or discrete binary (Gambetta, 1988). Our experimental design is based on a discrete version of the game with binary choices and no particular framing. This strategy aims at reducing the space of investment options in order to facilitate the detection of the cognitive and emotional spectra activated by concentrating the observations on just two possible actions. This has led to more clear-cut data analysis, especially regarding the stimuli homogeneity for emotional arousal studied through the physiological part of our design. In this context, half of the participants acted as

<sup>5</sup>Attanasi et al. (2016) find no significant correlation between the two methods.

<sup>6</sup>García-Gallego et al. (2012) provide detailed discussion on the multidimensionality of the test and its implications under expected utility and alternative theories of decision making under risk. Three different aspects of a subject's risk attitude could be of interest here. First, whether a subject chooses safer choices. This would reflect a subject's risk aversion. Second, the sensitivity of the subject's choice to variations in  $t$ , measuring the incentive to take higher risks. Third, choice differences among gain (panels 1-4) and mixed-domain (panels 5-8) gambles, attributed to a subject's loss aversion.

P1 players (trustors,  $N = 110$ ), whereas the rest acted as P2 players (trustees).<sup>7</sup> Instructions to the subjects never mentioned trust, investment or reciprocity, in order to avoid undesirable experimenter demand effects. **Figure 1** presents the payoffs implemented in the game and the number of subjects who chose each strategy.

If the P1 player decides not to trust, both players earn with certainty an amount of 10€ each. But if the P1 player trusts P2, the latter will have to choose whether to reciprocate, raising each players' earnings to 20€, or to behave individualistically, raising own payoffs to 30€ and letting the trusting player down (5€). Pairs were randomly formed and the game was played once in its genuine sequential form. Each P1 players made the decision whether to trust or not before P2 made the second stage decision, provided that P1 had decided to trust in the first place. As shown in **Figure 1**, 52 (35 females) out of 110 P1 subjects decided to trust. From the 52 active P2 players, 33 (22 females) reciprocated and 19 (11 females) exploited P1's trust toward them.

## Data Analyses

We conducted the descriptive analyses and calculated correlations among all variables. In order to integrate the highly inter-correlated personality measures and to identify the basic personality domains underlying them, an Exploratory Factor Analysis with the assessed personality dimensions from different bio-dispositional models (NEO-PI-R and EPQ-RS), the measure of psychopathy (LSRP), and the measure of specific facets of impulsivity (UPPS-P) was performed.<sup>8</sup> We used principal axis factor analysis with varimax rotation. A parallel analysis with the Monte Carlo PA program was carried out to select the number of retained factors. The regression scores for each factor were kept as variables in the database and used later in the regression analysis.

In order to study the relationship among personality, cognitive ability and risk aversion variables on TG behaviors, mean comparison and regression analysis were performed. Thus,  $t$ -tests were calculated in order to determine whether the differences in personality and intelligence scores between trust vs. no trust groups, and reciprocate vs. no reciprocate groups were statistically significant. In order to examine the role of personality traits on the dichotomous choices in the TG, a Binary Logistic Regression analysis was performed. In a first step, we controlled for potentially confounding variables as age and gender; next, we included the scores on the AR scale of DAT; last, we included factor scores of personality traits. Factor scores were

<sup>7</sup>Whether the continuous version of the TG and its framing as a potentially reciprocal investment situation is more realistic, is a matter of the real-life example one has in mind. For example, there are situations in which a continuum of actions is not available and trust comes in the form of discrete events, like for example, signing first a contract or proposing marriage. In any case, the use of continuous vs. discrete versions could not be fully equivalent, and may lead to somewhat different results. Thus, as observed by Schniter et al. (2016) when comparing this binary version of the TG with a continuous version, although investments are higher in the all-or-nothing game than in the continuous game, higher investments in the binary game do not lead to higher returns. This suggests that subjects perceive intentions not only by evaluating what others do but also by evaluating what others choose not to do.

<sup>8</sup>See Markon et al. (2005) for a similar procedure.

used instead of the 15 direct scores in order to capture the basic personality domains underlying the highly inter-correlated personality scales.<sup>9</sup> All analyses were performed with the SPSS statistic package, version 21.

## RESULTS

### Descriptive Statistics

In **Table 1** we present descriptive statistics (median and standard deviation) of the explanatory variables included in our study. As usual, women presented higher scores in N, A, and lower scores in psychopathy, P, and several facets of impulsivity (Costa and McCrae, 1999; Ortet et al., 2001; Verdejo-García et al., 2010). In our sample, women also presented lower scores in E and AR. Last, and following Croson and Gneezy (2009) meta-analysis we find that women are in general more risk averse than men in lottery experiments.

### Factor Analysis

When the factor analysis was performed, the first four factors presented eigenvalues greater than 1, and the parallel analysis

<sup>9</sup>For a similar rationale and procedure, see Ibáñez et al. (2010).

suggested retaining four factors. The Barlett's test for sphericity ( $\chi^2 = 1654, 563$ ; df = 105,  $p < 0.000$ ) and the Kaiser-Meyer-Olkin (KMO = 0.729) indicated that the extraction method used was adequate to the data. **Table 2** shows the factor loadings of the personality scales in the factor solution. The factors corresponded to Unconscientious disinhibition, Neuroticism/negative emotionality, Extraversion/positive emotionality and Disagreeable disinhibition and accounted for 60% of the total variance.

### Mean Comparisons

First, we split the sample of P1 players according to their strategy. Factor scores presented statistical differences between those participants who trust vs. those that do not trust in the *Extraversion/positive emotionality* factor ( $t = 2.117$ ;  $p = 0.037$ ). **Figure 2** shows that trusting and non-trusting P1 subjects exhibited similar means in all personality characteristics except in positive urgency, in which players who trust scored higher than non-trusting players. In addition, trustors also presented a non-significant tendency in the E dimension of both EPQ-R and NEO PI-R questionnaires ( $p = 0.06$  and  $p = 0.10$ , respectively). When focusing on specific facets, trusting participants scored significantly higher in the *Warmth* facet of the E dimension than

**TABLE 1 | Means, standard deviations and test of differences between men and women (*t*-test for personality variables and MW test for SGG and HL scores on risk attitudes) of the variables included in the study.**

	Total (N = 222)		Men (M = 84)		Women (F = 138)		Mean comparison
	M	SD	M	SD	M	SD	
LSRP primary	15.41	6.52	18.49	6.73	13.54	5.64	-5.891**
LSRP secondary	9.41	3.64	9.51	3.64	9.36	3.65	-0.311
Neuroticism-NEO	92.92	23.40	85.69	24.47	97.33	21.65	3.695**
Extraversion-NEO	116.80	19.71	120.50	18.86	114.54	19.94	-2.203*
Openness-NEO	116.59	18.38	114.76	19.70	117.70	17.50	1.157
Agreeableness-NEO	116.50	18.61	107.61	17.29	121.91	17.30	5.975**
Conscientiousness-NEO	113.68	23.40	112.93	20.14	114.14	25.24	0.373
Extraversion -EPQ	8.61	3.12	9.26	2.63	8.22	3.33	-2.445*
Neuroticism -EPQ	4.78	3.57	4.12	3.36	5.19	3.65	2.181*
Psychoticism-EPQ	3.00	2.46	3.69	2.56	2.59	2.30	-3.317*
Premeditation-UPPS	31.01	5.21	30.44	4.41	31.36	4.38	-1.505
Negative urgency-UPPS	27.00	3.12	26.48	5.22	27.33	5.19	-1.181
Sensation seeking-UPPS	31.59	7.58	34.75	7.24	29.66	7.14	-5.123**
Perseverance-UPPS	25.35	3.18	25.02	3.15	25.55	3.20	1.198
Positive urgency-UPPS	26.48	7.62	27.96	7.32	25.58	7.68	-2.284*
DAT-RA	23.95	6.58	25.51	6.00	22.99	6.76	-2.808**
SGG Panel 1 probability	0.48	0.25	0.44	0.26	0.51	0.24	-2.377*
SGG Panel 2 probability	0.43	0.20	0.42	0.20	0.44	0.21	0.787
SGG Panel 3 probability	0.42	0.20	0.37	0.21	0.45	0.20	-2.613**
SGG Panel 4 probability	0.40	0.21	0.36	0.20	0.43	0.21	-2.267*
SGG Panel 5 probability	0.54	0.28	0.52	0.32	0.55	0.26	-0.813
SGG Panel 6 probability	0.50	0.27	0.49	0.28	0.51	0.26	-0.524
SGG Panel 7 probability	0.48	0.25	0.46	0.28	0.49	0.23	-1.205
SGG Panel 8 probability	0.44	0.23	0.42	0.24	0.45	0.22	-0.962
HL Number of Safe Lotteries	6.79	1.97	6.39	2.27	7.03	1.73	-1.903 <sup>+</sup>

<sup>+</sup> $p < 0.10$ ; \* $p < 0.05$ ; \*\* $p < 0.01$ .

**TABLE 2 | Principal axis factor analysis with varimax rotation of the personality scales.**

	<b>Factor 1</b>	<b>Factor 2</b>	<b>Factor 3</b>	<b>Factor 4</b>
Conscientiousness NEO	<b>-0.86</b>	-0.28	0.01	-0.06
Premeditation UPPS	<b>-0.78</b>	0.07	-0.18	0.02
Perseverance UPPS	<b>-0.73</b>	0.04	0.08	0.02
Psychopathy secondary LSRP	<b>0.56</b>	<b>0.48</b>	-0.02	0.26
Psychoticism EPQ	<b>0.51</b>	0.10	0.16	<b>0.33</b>
Neuroticism NEO	0.07	<b>0.81</b>	<b>-0.34</b>	-0.07
Neuroticism EPQ	-0.03	<b>0.77</b>	-0.25	0.06
Negative urgency UPPS	0.10	<b>0.74</b>	0.15	0.20
Positive urgency UPPS	0.19	<b>0.43</b>	0.25	<b>0.34</b>
Extraversion NEO	-0.12	-0.16	<b>0.89</b>	-0.03
Extraversion EPQ	-0.01	-0.20	<b>0.77</b>	-0.01
Sensation seeking UPPS	0.19	0.06	<b>0.53</b>	0.27
Openness NEO	0.16	0.06	<b>0.46</b>	-0.25
Psychopathy primary LSRP	0.07	0.07	0.02	<b>0.92</b>
Agreeableness NEO	-0.09	-0.16	0.09	<b>-0.70</b>
Exp. Var.	24.26%	17.16%	10.48%	8.16%

*Bold*, loadings higher than 0.30. Exp. Var., percentage of variance explained. Factor 1, unconscious disinhibition; Factor 2, negative emotionality; Factor 3, positive emotionality; Factor 4, disagreeable disinhibition.

non-trusting participants ( $t = 2.020$ ;  $p = 0.046$ ) and showed a non-significant tendency in scoring lower on *Angry-hostility* facet of the N dimension ( $t = -1.820$ ;  $p = 0.072$ ).

We split now the sample of active, deciding ( $N = 52$ ) P2 players according to their strategy in the second stage of the game. Factor scores presented statistical differences between trustees that reciprocate vs. non-reciprocate in the *Disagreeable disinhibition* factor ( $t = -2.885$ ;  $p = 0.006$ ) and *Negative emotionality* factor ( $t = -2.449$ ;  $p = 0.018$ ), whereas *Unconscious disinhibition* factor also presented a non-significant tendency ( $t = -1.911$ ;  $p = 0.062$ ). **Figure 3** depicts the mean differences in specific scales. Thus, it can be observed that trustees who display a reciprocal behavior have significantly lower levels in psychopathy-related traits than subjects who have opted for the individualistic reaction to their trusting counterpart. These differences are evident on the primary and secondary psychopathy and on P. Players who reciprocate also presented a non-significant tendency in A, mainly attributed to the significant mean differences found in the A facet of *Straightforwardness* ( $t = 2.611$ ;  $p = 0.012$ ). In addition, players who did not reciprocate presented higher scores on disinhibition-related traits, as positive and negative urgency, low persistence, sensation seeking, low C and the *Impulsivity* scale of N ( $t = 2.129$ ;  $p = 0.038$ ).

## Regression Analysis and Correlations

We present in **Table 3** the predictive power of the factors underlying the questionnaires on trust and reciprocity behaviors. Despite gender differences found in predictors, neither age nor gender associate to any dependent variable. Once controlled for these variables, neither cognitive ability nor risk aversion associate with trust, but higher AR predicted higher reciprocation. Regarding personality, the *Positive emotionality*

factor that included E scales, predicted trust behavior, whereas *Disagreeable disinhibition* factor, which included primary psychopathy, P, positive urgency and low A scales, predicted non-reciprocal behavior. In addition, *Unconscious disinhibition* and *Negative Emotionality* factors presented a marginally non-significant association with no reciprocation behavior, probably reflecting the role of impulsivity on this behavior.

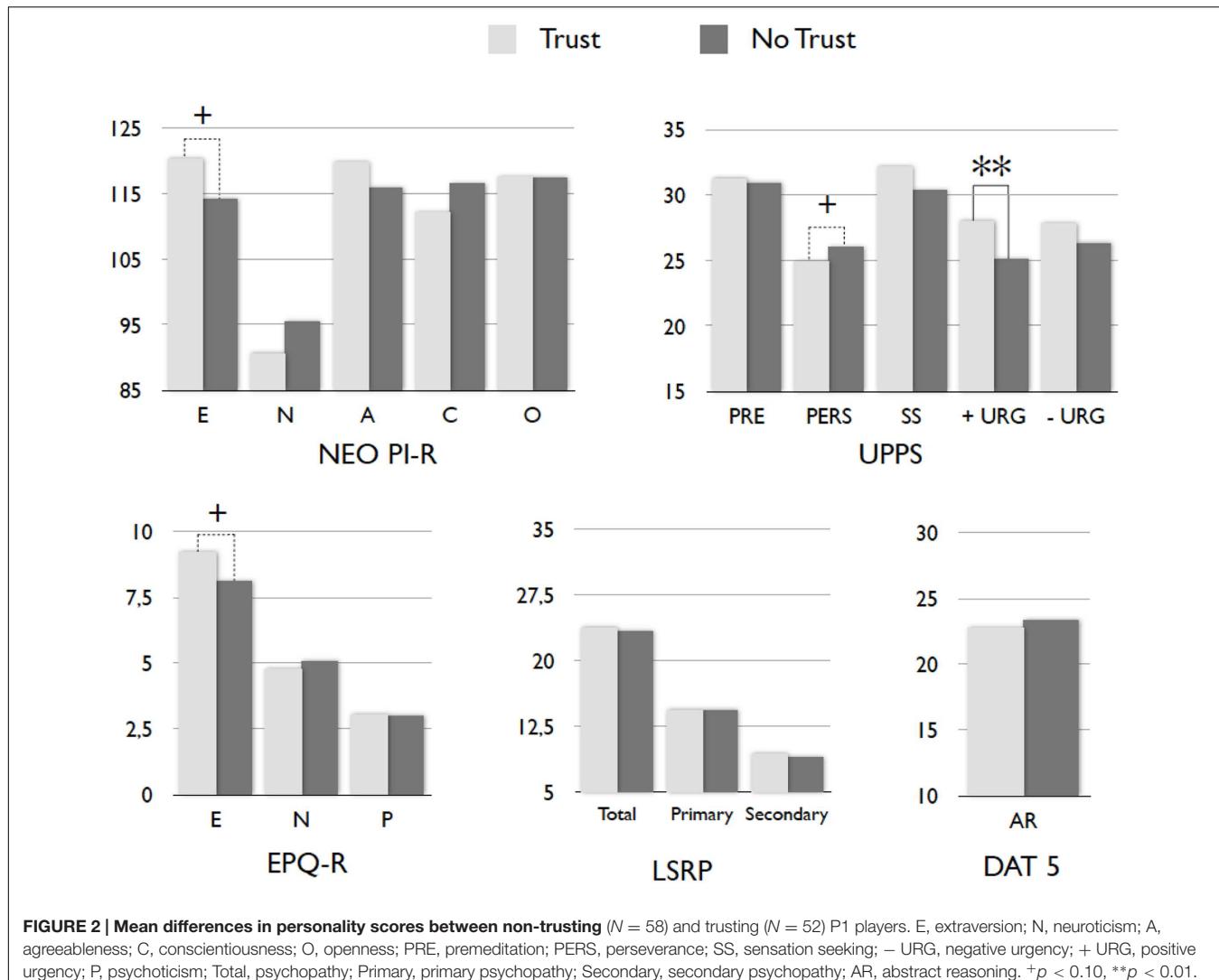
We look now at the results obtained from the physiological data. Interbeat intervals, measured one second after a screen is shown to P1 asking them to make a decision, significantly and negatively correlate with primary (Spearman,  $-0.338$ ,  $p = 0.007$ ) and total (Spearman,  $-0.314$ ,  $p = 0.013$ ) LSRP scores. Also, the amplitude of SCR corresponding to the same moment significantly correlates with primary (Spearman,  $0.267$ ,  $p = 0.015$ ) and total (Spearman,  $0.235$ ,  $p = 0.033$ ) LSRP scores. Both patterns indicate the relevance of the decision to trust in terms of attentional resources involved, and the emotions triggered in conjunction with the decision makers personality.

## DISCUSSION

The present study addresses factors that can account for individual differences in behavior of participants in the TG. To this end, we selected a wide range of personality constructs that might be useful in explaining the heterogeneity observed.

In order to integrate the different personality characteristics assessed within the FFM framework, we performed an exploratory factor analysis. We found a four-factor structure virtually identical to the one described by Markon et al. (2005) and similar to the ones found in other studies with a wide variety of personality scales (e.g., Zuckerman et al., 1993; Ortet et al., 2002; Aluja et al., 2004; Ibáñez et al., 2010). According to the nomenclature in Markon et al. (2005), the four factors we obtained were labeled Positive Emotionality, Negative Emotionality, Disagreeable Disinhibition and Unconscious Disinhibition. These factors are closely linked to the FFM of personality except for O, probably because this domain is not well represented in other personality models apart from the FFM (Markon et al., 2005).

Particularly relevant for the present research was the location of impulsivity and psychopathy scales within the FFM space. In reference to psychopathy, we found that subscales of the LSRP, although interrelated, loaded in two different factors: primary psychopathy characterized as manipulation, cheating, callousness and lack of remorse loaded in the Disagreeable Disinhibition factor, and would be mainly related to low A; while secondary psychopathy, characterized by impulsivity and deviant behavior, loaded in the Unconscious Disinhibition factor and would be mainly related to low C, in line with previous findings (Miller et al., 2008). In relation to impulsivity, it constitutes a complex multifaceted construct of pervasive importance in psychology (Evenden, 1999). In an attempt to add clarity to the impulsivity concept, Whiteside and Lynam (2001) identified four distinct components of impulsivity (i.e., urgency, sensation seeking, perseverance, and deliberation) and located them within the FFM framework. Posterior studies subdivided



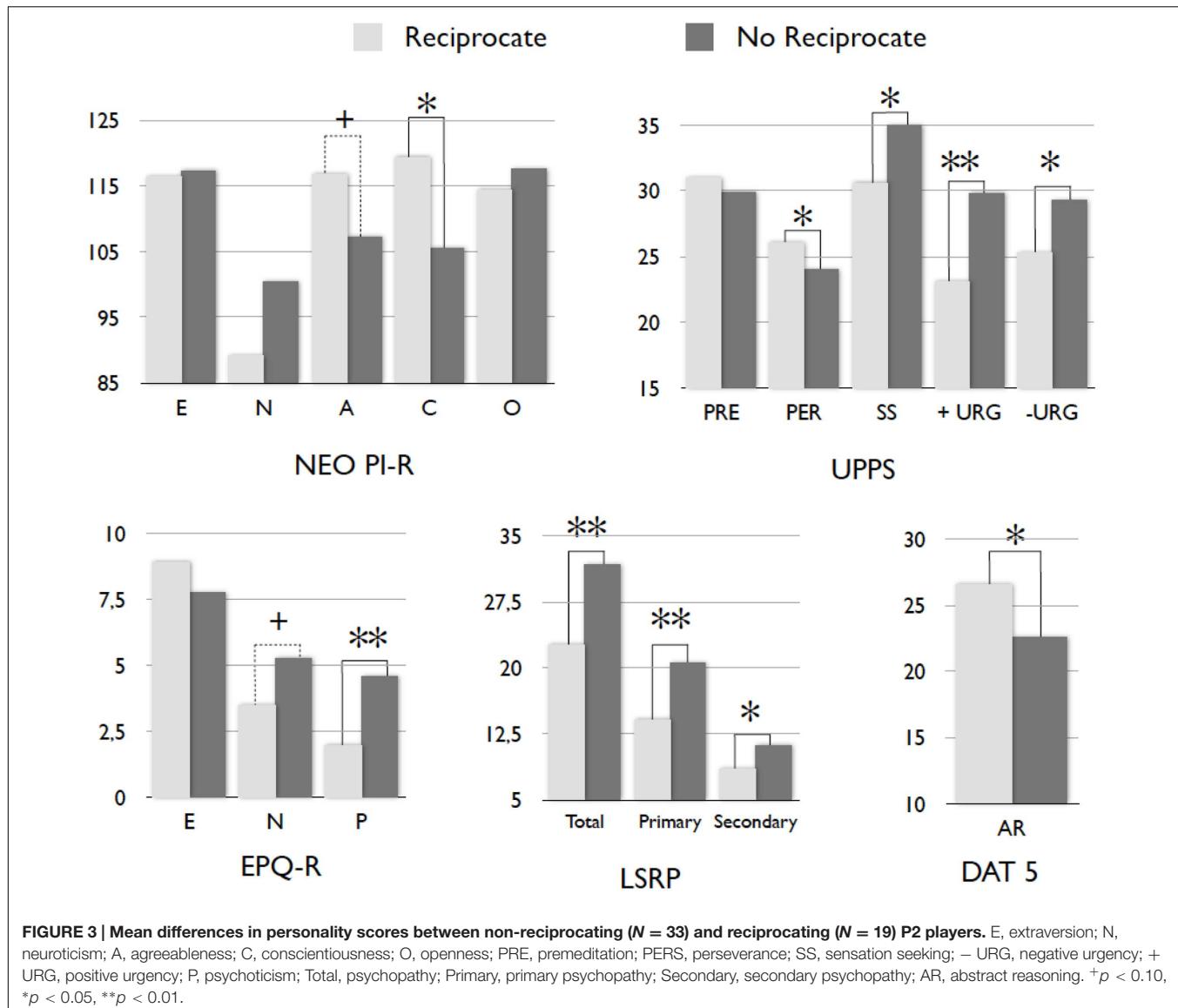
**FIGURE 2 | Mean differences in personality scores between non-trusting ( $N = 58$ ) and trusting ( $N = 52$ ) P1 players. E, extraversion; N, neuroticism; A, agreeableness; C, conscientiousness; O, openness; PRE, premeditation; PERS, perseverance; SS, sensation seeking; - URG, negative urgency; + URG, positive urgency; P, psychoticism; Total, psychopathy; Primary, primary psychopathy; Secondary, secondary psychopathy; AR, abstract reasoning.  $+p < 0.10$ ,  $**p < 0.01$ .**

urgency in two facets, negative urgency, and positive urgency (Cyders et al., 2007; Cyders and Smith, 2008). These facets were conceived as reflecting different ‘pathways’ to impulsive behavior. Accordingly, we found perseverance and deliberation to be closely linked to C, sensation seeking to E, and negative and positive urgency to N, although positive urgency would also be associated to low A and low C, in line with past research (Cyders and Smith, 2008).

With respect to the individual differences in the TG, first we deal with trusting behavior. Different approaches have been proposed to define and explain trust behavior (see Bauer, 2015). Recently Thielmann and Hilbig (2015b) have systematically reviewed the multiple basic processes underlying trusting behavior among strangers and its relationship to personality characteristics. They proposed that four main components would be relevant in the decision to trust: (a) attitudes toward risky prospects (i.e., risk aversion and loss aversion), (b) betrayal sensitivity, (c) trustworthiness expectations, and (d) sensitivity to reward. Importantly, individual differences in these

processes would be casually linked to personality characteristics, so examining the relationship between personality and trust behavior would help in determining which of these mechanisms could be more relevant in the TG.

According to our results, the main mechanisms involved in trusting behavior in our experiment would be Reward sensitivity. Thielmann and Hilbig (2015b) suggested that some individuals might place attention on the potential reward inherent in a positive social interaction, so, individuals more sensible to reward, i.e., scoring high in Extraversion-related traits, should perceive social interactions as particularly rewarding *per se* and therefore be highly motivated to approach such interactions (Depue and Collins, 1999; Denissen and Penke, 2008). Accordingly, we found that *Extraversion/positive emotionality*, and specifically the facet of *warmth* associate to trust. People scoring high in warmth are friendly, easily forming close attachment to others (Costa and McCrae, 1999). In accordance to our results, some other studies have also found a similar role of E on trusting behavior (Swope et al., 2008;



Ben-Ner and Halldorsson, 2010; Haring et al., 2013), suggesting that trustors' investments have a component of facilitation of social relations by expecting a large gain from trust. This interpretation would be reaffirmed by the fact that we have also found an association of trust and positive urgency, the tendency to engage in rash action in response to high positive affect (Cyders and Smith, 2008), suggesting that part of this behavior is linked to a non-deliberative rash behavior in front of a perceived appetitive situation.

In contrast to our hypothesis, we have not found any association between A and trust. The hypothetical process underlying the relevance of A on trust would be the development of trustworthiness expectations via social projection. To form an expectation about the other's likely behavior, the trustor can consider different sources of information, as trust cues (i.e., reputation), prior trust experiences, or social projection (Thielmann and Hilbig, 2015b). Social projection implies that

people would predict others cooperativeness by projecting their own cooperative preferences onto them (Krueger, 2013). In terms of the FFM, one's cooperation and trustworthiness should be mainly covered by the A domain, so agreeable people would expect others to behave more cooperatively and reciprocate. Accordingly, Evans and Revelle (2008) and Becker et al. (2012) found a slight but significant effect of A on the amounts invested in the TG, and Müller and Schwieren (2012) confirmed the relevance of *trust* and *straightforwardness* for this behavior. However, in line of our results, other studies have not found association between A and investment behavior (Swope et al., 2008; Ben-Ner and Halldorsson, 2010; Haring et al., 2013). The fact that we and others have failed to find significant associations could be reflecting the difficulty in detecting modest effect sizes, as those described for the associations between A and investment behavior (Zhao and Smillie, 2015).

**TABLE 3 | Hierarchical Logistic Regression analysis with Trust and Reciprocate behavior as dependent variables.**

		Trust (52 vs. 58)		Reciprocate (33 vs. 19)	
		R <sup>2</sup> (Cox and Snell-Nagelkerke)	B	R <sup>2</sup> (Cox and Snell-Nagelkerke)	B
Step 1	Gender (0 = male; 1 = female)	0.003–0.004	−0.232	0.003–0.004	−0.360
	Age		0.000		−0.086
Step 2	Fluid intelligence	0.005–0.007	−0.015	0.107–0.146*	0.102*
	Unconscious disinhibition	0.056–0.075	−0.182	0.322–0.441**	−0.738 <sup>+</sup>
Step 3	Negative emotionality		−0.003		−0.693 <sup>+</sup>
	Positive emotionality		0.485*		0.276
	Disagreeable disinhibition		−0.096		−1.057*
	Risk aversion HL	0.066–0.088	−0.022	0.337–0.462	0.038
Step 4	Risk aversion F1 SGG		−0.128		0.042
	Risk aversion F2 SGG		0.058		−0.432

Gender, age, fluid intelligence and personality factor scores as predictors (N in parenthesis). +p < 0.10, \*p < 0.05, \*\*p < 0.01.

Our data also indicate the minor role on trust of the other two proposed mechanisms, betrayal sensitivity and attitudes toward risky prospects (Thielmann and Hilbig, 2015b). In terms of FFM, individual differences in these mechanisms would be linked to N, mainly the facet of *angry hostility* for betrayal sensitivity ((Maltby et al., 2008; Thielmann and Hilbig, 2015b) and the facet of *anxiety* for attitudes toward risk (Thielmann and Hilbig, 2015b). However, in line with previous findings (Evans and Revelle, 2008; Swope et al., 2008; Becker et al., 2012), no association between trust behavior and N-domain nor its facets are found. In addition, no association between trusting behavior and risk aversion measures have been found (Bohnet and Zeckhauser, 2004; Bohnet et al., 2008; Houser et al., 2010). These findings are important because they reinforce the idea that risk attitudes would not be successful in organizing trust behavior (Fehr, 2009). Thus, and to sum up, our data suggest that the more important mechanism underlying individual differences in the TG was sensitivity to reward. Attitudes toward risky prospects, betrayal sensitivity or trustworthiness expectations would exert a minor role, presenting low effect sizes that would be difficult to detect with the sample size used in the present research.

Once P1 has decided to cooperate (i.e., trust), P2 can exploit the other's trust or can correspond with reciprocity. Reciprocity constitutes a key mechanism for explaining cooperative behavior among non-relatives, receiving strong attention from several disciplines, especially economics and evolutionary biology (Trivers, 1971; Fehr and Fischbacher, 2003; Falk and Fischbacher, 2006; Nowak, 2006; Tooby et al., 2006; Guala, 2012). Reciprocity could be understood as the tendency to respond "nicely" to nice actions (positive reciprocity) and "nastily" to nasty actions (negative reciprocity) when interacting with other players. Reciprocity can be beneficial for both parts (weak reciprocity), or even may involve a cost for responders (strong reciprocity). Cooperation usually emerges in repeated encounters within the same pair of individuals, helping each other (direct reciprocity). Nevertheless, cooperation is also extensively observed between strangers, probably because of an expected indirect gain (indirect reciprocity) like good reputation.

Conversely, a non-reciprocal subject may benefit from exploiting others' trust. Exploitative behavior has received some

attention recently, specifically from an evolutionary perspective (Mealey, 1995; Buss and Duntley, 2008; Lalumière et al., 2010; Glenn et al., 2011). According to this view, exploitation is a main class of strategies for acquiring reproductively relevant resources that consist in expropriating the resources of others through exploitation. This class of strategies ranges from mild, such as failing to reciprocate a minor favor in a social exchange, to extreme, such as coalitional warfare to expropriate all of an opposing group's reproductively relevant assets (Buss and Duntley, 2008). The personality characteristic most strongly associated to exploitation would be low A and its extreme, psychopathy (Buss, 2009). From an evolutionary point of view, psychopathic behavior would constitute a successful alternative strategy at a low relative frequency in the population, whereby a small number of individuals take advantage of their more populous, cooperative counterparts by defecting in social interactions (Mealey, 1995; Lalumière et al., 2010; Glenn et al., 2011). Surprisingly, psychopathic traits had not been formerly explored in the TG so far.

As a result of our approach, we obtained the novel finding that those individuals that did not reciprocate were higher in *Disagreeable disinhibition*. Specifically, non-reciprocators scored higher in both primary and secondary Levenson psychopathy scales, P, and low C. Conversely, the decision to reciprocate in order to reward a kind action would depend on the A FFM dimension, and, specifically, on *straightforwardness*. Individuals scoring high in *straightforwardness* would be honest, sincere and ingenuous, whereas low scorers would be dishonest and would tend to manipulate others through flattery or deception (Costa and McCrae, 1999). Along this line, some studies have found that the most relevant personality domain for reciprocation is A (Ben-Ner and Halldorsson, 2010; Becker et al., 2012; Lönnqvist et al., 2012), especially its honesty aspects (Thielmann and Hilbig, 2015a). Thus, from an evolutionary personality perspective, reciprocal-exploitative behaviors would be located on a continuum of opposite strategies regarding behavior in cooperative situations, and the personality domain linked to this continuum would be the dimension of A.

In addition, our results also suggest that impulsivity would play a relevant role in trust and, especially in reciprocal behavior.

To our knowledge, the present study is the first to systematically examine the role of this complex trait in the TG. We have found that a specific facet of impulsivity, positive urgency, is related to trusting behavior. Positive urgency refers to the tendency to engage in rash action as a response to high positive affect. This suggests that trusting behavior would be considered as a positive and potentially rewarding situation and that the decision to trust is partially guided by impulsive tendencies. In the same vein, reciprocal behavior also involves a non-reflexive component of the take-the-money-and-run type behavior, with individuals who are more sensitive to reward (sensation seeking), less perseverant, and score higher in urgency, both positive and negative, presenting rash responses of non-reciprocal behavior. We think that these results, if replicated, could be theoretically relevant since they point to a the role of hot impulsive and non-reflexive mechanisms at the basis of trust (e.g., Murray et al., 2011) and reciprocity, in contrast to a more classical view of economic decisions associated with a more cold reflexive and calculative vision of human behavior.

The physiological results show that P1 participants scoring high in psychopathy exhibit increased EDA at the moment in which they are asked to decide whether to trust. At the same time, the P1 group show reduced evoked HR deceleration, indicating decreased attentional engagement during the decision-making process. Taken together, these two findings suggest that high psychopathy scorers perceive the decision-making task as less demanding compared to low-scorers, despite physiological changes signaling increased emotional arousal. No significant differences in EDA or HR variation arise between trusting vs. non-trusting or reciprocating vs. non-reciprocating participants.

According to the somatic marker hypothesis, decision-making is influenced by physiological signals that arise in bioregulatory processes, including those expressed as emotions (Damasio, 1996). Numerous studies have shown that emotional activation guides decision making in healthy subjects, while this effect is reduced in patients with orbitofrontal dysfunction (Bechara et al., 2000). Interestingly, psychopathic personality traits and antisocial behavior (clinical and sub-clinical) have been linked to orbitofrontal dysfunction (Dinn et al., under review). While previous research associated psychopathic behavior with reduced EDA (Lorber, 2004; Casey et al., 2013), our findings may indicate an alternative mechanism to promote antisocial behavior by suppressing the influence of somatic markers in decision making.

This study has several limitations. First, the magnitude of personality association with trust is modest and, therefore, some effects may not have been detected due to the relatively small sample size. Although the effects were greater in magnitude for reciprocal behavior, the reduced number of participants in the reciprocating and non-reciprocating groups led to a low statistical power in part of our analysis. Also in relation to the sample, it is important to highlight that our results are referred to non-clinical population, and therefore, the generalization to clinically relevant samples such as psychopaths should be made with caution. Another limitation, and a potentially source of discrepancies with other studies, is the discrete TG version used in present experiment, in contrast to the more usual continuous

version used. Nevertheless, one strength of the present analysis is the inclusion not only of many personality domains, but also of specific traits relevant for particular behaviors (such as psychopathy for non-reciprocal behavior). However, and even though fluid intelligence has been used as a marker of general cognitive ability (Colom et al., 2007), other cognitive abilities have not been examined (McGrew, 2009). Thus, future research would benefit from including a larger number of participants, the use of clinical samples, and a broader selection of personality, economic and cognitive variables.

To conclude, although A and E are primarily dimensions of interpersonal behavior, E is related to the preferred quantity of social stimulation and A represents the characteristic quality of the interaction (Costa et al., 1991). Accordingly, the present study suggests that E could be relevant for initiating cooperation, whereas A could be relevant for maintaining it. That is, different personality domains would represent different strategies in the social domain, one based in the number of social contacts and the other in the cohesion of such contacts. With respect to the E domain, high E would favor a risky behavior that may increase the number of social partners. On the other hand, individuals scoring high in A would reward kind actions, even if this reward involves some cost for them. Conversely, low agreeable/high psychopathic and disinhibited/rash impulsive individuals would benefit from this situation, by taking the money and running!

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the ethical committee at the Universitat Jaume I. The deputy chair of the LEE ethics committee, Dr. Eva Camacho led the process in this specific case. Participants gave written informed consent in accordance with the Declaration of Helsinki.

## AUTHOR CONTRIBUTIONS

NG, GO, GS-G, and MI designed the general study. NG had the original idea of this specific paper. GS-G, SL-O, LM, and HV collected the data. MI, AG-G, IB-T, and LM performed the statistical analyses. PP designed, collected and analyzed the physiological data. MI, GS-G, and NG wrote the first manuscript draft. AG-G organized the database and coordinated the final version. All the authors contributed to and approved the final manuscript.

## FUNDING

Financial support by Universitat Jaume I (project P1.1B2015-48), the Spanish Ministry of Economics and Competitiveness (projects ECO2013-44409-P, ECO2015-68469-R and PSI2015-67766-R), the Bank of Spain Excellence Chair in Computational Economics (project 11I229.01/1) and the Generalitat Valenciana (project GV/2016/158) is gratefully acknowledged.

## REFERENCES

- Aluja, A., García, O., and García, L. F. (2004). Replicability of the three, four and five Zuckerman's personality super-factors: exploratory and confirmatory factor analysis of the EPQ-RS, ZKPQ and NEO-PI-R. *Pers. Ind. Diff.* 36, 1093–1108. doi: 10.1016/S0191-8869(03)00203-4
- Attanasi, G., Georgantzis, G., Rotondi, V., and Vigani, D. (2016). *Lottery- and Survey-based Risk Attitudes Linked Through A Multichoice Elicitation Task*. Technical report, Bureau d'Economie Théorique et Appliquée. Strasbourg: Unified Diagnostic Services.
- Balliet, D., and van Lange, P. A. M. (2013). Trust, conflict, and cooperation: a meta-analysis. *Psychol. Bull.* 139, 1090–1112.
- Bauer, P. C. (2015). *Three Essays on the Concept of Trust and its Foundations*. Dissertation, University of Bern, Bern.
- Bechara, A., Damasio, H., and Damasio, A. R. (2000). Emotion, decision making, and the orbitofrontal cortex. *Cereb. Cortex* 10, 295–307. doi: 10.1093/cercor/10.3.295
- Becker, A., Deckers, T., Dohmen, T., Falk, A., and Kosse, F. (2012). The relationship between economic preferences and psychological personality measures. *Annu. Rev. Econ.* 4, 453–478. doi: 10.1146/annurev-economics-080511-110922
- Ben-Ner, A., and Halldorsson, F. (2010). Trusting and trustworthiness: what are they, how to measure them, and what affects them. *J. Econ. Psychol.* 31, 64–79. doi: 10.1016/j.jeop.2009.10.001
- Bennett, G. K., Seashore, H. G., and Wesman, A. G. (2005). *Differential Aptitude Test (DAT-5)*. Madrid: TEA Ediciones
- Benning, S. D., Patrick, C. J., Hicks, B. M., Blonigen, D. M., and Krueger, R. F. (2003). Factor structure of the psychopathic personality inventory: validity and implications for clinical assessment. *Psychol. Assess.* 15, 340–350. doi: 10.1037/1040-3590.15.3.340
- Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games Econ. Behav.* 1, 122–142. doi: 10.1006/game.1995.1027
- Bohnet, I., and Zeckhauser, R. (2004). Trust, risk and betrayal. *J. Econ. Behav. Organ.* 55, 467–484. doi: 10.1016/j.jebo.2003.11.004
- Bohnet, I., Greig, F., Herrmann, B., and Zeckhauser, R. (2008). Betrayal aversion: evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *Am. Econ. Rev.* 98, 294–310. doi: 10.1257/aer.98.1.294
- Borghans, L., Duckworth, A. L., Heckman, J. J., and ter Weel, B. (2008). The economics and psychology of personality traits. *J. Hum. Resour.* 43, 972–1059. doi: 10.2307/40057376
- Burks, S. V., Carpenter, J. P., Goette, L., and Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7745–7750. doi: 10.1073/pnas.0812360106
- Buss, D. M. (2009). How can evolutionary psychology successfully explain personality and individual differences? *Perspect. Psychol. Sci.* 4, 359–366. doi: 10.1111/j.1745-6924.2009.01138.x
- Buss, D. M., and Duntley, J. D. (2008). Adaptations for exploitation. *Group Dynam. Theor. Res. Pract.* 12, 53–62. doi: 10.1037/1089-2699.12.1.53
- Camerer, C. F., Loewenstein, G., and Rabin, M. (eds.). (2011). *Advances in Behavioral Economics*. Princeton, NJ: Princeton University Press.
- Carl, N., and Billari, F. C. (2014). Generalized trust and intelligence in the United States. *PLoS ONE* 9:e91786. doi: 10.1371/journal.pone.0091786
- Carvalho, J. L. A., da Rocha, A. F., Nascimento, F. A. O., Souza Neto, J., and Junqueira, L. F. Jr. (2002). "Development of a Matlab software for analysis of heart rate variability," in *Proceedings of the ICSP'02, 6th International Conference on Signal Processing*, eds Y. Baozong and T. Xiaofang (Beijing: IEEE Press), 1488–1491.
- Casey, H., Rogers, R. D., Burns, T., and Yiend, J. (2013). Emotion regulation in psychopathy. *Biol. Psychol.* 92, 541–548. doi: 10.1016/j.biopsych.2012.06.011
- Cesarini, D., Dawes, C. T., Fowler, J. H., Johannesson, M., Lichtenstein, P., and Wallace, B. (2008). Heritability of cooperative behavior in the trust game. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3721–3726. doi: 10.1073/pnas.0710069105
- Coleman, J. (1990). *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.
- Colom, R., Escorial, S., Shih, P. C., and Privado, J. (2007). Fluid intelligence, memory span, and temperament difficulties predict academic performance of young adolescents. *Pers. Ind. Differ.* 42, 1503–1514. doi: 10.1016/j.paid.2006.10.023
- Corgnet, B., Espin, A., Hernan-Gonzalez, R., Kujal, P., and Rassenti, S. (2015). To trust, or not to trust: cognitive reflection in trust games. *J. Behav. Exp. Econ.* 64, 20–27. doi: 10.1016/j.soec.2015.09.008
- Costa, P. T., McCrae, R., and Dye, D. A. (1991). Domains and facets scales for agreeableness and conscientiousness: a revision of the NEO personality inventory. *J. Pers. Assess.* 12, 887–898.
- Costa, P. T., and McCrae, R. R. (1999). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI). Professional Manual*. Madrid: TEA Ediciones.
- Costa-Gomes, M. A., Huck, S., and Weizsäcker, G. (2014). Belief and actions in the trust game: creating instrumental variables to estimate the causal effect. *Games Econ. Behav.* 88, 298–309. doi: 10.1016/j.geb.2014.10.006
- Crosen, R., and Gneezy, U. (2009). Gender differences in preferences. *J. Econ. Literat.* 47, 448–474. doi: 10.1257/jel.47.2.448
- Curry, O., Chesters, M. J., and Viding, E. (2011). The psychopath's dilemma: the effects of psychopathic personality traits in one-shot games. *Pers. Ind. Differ.* 50, 804–809. doi: 10.1016/j.paid.2010.12.036
- Cyders, M. A., and Smith, G. T. (2008). Emotion-based dispositions to rash action: positive and negative urgency. *Psychol. Bull.* 134, 807–828. doi: 10.1037/a0013341
- Cyders, M. A., Smith, G. T., Spillane, N. S., Fischer, S., Annus, A. M., and Peterson, C. (2007). Integration of impulsivity and positive mood to predict risky behavior: development and validation of a measure of positive urgency. *Psychol. Assess.* 19, 107–118. doi: 10.1037/1040-3590.19.1.107
- Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philos. Trans. R. Soc. B Biol. Sci.* 351, 1413–1420. doi: 10.1098/rstb
- Denissen, J. J. A., and Penke, L. (2008). Motivational individual reaction norms underlying the Five-Factor model of personality: first steps towards a theory-based conceptual framework. *J. Res. Pers.* 42, 1285–1302. doi: 10.1016/j.jrp.2008.04.002
- Depue, R. A., and Collins, P. F. (1999). Neurobiology of the structure of personality: dopamine, facilitation of incentive motivation, and extraversion. *Behav. Brain Sci.* 22, 491–569. doi: 10.1017/S0140525X99002046
- Dinh, T., Perrault, H., Calabrese, P., Eberhard, A., and Benchettit, G. (1999). New statistical method for detection and quantification of respiratory sinus arrhythmia. *IEEE Trans. Biomed. Eng.* 46, 1161–1165. doi: 10.1109/10.784148
- Evans, A. M., and Krueger, J. I. (2009). The psychology (and economics) of trust. *Soc. Pers. Psychol. Compass* 3, 1003–1017. doi: 10.1111/j.1751-9004.2009.00232.x
- Evans, A. M., and Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *J. Res. Pers.* 42, 1585–1593. doi: 10.1016/j.jrp.2008.07.011
- Evenend, J. L. (1999). Varieties of impulsivity. *Psychopharmacology* 146, 348–361. doi: 10.1007/PL000005481
- Eysenck, H. J. (1992). The definition and measurement of psychotism. *Pers. Ind. Differ.* 13, 757–785. doi: 10.1016/0191-8869(92)90050-Y
- Falk, A., and Fischbacher, U. (2006). A theory of reciprocity. *Games Econ. Behav.* 54, 293–315. doi: 10.1016/j.geb.2005.03.001
- Fehr, E. (2009). On the economics and biology of trust. *J. Eur. Econ. Assoc.* 7, 235–266. doi: 10.1162/JEEA.2009.7.2-3.235
- Fehr, E., and Fischbacher, U. (2003). The nature of human altruism. *Nature* 425, 785–791. doi: 10.1038/nature02043
- Ferguson, E., Heckman, J. J., and Corr, P. (2011). Personality and economics: overview and proposed framework. *Pers. Ind. Differ.* 51, 201–209. doi: 10.1016/j.paid.2011.03.030
- Gambetta, D. (1988). "Can we trust?" in *Trust: Making and Breaking Cooperative Relations*, ed. D. Gambetta (Cambridge: Blackwell), 213–237.
- García-Gallego, A., Georgantzis, N., Jaramillo, A., and Parravano-Gutiérrez, M. (2012). The lottery-panel task for bi-dimensional parameter-free elicitation of risk attitudes. *Rev. Int. Sociol.* 70, 53–72. doi: 10.3989/ris.2011.10.30
- Gaughan, E. T., Miller, J. D., and Lynam, D. R. (2012). Examining the utility of general models of personality in the study of psychopathy: a comparison of the HEXACO-PI-R and NEO PI-R. *J. Personal. Disord.* 26, 513–523. doi: 10.1521/pedi\_2012\_26\_017
- Gillespie, S. M., Mitchell, I. J., Johnson, I., Dawson, E., and Beech, A. R. (2013). Exaggerated intergroup bias in economical decision making games: differential effects of primary and secondary psychopathic traits. *PLoS ONE* 8:e69565. doi: 10.1371/journal.pone.0069565

- Glenn, A. L., Kurzban, R., and Raine, A. (2011). Evolutionary theory and psychopathy. *Aggress. Violent Behav.* 16, 371–380. doi: 10.1016/j.avb.2011.03.009
- Guala, F. (2012). Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate. *Behav. Brain Sci.* 35, 1–15. doi: 10.1017/S0140525X11000069
- Gunnthorsdottir, A., McCabe, K., and Smith, V. (2002). Using the Machiavellianism instrument to predict trustworthiness in a bargaining game. *J. Econ. Psychol.* 23, 49–66. doi: 10.1016/S0167-4870(01)00067-8
- Haring, K. S., Matsumoto, Y., and Watanabe, K. (2013). “How do people perceive and trust a lifelike robot,” in *Proceedings of the World Congress of Engineering and Computer Science 2013*, San Francisco, CA, 23–25.
- Heckman, J. J. (2011). *Integrating Personality Psychology Into Economics*. Available at: <http://www.nber.org/papers/w17378> and <http://www.nber.org/papers/w17378.ack>
- Holt, C. A., and Laury, S. K. (2002). Risk aversion and incentive effects. *Am. Econ. Rev.* 92, 1644–1655. doi: 10.1257/000282802762024700
- Hooghe, M., Marien, S., and de Vroome, T. (2012). The cognitive basis of trust. The relation between education, cognitive ability, and generalized and political trust. *Intelligence* 40, 604–613. doi: 10.1016/j.intell.2012.08.006
- Houser, D., Schunk, D., and Winter, J. (2010). Distinguishing trust from risk: an anatomy of the investmentgame. *J. Econ. Behav. Organ.* 74, 72–81. doi: 10.1016/j.jebo.2010.01.002
- Ibáñez, M. I., Moya, J., Villa, H., Mezquita, L., Ruipérez, M. Á., and Ortet, G. (2010). Basic personality dimensions and alcohol consumption in young adults. *Pers. Ind. Differ.* 48, 171–176. doi: 10.1016/j.paid.2009.09.017
- John, O. P., Naumann, L. P., and Soto, C. J. (2008). “Paradigm shift to the integrative big-five trait taxonomy: history, measurement, and conceptual issues,” in *Handbook of Personality: Theory and Research*, 3rd Edn, eds O. P. John, R. W. Robins, and L. A. Pervin (New York, NY: Guilford Press), 114–153.
- Johnson, N. D., and Mislin, A. A. (2011). Trust games: a meta-analysis. *J. Econ. Psychol.* 32, 865–889. doi: 10.1016/j.joep.2011.05.007
- Johnston, L., Hawes, D. J., and Straiton, M. (2014). Psychopathic traits and social cooperation in the context of emotional feedback. *Psychiatry Psychol. Law* 21, 767–778. doi: 10.1080/13218719.2014.893550
- Jones, G. (2008). Are smarter groups more cooperative? Evidence from prisoner’s dilemma experiments, 1959–2003. *J. Econ. Behav. Organ.* 68, 489–497. doi: 10.1016/j.jebo.2008.06.010
- Kahneman, D. (2003). Maps of bounded rationality: psychology for behavioral economics. *Am. Econ. Rev.* 93, 1449–1475. doi: 10.1257/000282803322655392
- King-Casas, B., and Chiu, P. H. (2012). Understanding interpersonal function in psychiatric illness through multiplayer economic games. *Biol. Psychiatry* 72, 119–125. doi: 10.1016/j.biopsych.2012.03.033
- King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., and Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science* 321, 806–810. doi: 10.1126/science.1156902
- Koenigs, M., Kruepke, M., and Newman, J. P. (2010). Economic decision-making in psychopathy: a comparison with ventromedial prefrontal lesion patients. *Neuropsychologia* 48, 2198–2204. doi: 10.1016/j.neuropsychologia.2010.04.012
- Krueger, J. I. (2013). Social projection as a source of cooperation. *Curr. Dir. Psychol. Sci.* 22, 289–294. doi: 10.1177/0963721413481352
- Lalumière, M. L., Mishra, S., and Harris, G. T. (2010). “In cold blood: the evolution of psychopathy,” in *Evolutionary Forensic Psychology: Darwinian Foundations of Crime and Law*, eds J. D. Duntley and T. K. Shackelford (New York, NY: Oxford University Press), 176–197.
- Loewenstein, G., Rick, S., and Cohen, J. D. (2008). Neuroeconomics. *Annu. Rev. Psychol.* 59, 647–672. doi: 10.1146/annurev.psych.59.103006.093710
- Lönnqvist, J.-E., Verkasalo, M., Wichardt, P. C., and Walkowitz, G. (2012). Personality disorder categories as combinations of dimensions: translating cooperative behavior in borderline personality disorder into the Five-Factor framework. *J. Personal. Disord.* 26, 298–304. doi: 10.1521/pedi.2012.26.2.298
- Lorber, M. F. (2004). Psychophysiology of aggression, psychopathy, and conduct problems: a meta-analysis. *Psychol. Bull.* 130, 531–552. doi: 10.1037/0033-2909.130.4.531
- Luhmann, N. (1988). “Familiarity, confidence, trust, problems and alternatives,” in *Trust: Making and Breaking Cooperative Relations*, ed. D. Gambetta (Oxford: Basil Blackwell).
- Lynam, D. R., Gaughan, E. T., Miller, J. D., Miller, D. J., Mullins-Sweatt, S., and Widiger, T. (2011). Assessing the basic traits associated with psychopathy: development and validation of the elemental psychopathy assessment. *Psychol. Assess.* 23, 108–124. doi: 10.1037/a0021146
- Lynam, D. R., Whiteside, S., and Jones, S. (1999). Self-reported psychopathy: a validation study. *J. Pers. Assess.* 73, 110–132. doi: 10.1207/S15327752JPA730108
- Maltby, J., Wood, A. M., Day, L., Kon, T. W. H., Colley, A., and Linley, P. A. (2008). Personality predictors of levels of forgiveness two and a half years after the transgression. *J. Res. Pers.* 42, 1088–1094. doi: 10.1016/j.jrp.2007.12.008
- Markon, K. E., Krueger, R. F., and Watson, D. (2005). Delineating the structure of normal and abnormal personality: an integrative hierarchical approach. *J. Pers. Soc. Psychol.* 88, 139–157. doi: 10.1037/0022-3514.88.1.139
- McCrae, R. R., and Costa, P. T. (2008). “Empirical and theoretical status of the Five-Factor Model of personality traits,” in *Personality Theory and Assessment: Personality Theories and Models*, Vol. 1, eds G. J. Boyle, G. Matthews, and D. H. Saklofske (London: Sage), 273–294.
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: standing on the shoulders of the giants of psychometric intelligence research. *Intelligence* 37, 1–10. doi: 10.1016/j.intell.2008.08.004
- Mealey, L. (1995). The sociobiology of sociopathy: an integrated evolutionary model. *Behav. Brain Sci.* 18, 523–599. doi: 10.1017/S0140525X00039595
- Miller, J. D., Gaughan, E. T., and Pryor, L. R. (2008). The levenson self-report psychopathy scale: an examination of the personality traits and disorders associated with the LSRP factors. *Assessment* 15, 450–463. doi: 10.1177/1073191108316888
- Miller, J. D., and Lynam, D. R. (2003). Psychopathy and the five-factor model of personality: a replication and extension. *J. Pers. Assess.* 81, 168–178. doi: 10.1207/S15327752JPA8102\_08
- Miller, J. D., Lynam, D. R., Widiger, T. A., and Leukefeld, C. (2001). Personality disorders as extreme variants of common personality dimensions: can the Five-Factor Model adequately represent psychopathy? *J. Pers.* 69, 253–276. doi: 10.1111/1467-6494.00144
- Mokros, A., Menner, B., Eisenbarth, H., Alpers, G. W., Lange, K. W., and Osterheider, M. (2008). Diminished cooperativeness of psychopaths in a prisoner’s dilemma game yields higher rewards. *J. Abnorm. Psychol.* 117, 406–413. doi: 10.1037/0021-843X.117.2.406
- Montañés-Rada, F., de Lucas-Taracena, M. T., and Martín-Rodríguez, M. A. (2003). Evaluación del trastorno antisocial de personalidad con el dilema del prisionero [Antisocial personality disorder evaluation with the prisoner’s dilemma game]. *Actas Españo. Psiquiatr.* 31, 307–314.
- Müller, J., and Schwieren, C. (2012). *What can the Big Five Personality Factors Contribute to Explain Small-Scale Economic Behavior?* Amsterdam: Tinbergen Institute Discussion Paper, 1–24.
- Murray, S. L., Pinkus, R. T., Holmes, J. G., Harris, B., Gomillion, S., Aloni, M., et al. (2011). Signaling when (and when not) to be cautious and self-protective: impulsive and reflective trust in close relationships. *J. Pers. Soc. Psychol.* 101, 485–502. doi: 10.1037/a0023233
- Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science* 314, 1560–1563. doi: 10.1126/science.1133755.Five
- Ortet, G., Ibáñez, M. I., Llerena, A., and Torrubia, R. (2002). Underlying traits of the karolinska scales of personality (KSP). *Eur. J. Psychol. Assess.* 18, 139–148. doi: 10.1027//1015-5759.18.2.139
- Ortet, G., Ibáñez, M. I., Moro, M., and Silva, F. (2001). “Versión corta del EPQ-R (EPQ-RS),” in *Short Version of the EPQ-R (EPQ-RS)*, eds H. J. Eysenck and S. B. G. Eysenck (Madrid: TEA Ediciones).
- Perakakis, P., Joffily, M., Taylor, M., Guerra, P., and Vila, J. (2010). KARDIA: a Matlab software for the analysis of cardiac interbeat intervals. *Comput. Methods Prog. Biomed.* 98, 83–89. doi: 10.1016/j.cmpb.2009.10.002
- Perakakis, P., Guinot, J. V., Conde, A., Jaber-López, T., García-Gallego, A., and Georgantzis, N. (2013). *A Technical Note on the Precise Timing of Behavioral Events in Economic Experiments, Working Papers 2013/12, Economics Department*. Castelló: Jaume I University.
- Rilling, J. K., Glenn, A. L., Jairam, M. R., Pagnoni, G., Goldsmith, D. R., Elfenbein, H. A., et al. (2007). Neural correlates of social cooperation and non-cooperation as a Function of Psychopathy. *Biol. Psychiatry* 61, 1260–1271. doi: 10.1016/j.biopsych.2006.07.021

- Sabater-Grande, G., and Georgantzis, N. (2002). Accounting for risk aversion in repeated prisoners' dilemma games: an experimental test. *J. Econ. Behav. Organ.* 48, 37–50. doi: 10.1016/S0167-2681(01)00223-2
- Samuel, D. B., and Widiger, T. A. (2008). A meta-analytic review of the relationships between the five-factor model and DSM-IV-TR personality disorders: a facet level analysis. *Clin. Psychol. Rev.* 28, 1326–1342. doi: 10.1016/j.cpr.2008.07.002
- Saulsman, L. M., and Page, A. C. (2004). The five-factor model and personality disorder empirical literature: a meta-analytic review. *Clin. Psychol. Rev.* 23, 1055–1085. doi: 10.1016/j.cpr.2002.09.001
- Schniter, E., and Sheremeta, R. M., and Shields, T. W. (2016). *The Problem with All-Or-Nothing Trust Games: What Others Choose Not To Do Matters in Trust-Based Exchange*. Available at: <https://ssrn.com/abstract=2708869>
- Sharp, C. (2012). The use of neuroeconomic games to examine social decision making in child and adolescent externalizing disorders. *Curr. Dir. Psychol. Sci.* 21, 183–188. doi: 10.1177/0963721412444726
- Sturgis, P., Read, S., and Allum, N. (2010). Does intelligence foster generalized trust? An empirical test using the UK birth cohort studies. *Intelligence* 38, 45–54. doi: 10.1016/j.intell.2009.11.006
- Swope, K. J., Cadigan, J., Schmitt, P. M., and Shupp, R. (2008). Personality preferences in laboratory economics experiments. *J. Socio-Econ.* 37, 998–1009. doi: 10.1016/j.socloc.2006.12.065
- Thielmann, I., and Hilbig, B. E. (2015a). The traits one can trust: dissecting reciprocity and kindness as determinants of trustworthy behavior. *Pers. Soc. Psychol. Bull.* 41, 1523–1536. doi: 10.1177/0146167215600530
- Thielmann, I., and Hilbig, B. E. (2015b). Trust : an integrative review from a person – situation perspective. *Rev. Gen. Psychol.* 19, 249–277. doi: 10.1037/gpr0000046
- Tooby, J., Cosmides, L., and Price, M. E. (2006). Cognitive adaptations for n-person exchange: the evolutionary roots of organizational behavior. *Manage. Decis. Econ.* 27, 103–129. doi: 10.1002/mde.1287
- Trivers, R. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57. doi: 10.1086/406755
- Unoka, Z., Seres, I., Aspán, N., Bódi, N., and Kéri, S. (2009). Trust game reveals restricted interpersonal transactions in patients with borderline personality disorder. *J. Personal. Disord.* 23, 399–409. doi: 10.1521/pedi.2009.23.4.399
- Verdejo-García, A., Lozano, O., Moya, M., Alcázar, M. A., and Pérez-García, M. (2010). Psychometric properties of a Spanish version of the UPPS-P impulsive behavior scale: reliability, validity and association with trait and cognitive impulsivity. *J. Pers. Assess.* 92, 70–77. doi: 10.1080/00223890903382369
- Vukasovic, T., and Bratko, D. (2015). Heritability of personality : a meta-analysis of behavior genetic studies. *Psychol. Bull.* 141, 769–785. doi: 10.1037/bul0000017
- Walker, J., and Ostrom, E. (2009). “Trust and reciprocity as foundations for cooperation,” in *Whom Can We Trust? How Groups, Networks, and Institutions Make Trust Possible*, Chap. 4, eds K. Cook, L. Levi, and R. Hardin (New York, NY: Russell Sage Foundation), 91–124.
- Whiteside, S. P., Lynam, D., Miller, J., and Reynolds, S. (2005). Validation of the UPPS impulsive behavior scale: a four factor model of impulsivity. *Eur. J. Pers.* 19, 559–574. doi: 10.1002/per.556
- Whiteside, S. P., and Lynam, D. R. (2001). The five factor model and impulsivity: using a structural model of personality to understand impulsivity. *Pers. Ind. Diff.* 30, 669–689. doi: 10.1016/S0191-8869(00)00064-7
- Widiger, T. A., and Mullins-Sweatt, S. N. (2009). Five-factor model of personality disorder: a proposal for DSM-V. *Annu. Rev. Clin. Psychol.* 5, 197–220. doi: 10.1146/annurev.clinpsy.032408.153542
- Zhao, K., and Smillie, L. D. (2015). The role of interpersonal traits in social decision making: exploring sources of behavioral heterogeneity in economic games. *Pers. Soc. Psychol. Rev.* 19, 277–302. doi: 10.1177/1088868314553709
- Zuckerman, M., Kuhlman, D. M., Teta, P., Joireman, J., and Kraft, M. (1993). A comparison of three structural models of personality: the big three, the big five, and the alternative five. *J. Pers. Soc. Psychol.* 65, 757–768. doi: 10.1037/0022-3514.65.4.757

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer PB and the handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2016 Ibáñez, Sabater-Grande, Barreda-Tarazona, Mezquita, López-Ovejero, Villa, Perakakis, Ortet, García-Gallego and Georgantzis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Personality Trait of Intolerance to Uncertainty Affects Behavior in a Novel Computer-Based Conditioned Place Preference Task

Milen L. Radell<sup>1,2</sup>, Catherine E. Myers<sup>1,2</sup>, Kevin D. Beck<sup>1,2</sup>, Ahmed A. Moustafa<sup>3</sup> and Michael Todd Allen<sup>4\*</sup>

<sup>1</sup> Department of Veterans Affairs, Veterans Affairs New Jersey Health Care System, East Orange, NJ, USA, <sup>2</sup> Department of Pharmacology, Physiology and Neuroscience, New Jersey Medical School, Rutgers University, Newark, NJ, USA, <sup>3</sup> School of Social Sciences and Psychology and Marcs Institute for Brain and Behaviour, University of Western Sydney, Sydney, NSW, Australia, <sup>4</sup> School of Psychological Sciences, University of Northern Colorado, Greeley, CO, USA

## OPEN ACCESS

**Edited by:**

Manuel Ignacio Ibáñez,  
Jaume I University, Spain

**Reviewed by:**

Karl Friston,  
University College London, UK  
Patrick Anselme,  
University of Liège, Belgium

**\*Correspondence:**

Michael Todd Allen  
michael.allen@unco.edu

**Specialty section:**

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

**Received:** 01 June 2016

**Accepted:** 25 July 2016

**Published:** 09 August 2016

**Citation:**

Radell ML, Myers CE, Beck KD, Moustafa AA and Allen MT (2016) The Personality Trait of Intolerance to Uncertainty Affects Behavior in a Novel Computer-Based Conditioned Place Preference Task. *Front. Psychol.* 7:1175.  
doi: 10.3389/fpsyg.2016.01175

Recent work has found that personality factors that confer vulnerability to addiction can also affect learning and economic decision making. One personality trait which has been implicated in vulnerability to addiction is intolerance to uncertainty (IU), i.e., a preference for familiar over unknown (possibly better) options. In animals, the motivation to obtain drugs is often assessed through conditioned place preference (CPP), which compares preference for contexts where drug reward was previously received. It is an open question whether participants with high IU also show heightened preference for previously rewarded contexts. To address this question, we developed a novel computer-based CPP task for humans in which participants guide an avatar through a paradigm in which one room contains frequent reward (i.e., rich) and one contains less frequent reward (i.e., poor). Following exposure to both contexts, subjects are assessed for preference to enter the previously rich and previously poor room. Individuals with low IU showed little bias to enter the previously rich room first, and instead entered both rooms at about the same rate which may indicate a foraging behavior. By contrast, those with high IU showed a strong bias to enter the previously rich room first. This suggests an increased tendency to chase reward in the intolerant group, consistent with previously observed behavior in opioid-addicted individuals. Thus, the personality factor of high IU may produce a pre-existing cognitive bias that provides a mechanism to promote decision-making processes that increase vulnerability to addiction.

**Keywords:** uncertainty, decision making, conditioned place preference (CPP), personality, addiction, humans

## INTRODUCTION

Some individuals exposed to drugs of abuse develop addiction while others do not. One factor mediating this difference in outcomes may be personality traits that confer biases in decision making, such as a tendency to pursue familiar sources of reward at the expense of exploring other (possibly more rewarding) options. Such individual differences have been studied in the context of anxiety, but some of the same personality traits may also confer vulnerability to addiction. Addiction has a high comorbidity rate with anxiety disorders (Merikangas et al., 1998;

Grant et al., 2004). Based on their comorbidity, it is not surprising that both types of disorders share other common features, including behaviors such as withdrawal or avoidance, changes in learning, and maladaptive decision making (e.g., risk taking, chasing reward). This alteration in decision making is not limited to decisions about drugs, but can also affect reward in general (Clark and Robbins, 2002). Drug use continues regardless of the negative consequences (e.g., to health, income, family), as do anxiety behaviors. Addiction and anxiety also share some common neural mechanisms. Both come about through some form of associative learning to a maladaptive stimulus, specifically, anxiety via altered associative learning in the amygdala (Packard and Cahill, 2001; Packard, 2009) and addiction through altered reward learning in the mesolimbic dopamine system (Robinson and Berridge, 2001; Everitt and Robbins, 2005; Volkow et al., 2010). In addition, stress and anxiety can lead to increased drug use and relapse (Jacobsen et al., 2001; Sinha, 2001).

## Learning, Personality and Vulnerability

Some recent work has examined the effects of personality on vulnerability for anxiety disorders, and to a lesser extent, addiction. Overall, the results suggest that personality factors, including behavioral inhibition (BI) and harm avoidance, hypothesized to be risk factors for anxiety disorders, are associated with enhanced learning in a variety of tasks (Sheynin et al., 2013, 2014; Allen et al., 2014; Holloway et al., 2014). For example, BI is a temperamental tendency to withdraw from or avoid novel social and non-social situations (Kagan et al., 1987; Morgan, 2006). In addition to avoidance, BI includes social reticence and enhanced reactivity to novelty, threat, and uncertainty (Hirshfeld et al., 1992; Schwartz et al., 2003). BI has long been considered a vulnerability factor for the development of anxiety-related disorders including post-traumatic stress disorder (Myers et al., 2012; Clauss et al., 2015). Behaviorally inhibited individuals exhibit enhanced associative learning as measured by eyeblink conditioning with a tone conditioned stimulus (CS) and a corneal air puff unconditioned stimulus (US) (Allen et al., 2014, 2016; Holloway et al., 2014), and with increased avoidance in a computer-based task (Sheynin et al., 2014). Enhanced avoidance learning was also observed in male, but not female, opioid addicts undergoing methadone maintenance therapy, when compared to controls, using the same task (Sheynin et al., 2016).

In addition to these findings with basic classical conditioning and avoidance learning, the effects of personality factors have been examined with computer-based tasks involving economic decision making. For example, Radell et al. (2016) used a cognitive economic decision making task based on social interactions (i.e., the trust game) with behaviorally inhibited individuals. This task, based on the version used by Delgado et al. (2005), had participants read the biographies of partners in the game that portrayed them as morally trustworthy ("good partner"), untrustworthy ("bad partner"), or neutral ("neutral partner"). On each trial, participants were shown a partner and were given a choice of keeping \$1 or sharing \$3 with that partner. If the money was shared, the partner had the choice of

keeping it all or reciprocating by returning half of the money (\$1.50). On any trial in which the participant chose to share, the partner always reciprocated with 50% probability, irrespective of how they were portrayed in the biography. Inhibited individuals tended to share with the neutral partner less than uninhibited individuals; however, this behavioral difference was not evident in the ratings of trustworthiness for the "neutral partner." These results suggest that inhibited individuals may be predisposed to interpret neutral or ambiguous information more negatively, which may contribute to the tendency to avoid unfamiliar people characteristic of behaviorally inhibited temperament, and its relationship to anxiety disorders.

Probabilistic category learning tasks that include both reward and punishment trials have also revealed a role for anxiety vulnerability factors in economic decision making (Sheynin et al., 2013). On each trial, participants view a stimulus and are asked to categorize it. The categories are probabilistic in that each stimulus is a member of one category 80% of the time and a member of the other category 20% of the time. For some stimuli, correct categorization results in a reward (point gain) and incorrect categorization results in no feedback; for other stimuli, incorrect categorization results in a punishment (point loss) and correct categorization results in no feedback. Thus, performance on reward and punishment trials can be directly contrasted, as can the interpretation of the ambiguous "no-feedback" outcome, which can signal either failure to obtain reward or successful avoidance of punishment. Behaviorally inhibited individuals demonstrated better associative learning on both reward and punishment trials. Given the option to opt out of individual trials to avoid any chance of being punished or rewarded, inhibited individuals also preferred to opt out to avoid punishment (Sheynin et al., 2013). In a follow-up study, using this task, participants with severe symptoms of post-traumatic stress disorder exhibited enhanced learning, specifically on reward trials, relative to peers with fewer or no symptoms (Myers et al., 2013).

Extending this task to the topic of addiction, Myers et al. (2016) found that opioid-addicted individuals undergoing methadone maintenance therapy were more likely to abandon previous response rules and explore new alternatives when expectancies were violated (i.e., increased lose-shift behavior), relative to controls. Thus, addicted participants tended to respond based on immediate feedback, which may explain why they continue to pursue short-term reward while ignoring the long-term negative consequences of drug use (Myers et al., 2016). Likewise, in other decision-making tasks, addicts tend to choose small immediate rewards over larger delayed rewards, and display a number of other changes in decision making compared to control participants (Petry et al., 1998; Clark and Robbins, 2002). Additionally, at least some of these changes appear to persist even after long-term abstinence (Li et al., 2013). However, it is important to note that a preference for small immediate rewards over large delayed rewards is not specific to addicts – it has been shown in both humans and animals, and is a function of multiple factors including the length of the delay, age, intelligence (Mischel and Metzner, 1962), and the amount of reward (Green et al., 1997). Thus, addiction is only associated with an exaggeration

of this preference, which may reflect increased impulsivity or reduced self-control (Madden et al., 2003).

## The Role of Uncertainty

One common feature of most of the tasks discussed above is some aspect of uncertainty that was associated with performance improvements. Acquisition of eyeblink conditioning was enhanced in anxiety-vulnerable individuals under protocols which included schedules of partial reinforcement with 50% CS alone and 50% US alone trials (Allen et al., 2014), and variability in trial timing (Allen et al., 2016). In contrast, vulnerability did not modulate performance on a standard 100% CS-US paired trials protocol. In the computer avoidance task (Sheynin et al., 2014), participants were given no instructions and had to learn, through trial-and-error, what behavior resulted in avoiding point loss. In the trust game (Radell et al., 2016), all partners shared 50% of the time regardless of the nature of their biographies but individuals with anxiety vulnerability only differed in how they treated the neutral partner. The probabilistic category learning task (Myers et al., 2016) involved uncertainty in that it was not possible to be correct 100% of the time based on the probabilistic nature of the categories. There was also a mix of reward and punishment trials, and no feedback was given on correct punishment trials and incorrect reward trials. Finally, tasks that pit immediate small rewards against larger delayed rewards (Petry et al., 1998; Clark and Robbins, 2002) may also involve perceived uncertainty in that there is no guarantee that the delayed reward will actually be received.

Given the possible role of uncertainty in most prior tasks examining the role of individual differences in anxiety and addiction vulnerability, the purpose of the current study was to test how personality can modulate economic decision making for rewards in healthy individuals, focusing on intolerance to uncertainty (IU) – another personality factor that has been linked to anxiety disorders (Dugas et al., 1997; Ladouceur et al., 1997; Birrell et al., 2011; Carleton, 2012; Grupe and Nitschke, 2013). IU can be defined as a tendency to perceive uncertain situations as aversive and stressful, and respond with BI and negative expectations about their possible consequences (Nelson et al., 2015). Initially, IU was linked to generalized anxiety disorder, and is a strong predictor of the tendency to worry (Dugas et al., 1997; Ladouceur et al., 1997). However, other studies suggest that it is not specific to that disorder, but constitutes a broader risk factor for the development and maintenance of anxiety and depression (Tolin et al., 2003; Carleton, 2012; Carleton et al., 2012).

In recent work, individuals undergoing treatment for opioid dependence had significantly higher IU, as measured with the IU scale (Carleton et al., 2007), compared to healthy controls, suggesting that IU may also be a risk factor in substance abuse and addiction. This evidence is, of course, correlational and a causal relationship, if any, remains to be established. Still, IU implies reduced risk taking, in contrast to substance abuse and addiction, associated with increased impulsivity and risk taking. Thus, if higher IU does contribute to addiction vulnerability, this relationship may be indirect and only appear in a subpopulation of individuals who, for example, may have started substance use as a form of self-medication for anxiety. Along the same

lines, pathological gambling – also associated with increased risk taking – is also often comorbid with anxiety disorders (Lorains et al., 2011), which are, in contrast, linked to higher risk aversion (Maner and Schmidt, 2006) and greater IU (Ladouceur et al., 1997). As with the relationship between IU and addiction, these contradictory findings might be resolved if pathological gamblers are not a homogenous group of individuals, but rather consist of multiple subtypes, only one of which represents impulsive risk-takers (Blaszczyński and Noweł, 2002). IU has also been linked to changes in economic decision making and reward system function (Nelson et al., 2015). Using a gambling task, Nelson et al. (2015) found IU could modulate event-related potential responses to gains and losses, which have been linked to activation in the ventral striatum and medial prefrontal cortex, and activation in the anterior cingulate cortex, respectively. Individuals with higher IU are more likely to perceive situations as uncertain, and have stronger emotional responses (e.g., increased anxiety) under those conditions (Ladouceur et al., 1997). They also tend to require additional information before making a decision, and paradoxically avoid cues that can lead to anxiety, which would in practice reduce the amount of information available for decision making (Ladouceur et al., 1997; Krain et al., 2006). Similar to drug addicted individuals, individuals with higher IU were more likely to choose small, low-probability rewards over larger but delayed high-probability rewards (Luhmann et al., 2011).

## Conditioned Place Preference

We sought to continue this line of research by investigating the role of IU on learning in a computer-based economic decision making task, similar to the conditioned place preference (CPP) paradigm widely applied to the study of addiction in animal models. CPP has been commonly used to measure the reward value of different drugs of abuse (for reviews, see Bardo and Bevins, 2000; Tzschentke, 2007). Here, drug-free subjects (typically rodents), are first allowed to explore an apparatus consisting of at least two distinct interconnected chambers to measure initial preference (i.e., by comparing time spent in each context). In subsequent conditioning sessions, the animal is injected with a drug and confined to one chamber. Similarly, the other context is paired with saline. Finally, drug-free subjects are once again allowed to choose between the compartments in order to assess preference. A large number of studies have shown animals spend more time in the drug-paired than in the saline-paired compartment for a wide variety of drugs, including opioids (e.g., heroin, methadone), and psychomotor stimulants (e.g., cocaine, amphetamine) (Bardo et al., 1995). CPP has also been observed for non-drug rewards including food (Sprakraki et al., 1982), water, and access to sexual interaction (Oldenburger et al., 1992) or a running wheel (Lett et al., 2000).

Here, we report results from a computer-based CPP task where humans guide an avatar through a paradigm in which one room contains frequent reward and one contains less frequent reward. Following exposure to both contexts, participants were assessed for preference to enter the previously rich and previously poor room. IU was assessed via a self-report questionnaire. An important limitation of animal CPP as a model of human

substance abuse and addiction is that rewards are simply administered by the experimenter, and are not contingent on behavior (e.g., animals are injected with drug or confined to a compartment containing reward). In contrast, humans choose to start taking the drug and control the frequency of administration. To address this concern, in the current task, obtaining reward was contingent on operant responding by the participants. We predicted that if IU contributes to decision making that can promote substance abuse and addiction, individuals with higher IU should show a stronger bias toward the previously rich room, compared to individuals with lower IU, who might be more prone to explore other options.

## MATERIALS AND METHODS

### Participants

A total of 88 participants were recruited from the University of Northern Colorado, and received research credit in a psychology class as payment for their participation. Data from 12 participants were lost due to computer failure. The remaining sample ( $n = 76$ ) contained 50 females and had a mean age of 20.7 ( $SD = 5.4$ , range = 18–56), and education of 13.8 years ( $SD = 1.4$ , range = 12–17). All participants provided informed consent before initiation of any behavioral testing. Procedures were approved by the Institutional Review Board at the University of Northern Colorado, and conformed to guidelines established by the Federal Government and the Declaration of Helsinki for the protection of human subjects.

### Procedure

Testing took place in a quiet room. All participants completed the brief, 12-item version of the Intolerance to Uncertainty Scale (IUS-12; Carleton et al., 2007), and a computer-based CPP task programmed in the Java 8 language (Oracle Corporation, Redwood City, CA, USA), administered on a desktop computer running Windows. The task, illustrated in **Figure 1**, consisted of a tutorial, pretest, training and a posttest phase. Participants controlled a cartoon avatar (a fox) and were instructed to help the fox collect as many golden eggs as possible. The exact instructions are provided in the Appendix. The task began with a tutorial where the fox was placed in a lobby area with a single door in the middle. Participants were told that they could click on the door to switch between rooms. Once they did, the fox entered a room with eight chests, and participants were prompted to click on the chests to collect two golden eggs. When the participant clicked on a chest, the fox moved to inspect that chest. The chest was then opened to reveal whether an egg was inside. During the tutorial, all chests always contained eggs. Therefore, the subject's first two choices were always rewarded. The total score (i.e., total eggs collected by participants) was always visible at the top of the screen.

Next, participants began the pretest, during which they once again started in the lobby area (**Figure 1B**), but were given a choice between two doors (blue and brown) on the sides of the room. The left or right placement of the two doors was counterbalanced across participants. The doors led to two visually

distinct rooms (blue and brown, **Figures 1C,D**), which contained eight chests each arranged in a circular pattern. Both rooms were visually distinct from each other, and from the room encountered in the tutorial. For the next 4 min, participants were allowed to freely explore the virtual environment, switching between rooms and clicking on chests to acquire eggs. During the pretest, each chest had an initial 5% chance of containing reward. Throughout the task, whenever an egg was found in a particular chest, the chest's subsequent chance of reward decreased to 0, and increased back to the maximum at increments of 1% every 4 s. Thus, repeatedly searching the same chest was not encouraged. Rather, the optimum strategy was to move around a room exploring different chests. Participants, however, were not told anything about reward contingencies and had to rely on trial-and-error. The amount of time spent in each of the rooms, the total number and order of chest clicks, and the total score was recorded. For each subject, the room (blue or brown) where that participant had spent more time during the pretest was defined as the "more preferred" room and the other as the "less preferred" room.

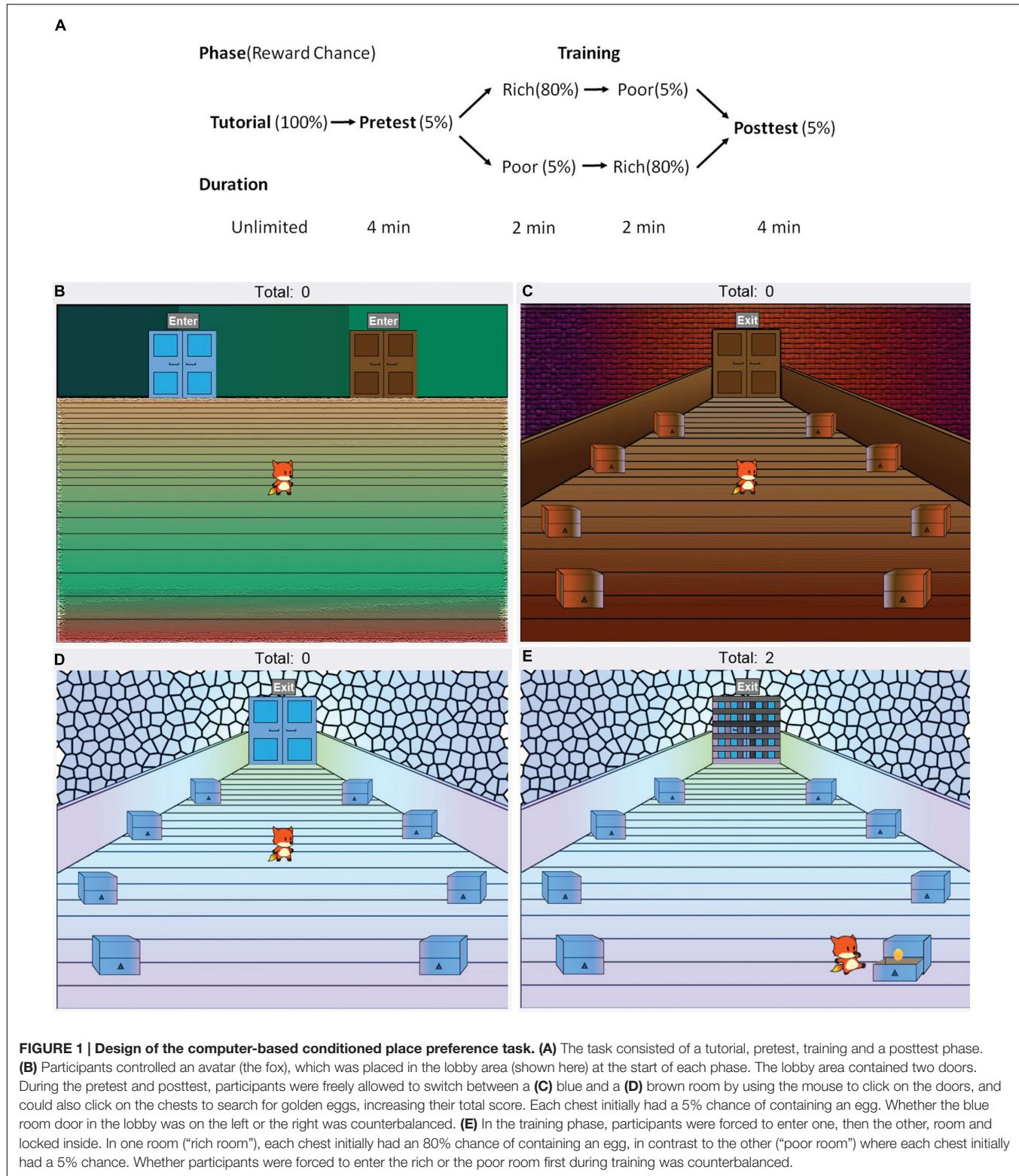
The pretest was followed by the training phase, which consisted of two parts (2 min each). At the start of each part of the training phase, the fox was placed in the lobby, but only one of the doors was available, forcing participants to enter one of the side rooms. Once they entered the room, they were locked in (**Figure 1E**) and had to remain there until the second part of training. The second part of training began in the same way, with the fox placed in the lobby and only the remaining door available. The less preferred room during the pretest was assigned to be the rich room, meaning that each chest had an initial 80% chance of containing an egg. The other room was assigned to be the poor room, where each chest had an initial 5% chance of containing an egg. As in the pretest, once an egg was found in a chest, reward chance decreased to 0 and gradually increased back to initial levels at increments of 10% (for the rich room) or 1% for the poor room, every 4 s. Whether participants were locked in the rich or the poor room first was counterbalanced. Again, the order and number of chests clicked was recorded, along with the number of eggs obtained.

Finally, participants completed a posttest, which was identical to the pretest. The fox was placed in the lobby with both blue and brown rooms freely available. All chests had an initial 5% chance of containing an egg. Here, the first room entered by participants, and the time spent in each room (previously rich vs. previously poor) were recorded, along with the order and number of chests clicked and the number of eggs obtained. After the task, all participants completed a questionnaire (see the Appendix) about their knowledge of reward contingencies, whether or not they had a strategy, and their computer or video game experience.

## RESULTS

### Questionnaires

The mean score on the IUS-12 was 32.25 ( $SD = 8.58$ , range = 14–57). For all analyses, subjects were split into high or low IU



groups based on the sample median of 32, with 37 participants (25 female) classed as low, and 39 (25 female) classed as high. The high and low IU groups did not differ in gender distribution,  $\chi^2(1) = 0.101, p = 0.750$ , or age,  $t(74) = 0.284, p = 0.778$ .

In the post-task questionnaire, in response to "Did you think that one of the rooms had more eggs in it?" 78.9% of participants responded "yes,"  $\chi^2(1) = 25.5, p < 0.001$ . Out of those who said "yes," 90% also correctly identified the rich room,  $\chi^2(1) = 38.4,$

$p < 0.001$ . Thus, most participants were explicitly aware of which room was more rewarding. Finally, 64.5% of the participants reported they had previously played computer or video games,  $\chi^2(1) = 6.4$ ,  $p = 0.012$ , and 61.8% reported they had followed a specific strategy while searching for eggs,  $\chi^2(1) = 4.3$ ,  $p = 0.039$ . Among the strategies mentioned were going in circles or zig zags and checking all of the chests once then switching rooms.

## Conditioned Place Preference Task

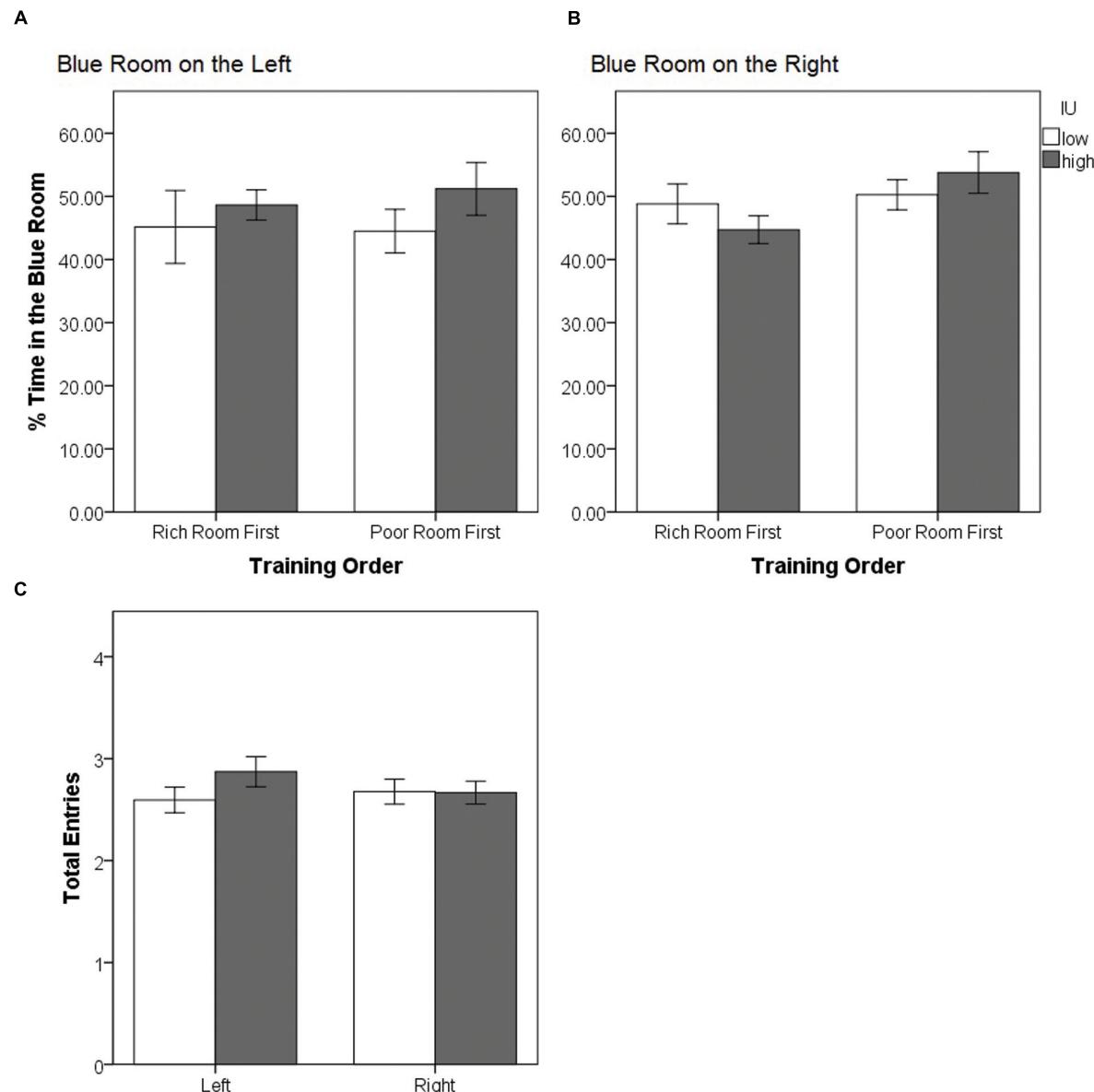
Since participants were assigned to one of four conditions to counterbalance which context (blue or brown) was on the left or right in the lobby, and whether the rich or the poor room was experienced first during training, we first examined whether this led to an initial preference bias as a function of IU. The mean percent of the time participants spent in the blue room during the pretest was computed as total time spent in the blue side room divided by sum of the total time spent in the blue and brown rooms (**Figures 2A,B**). A 2 (blue on left vs. right)  $\times$  2 (rich room first vs. second)  $\times$  2 (IU high vs. low) between-subjects ANOVA on the percent time spent in the blue room during the pretest confirmed that there were no significant main effects (all  $F < 1.720$ , all  $p > 0.19$ ) or interactions (all  $F < 1.320$ , all  $p > 0.25$ ). Thus, on average, participants tended to divide their time equally, spending approximately 50% of the time in each of the side rooms, eliminating initial bias as a potential explanation of the results in subsequent analyses. The average total number of entries made into each room by high and low IU participants was also examined (**Figure 2C**). A 2 (left vs. right room)  $\times$  2 (IU high vs. low) mixed-model ANOVA confirmed there were no significant main effects (both  $F < 0.900$  and  $p > 0.34$ ), and no significant interaction,  $F(1,74) = 1.618$ ,  $p = 0.207$ . Thus, both groups of participants had enough time to make multiple visits to each room during the pretest.

As our primary analysis, we examined whether participants tended to enter the previously rich or the previously poor room first at the start of the posttest, i.e., whether they first entered the room paired with a high chance of reward (maximum 80%) or a low chance of reward (maximum 5%) during the training phase. It is important to note that during the posttest, both rooms were once again equivalent and paired with a low chance of reward (maximum 5%) as in the pretest. As expected, most participants entered the previously-rich room first,  $\chi^2(1) = 10.32$ ,  $p = 0.001$  (**Figure 3A**). However, surprisingly, approximately 30% of participants instead chose to enter the previously poor room. This could be due to differences in personality between participants, or a function of whether the last room experienced during the training phase was the rich or the poor room. To examine this possibility, we performed log-linear analysis – an extension of the chi-square test used for more than two categorical variables – on the total number of participants with factors of the first room entered during the posttest, the last room (rich or poor) experienced during the training phase, and IU (high or low). A non-hierarchical (forced-entry) method was used to enter factors into the model. The log-linear analysis produced a model that retained only the main effects and two-way interactions, and had a perfect fit to the data. The only significant two-way interaction was between the first room entered in the posttest

and IU,  $X_p^2(1) = 4.578$ ,  $p = 0.032$ . The three-way interaction and the remaining two-way interactions (first room entered in posttest  $\times$  last room in training and IU  $\times$  last room in training) were not significant (all  $X^2 < 3.7$ , all  $p$ 's  $> 0.05$ ). **Figure 3B** shows the percent of the total participants as a function of whether they entered the rich or the poor room first, and IU. Based on the odds ratio, participants who had high IU had 3.87 times higher odds of first going to the rich room in the posttest compared to participants who had low IU. Thus, participants with high IU tended to show greater CPP by going back to the previously rewarded context (i.e., followed a win-stay strategy) while those with low IU instead explored a different room (i.e., followed a win-shift strategy). The absence of other significant effects in the loglinear analysis suggests that this behavior was specifically a function of IU rather than other variables, such as which room participants had most recently been in during the prior training phase.

Similar analyses were performed to eliminate other possible confounds. IU and the first room entered in the posttest were always included in the model, while the third factor was whether or not participants reported they knew which room had more eggs (i.e., knew the rich room), had previous game experience or reported following a specific strategy. There was a significant three-way interaction between knowledge of the rich room, IU and the first room entered in the posttest,  $\chi^2(1) = 6.028$ ,  $p = 0.014$ . To examine this interaction, a total of four Bonferroni-corrected two-sided Fisher's exact tests (alpha adjusted to  $0.05/4 = 0.0125$ ) were performed with factors of IU and first room entered in the posttest. The first two tests were performed separately on individuals who reported they knew vs. did not know which room had more eggs. The result was significant only for individuals who reported they knew the rich room ( $p = 0.001$  vs.  $p = 0.518$ ). The second set of tests were performed on the subset of individuals who reported they knew the rich room, split by whether they also correctly identified that room. This confirmed a significant difference only for those who could identify the room ( $p = 0.001$  vs.  $p = 0.467$ ). Therefore, the interaction between IU and the first room entered during the posttest appears driven by participants who could explicitly identify the rich room. To avoid confusion, note that test statistics are not generated for Fisher's exact test, therefore only  $p$ -values are reported. Finally, when game experience was examined, the model retained only the main effects and two-way interactions – the only significant two-way interaction was once again between IU and the first room entered in the posttest,  $X_p^2(1) = 8.684$ ,  $p = 0.003$ . Similarly this was the only significant two-way interaction when whether or not participants had a strategy was included as the third factor,  $X_p^2(1) = 7.3$ ,  $p = 0.007$ . Thus, neither game experience nor following a strategy were related to IU, or to which room participants entered first during the posttest. Across analyses, this depended on IU, and was also related to explicit knowledge of the rich room.

Having entered one room first in the posttest, we next examined whether participants tended to stay there, spending more time, overall, in that room. A 2 (rich vs. poor room entered first)  $\times$  2 (IU high vs. low) ANOVA was performed on the percent of the total time spent in the rich room during the posttest

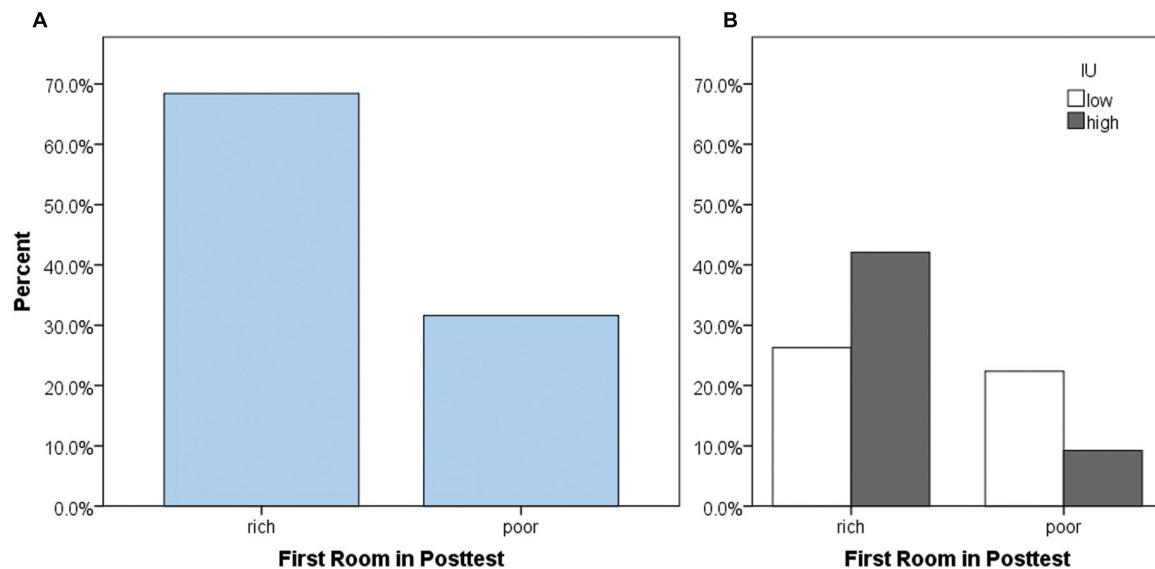


**FIGURE 2 | (A,B)** Mean percent time spent in the blue room during the pretest as a function of training order and IU. There were no significant differences confirming that initially, all groups had no strong preference and divided their time equally between the blue and brown rooms. This was irrespective of which room was on the right or on the left, and whether the less-preferred room (assigned to be the rich room) was experienced first or second during training. **(C)** Average total number of entries into the left or right room during the pretest as a function of IU. All participants, irrespective of IU, made a similar number of visits to each room. Error bars represent  $\pm$  SEM.

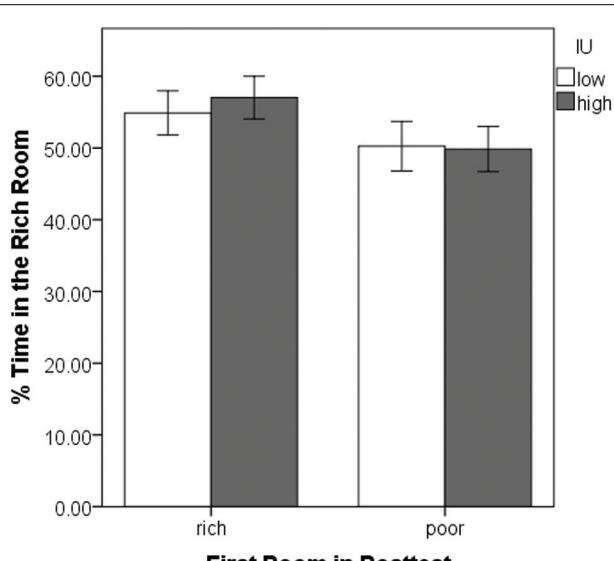
(Figure 4). This was calculated as total time in the rich room divided by total time in the rich plus the poor room. There were no significant differences (all  $p$ 's  $> 0.05$ ). Thus, despite the initial preference to enter the previously rich room, most participants did not simply remain in the originally chosen room. Rather, across the whole posttest, participants tended to divide their time equally between the two rooms.

Next, we assessed locomotion in the posttest, first considering movement between rooms (Figure 5), then total chest clicks within each room (Figure 6). A mixed-model ANOVA was

performed on total side room entries during the posttest with a within-subjects factor of the room entered (rich or poor), and between-subjects factors of the first room entered during the posttest (rich or poor) and IU (high or low). This yielded significant interactions between the first room entered and IU,  $F(1,72) = 4.71$ ,  $p = 0.033$ ,  $\eta_p^2 = 0.061$ , and between total entries into the rich or poor rooms and the first room entered,  $F(1,72) = 24.702$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.255$ . There were no other significant interactions or main effects (all  $p$ 's  $> 0.05$ ). Post hoc Bonferroni-corrected independent samples  $t$ -tests were



**FIGURE 3 | Percent of total participants who chose to enter the previously rich or previously poor room first, at the start of the posttest. (A)** While most participants chose to enter the previously rich room first, surprisingly, about 30% went to the previously poor room first. **(B)** This behavior depended on IU. Participants with high IU were more likely to enter the previously rich room first in the posttest, and less likely to enter the previously poor room first, compared to participants with low IU.



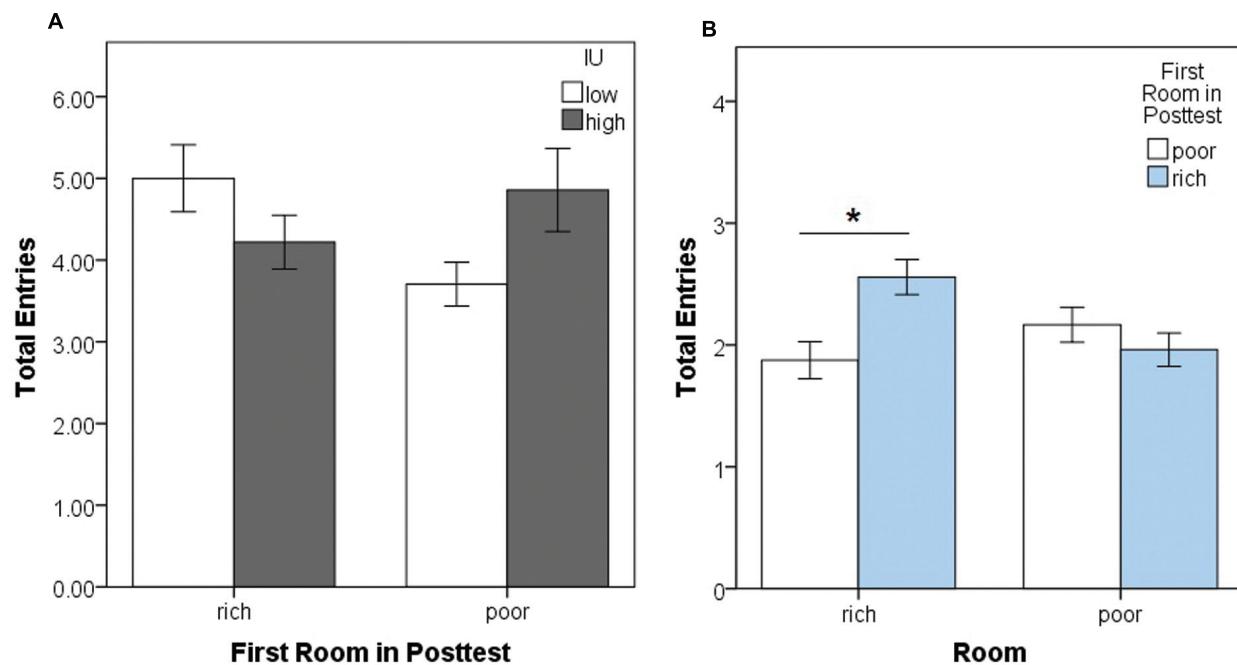
**FIGURE 4 | Mean percent time spent in the previously rich room during the posttest as a function of the first room entered and IU.** There were no significant differences. Error bars represent  $\pm$  SEM.

conducted to further examine the significant interactions (alpha adjusted to  $0.05/4 = 0.0125$ ). The interaction between the first room entered and total room entries appeared to be driven by individuals with high IU making more entries into the poor room (**Figure 5A**), however, the test comparing entries into the previously poor room entries by high vs. low IU participants

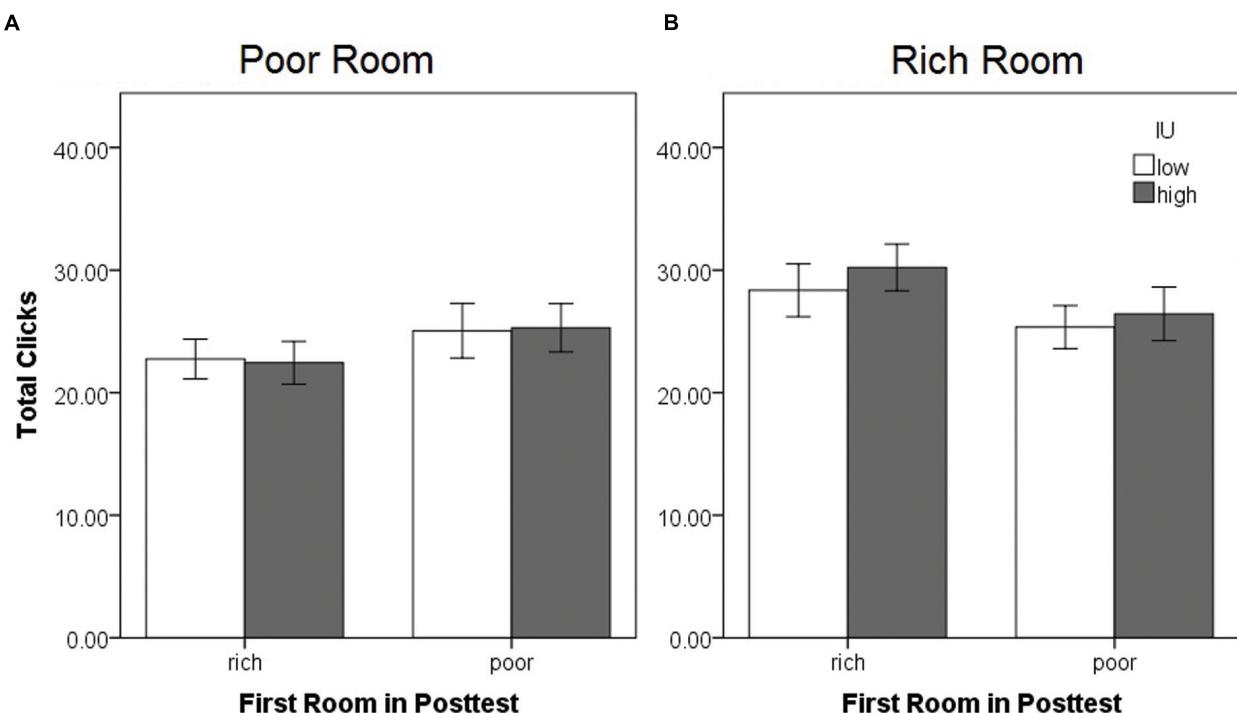
failed to reach corrected significance,  $t(22) = 2.18$ ,  $p = 0.04$ . The interaction between total entries and the first room entered was due to participants who first entered the rich room tending to make more re-entries into that same room throughout the posttest,  $t(60.92) = 3.27$ ,  $p = 0.002$ ,  $r = 0.39$  (**Figure 5B**). There was no significant difference in entries into the poor room as a function of which room was entered first during training. Overall, it is important to note that while there were some significant differences, effect sizes are small and the differences amounted to, on average, one or two additional room entries. More importantly, these data indicate participants remained active and continued to switch between rooms throughout the posttest.

A mixed-model ANOVA was also performed on the total number of chest clicks (sum of the clicks on all eight chests) within each side room (**Figure 6**), with within-subjects factor of the room (rich or poor), and between-subjects factors of the first room entered (rich or poor) and IU (high or low). There were no significant differences (all  $p > 0.05$ ). Sample graphs of the path taken by two individuals during the pretest and posttest, one from the low IU and one from the high IU groups (**Figure 7**), also indicate that participants remained motivated, continuing to switch rooms and check different reward locations, throughout the task. Note that while some individuals did show a strong preference for one room during the posttest (**Figure 7B**), on average, participants spent approximately equal amounts of time in both rooms, irrespective of IU. In contrast, as described earlier, individuals with high IU tended to visit the previously rich room first during the posttest.

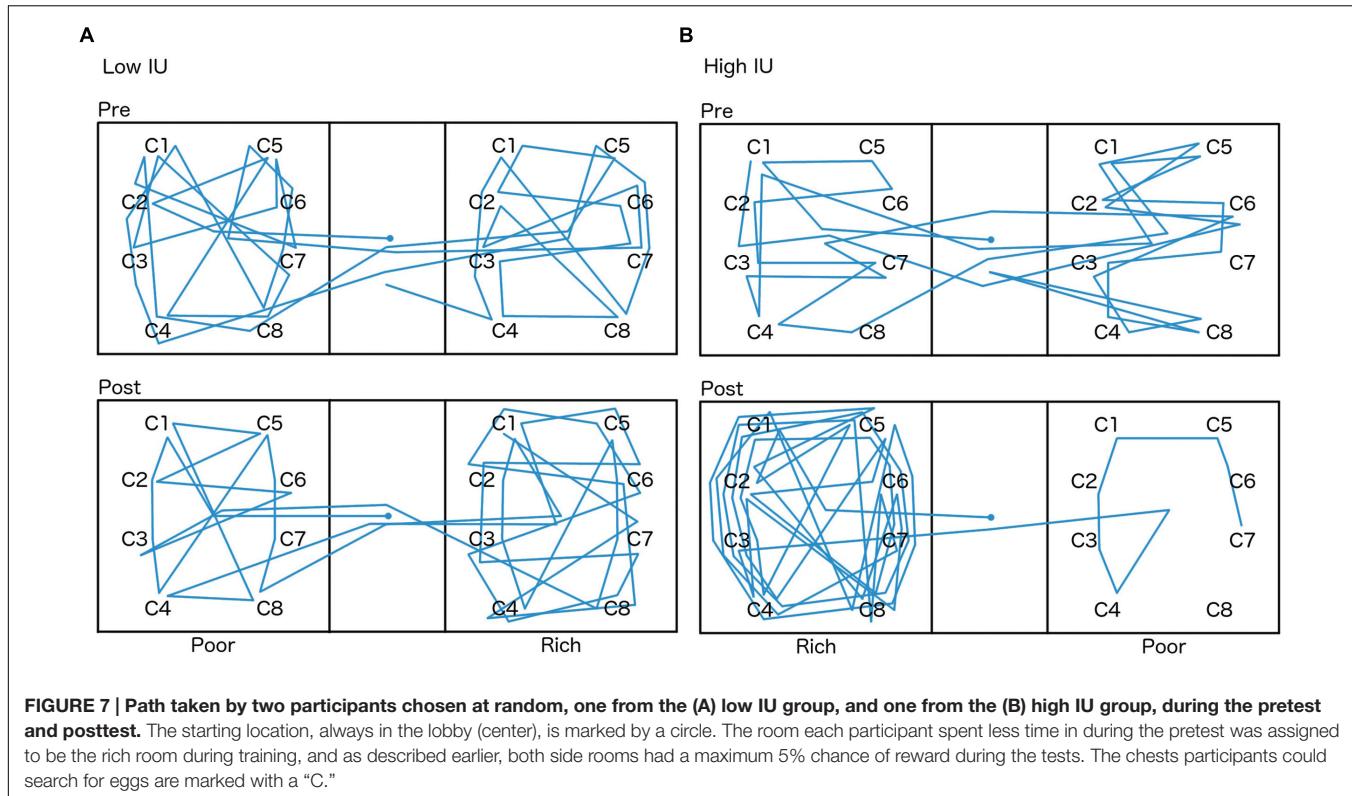
Finally, univariate ANOVA was performed on the total score (number of eggs collected) with between-subjects factors of



**FIGURE 5 | Movement between rooms in the computer-based task. (A)** Mean total side room (blue and brown room) entries during the posttest as a function of the first room entered and IU. There was a significant interaction; however, *post hoc* independent samples *t*-tests fell short of corrected significance. **(B)** Mean total entries into the previously rich and -poor rooms during the posttest as a function of the first room entered. Significantly more re-entries were made into the previously rich room when that room was also the one first entered during the posttest. In contrast, total entries into the previously poor room were similar irrespective of which room was entered first. Error bars represent  $\pm$  SEM. \* indicates significant difference,  $p < 0.01$ .



**FIGURE 6 | Movement within each side room during the posttest expressed as the mean total chest clicks in the (A) previously poor and (B) previously rich room, as a function of the first room entered and IU.** There were no significant differences. Error bars represent  $\pm$  SEM.



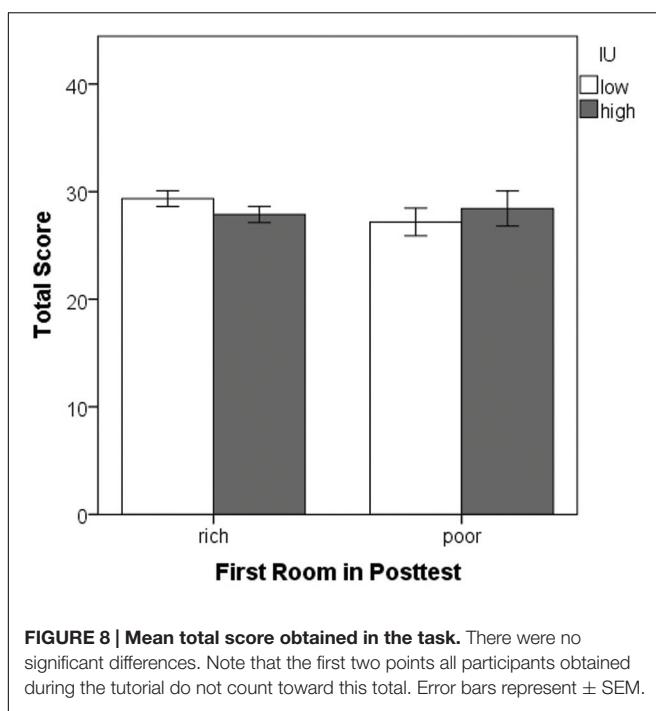
the first room entered during the posttest and IU. There were no significant differences (all  $p$ 's  $> 0.05$ ) indicating that on average, all participants obtained similar scores, irrespective of IU (Figure 8). Thus, differences in the amount of reward obtained during the training phase cannot account for the tendency of high IU individuals to choose to enter the previously rich room first during the posttest.

## DISCUSSION

The current study found that individuals with low IU showed little bias to enter the previously rich room first, and instead entered both rooms at about the same rate. In contrast, those with high IU had a strong bias to enter the previously rich room first (i.e., increased win-stay). This interaction appeared to be driven by participants who could identify the previously rich room. These findings could not be explained by differences in initial room preference, by prior video or computer game experience or in the total reward obtained by participants. It was also not related to whether or not participants reported following a specific strategy in the task. There are at least two possible interpretations of this result – first, individuals with high IU may have selected the safer, more certain choice, by returning to the previously rewarded context. Second, this could also indicate an increased tendency to chase reward, consistent with previously observed behavior in heroin addicts in a probabilistic category learning task (Myers et al., 2016). In either case, this tendency may represent a pre-existing cognitive bias, possibly based on

personality, which could promote decision-making processes that increase vulnerability to addiction. However, unlike the Myers et al. (2016) study, where the tendency to chase reward was expressed as exploration of new response options following expectancy violations (i.e., a lose-shift strategy), in the current task, participants chose the previously rewarded option (i.e., a win-stay strategy). This could be because, here, the response represents the first choice made in the posttest, precluding the influence of expectancy violations.

Another possible interpretation of this result, within an active inference framework (Friston et al., 2015), involves a change in the balance between pragmatic actions – actions that exploit previously rewarded strategies – and epistemic actions, which serve to discover new information that, long-term, may improve selection of pragmatic actions. According to this view, the value of an action is related to both its extrinsic value (i.e., expected reinforcement value) and its epistemic value (i.e., expected information gain). Here, actors seek out surprising outcomes, which will ultimately reduce uncertainty through information gain, and help construct a better internal model of the world. While short-term, this may require moving away from a goal (i.e., choosing to visit a previously less-rewarded location), the subsequent improved model would allow for better strategies to obtain a goal in the future (Friston et al., 2015). IU may involve a reduction in epistemic value in favor of increased extrinsic value, leading to behavior guided by preferences, i.e., prior beliefs about reinforcement contingencies. This interpretation is also consistent with the BI component of IU, which may paradoxically reduce information gain and therefore preclude



**FIGURE 8 | Mean total score obtained in the task.** There were no significant differences. Note that the first two points all participants obtained during the tutorial do not count toward this total. Error bars represent  $\pm$  SEM.

resolving uncertainty in the long-term (Ladouceur et al., 1997; Krain et al., 2006).

Overall, we did not observe a CPP effect. That is, despite the fact that a majority of participants entered the previously rich room first, when considering behavior across the entire posttest period, participants did not spend a majority of the time in the previously rich room. This behavioral pattern may reflect a different strategy than simple CPP. Some individuals may have been foraging or choosing where and how to seek reward, much like animals search for food in the natural environment. Decision making in a foraging scenario would involve deciding between a limited number of options which have different probabilities of reward and amount of reward (Platt and Huettel, 2008). Foraging also includes some risk or cost in choosing to look elsewhere for food. This cost may take the form of energy expenditure to travel to some other location where more food may be available. Another cost is the amount of time that it would take to arrive at the other location. Cognitive decisions to forage involve several factors including the value of each option, an estimate of the average value in the environment, as well as the cost of leaving the current location to search elsewhere (Kolling et al., 2012). In the current task, there was very little cost of time or energy in moving from room to room. Therefore, this scenario did not include risk of energy expenditure or much time lost. Thus, the freedom to switch rooms without penalty would have made foraging behavior a viable strategy to possibly achieve more reward.

A tendency from the foraging literature that may have been expressed by our high IU group is known as the ambiguity effect, where when given a choice between two options, one in which the probabilities are known and one in which the probabilities are unknown, most avoid the option with no probability information (Camerer and Weber, 1992). This avoidance of a choice with

unknown probability of outcomes could be an indication of IU. If high IU individuals knew which room was more rewarding, they may have tended to not shift their initial search to the other room, which in the past held less reward, but now may hold more. However, low IU persons exhibited a pattern of searching both rooms at similar rates. Low IU persons may have been more open to the risk of losing reward in the previously rich room if it was possible that more reward was available in the other room. Foraging has been tested in a computer environment (Goldstone and Ashpole, 2004), but the task involved a large number of participants interacting in real time in a virtual world. Based on a computational model, Goldstone and Ashpole (2004) suggested that some people tend to sample all locations with equal frequency while others tend to sample locations with greater rewards. Our current results would predict that these two tendencies may be found in two separate groups of individuals – those with lower IU would tend to sample all locations while those with higher IU would tend to sample locations of greater reward.

As noted earlier, individuals in the current study did not show an overall preference for the previously rich context. This could be because the current task differed in several ways from the CPP paradigms used in previous human and animal studies. In contrast to the work in animals, far fewer studies have attempted to examine CPP in humans. For example, in a study by Childs and de Wit (2009), humans received *d*-amphetamine or placebo in separate rooms. Participants reported higher liking for the drug-paired room. Molet et al. (2013) used a computer-based task where a distinct virtual environment was paired with either pleasant music or static noise. Analogous to animal studies, time spent in each context served as the dependent measure, and participants showed greater preference for the context paired with pleasant music. Finally, Astur et al. (2014) assessed preference for two distinct virtual rooms after one of them was paired with chocolate M&Ms. Similar to studies in animals, the participants spent more time in the chocolate-paired room, but only if they were food deprived.

Thus, similar to studies in animals, most human studies of CPP have employed either natural rewards (e.g., food, water) or drugs of abuse, while studies of economic decision making have used monetary gains. Nonetheless, most participants in the current study remained motivated throughout the task, despite only receiving golden eggs, as indicated by reliable movement and egg collection. The difference in the type of reinforcer, however, remains a possible explanation for the lack of overall preference in the current study, although note that Molet et al. (2013) were able to observe CPP to music. Similar to Childs and de Wit (2009), most participants in the current study reported that they knew which room was more rewarding, and were able to correctly identify that room. Despite this, there was no overall preference, and approximately 30% of participants first chose to enter the poor room during the posttest.

Another difference between this and other studies, which could account for the lack of overall preference, is in the duration of training. Animal studies typically involve multiple conditioning sessions, spread over several days, with training and testing on separate days. The study of Astur et al. (2014)

in humans employed six 6-min sessions, with a 5 min break between each session, with training and testing on separate days. On the other hand, as in the current study, Molet et al. (2013) employed only two 2-min conditioning sessions but still observed a preference. However, in contrast to the current study, Molet et al. (2013) used an unbiased procedure where each context is paired with a stimulus, in counterbalanced order – there was no pretest. Here, a biased procedure was used where the least preferred room during the pretest was paired with reward. Additionally, Molet et al. (2013) did not restrict the duration of the preference test, while here both tests were 4 min long. This was to ensure participants remained motivated and to reduce frustration given that the chance of reward was, at most, 5% during each test. It is possible that participants did not have enough time to explore each of the available locations and become familiar with the task. This is unlikely, given that on average, participants were able to switch between rooms several times despite the time limit (as shown by the average total room entries). Still, the duration of the test may have precluded observing an overall preference, which could be examined in future studies.

Finally, unlike prior studies of economic decision making, participants did not receive more or less reward depending on their decisions to stay in each location. This was by design since differences in the reinforcement value between locations during the test would confound interpretation of any preference observed (i.e., such preference could be due to experiences during training and testing). In the future, the task could be modified to examine the preference between a poor room with potentially higher gains, but lower gain on average, and a rich room with lower but reliable gains. Similarly, the task could be modified to examine the effect of IU on both reward and punishment learning by introducing a chance to lose points when foraging in particular locations (e.g., to assess preference for a location associated with high risk and high reward). These alternatives may have a strong effect on preference, and alter the foraging strategy used by participants as a function of IU.

In both human and animal studies of CPP, reward is not contingent on operant responding. In the context of substance abuse, however, humans choose to start taking the drug and control the frequency of administration. Similar choices are involved in the context of foraging and economic decision making. Animal studies have typically used an operant conditioning paradigm to study these processes, where subjects learn to press a lever to self-administer drugs or obtain other rewards (Balster and Lukas, 1985; Bardo and Bevins, 2000). Most standard self-administration studies, however, do not consider the role of contextual cues. Thus, the current task combined the two approaches in order to examine both contextual conditioning, and placed reward under the control of participants. In doing so, the task likely also taps into different mechanisms compared to traditional CPP paradigms. For example, behaviorally, the magnitude of rodent CPP is often dissociated from the rate of self-administration (Bardo et al., 1999). The two paradigms also appear to engage different neural substrates. For example, pretreatment with D2 dopamine receptor antagonists has no effect on CPP to cocaine (Cervo and

Samanin, 1995), but attenuates self-administration (Caine and Koob, 1994), suggesting that dopaminergic neurotransmission may only be involved in the primary reinforcing effects of cocaine, but not the secondary reinforcing properties acquired by contextual stimuli paired with cocaine (Bardo and Bevins, 2000). Finally, the ability of drugs of abuse to activate the mesolimbic dopamine system is also contingent on whether drug administration is under the operant control of the animal (Di Ciano et al., 1998).

Although the current task was probabilistic in that reward was not always guaranteed, the contrast between the two rooms during conditioning (5 vs. 80% chance of reward) should have been immediately apparent. When the rich room reverted to 5% chance of reward during the posttest, this may have led to rapid extinction, in particular since reward was under operant control, precluding observing a preference using time spent in the previously rich context as the dependent measure. While rodent CPP studies have used a 0 vs. 100% contrast, reward was not under operant control like it was in the current study. Regardless, the lack of an overall preference could also suggest that the effect of IU is not very strong in reality. Still, it may be possible to amplify this effect by increasing uncertainty (e.g., for example if the contrast is between 20 and 80% chance of reward). The number of chests in each context may have also played a role – the number was small enough to allow participants to explore all of the chests in one location before moving on and doing the same in the other. Thus, increasing the number of chests could influence how long participants choose to stay in one room, which could in turn impact overall preference.

In summary, we found a tendency for individuals who had high intolerance for uncertainty to first enter the previously rich reward room while individuals who had low intolerance showed no such bias, and first entered either of the rooms at equal rates. This initial decision may have been influenced by foraging strategies in addition to CPP. The results of the current study suggest that IU may have broader implications beyond the realm of anxiety, and is associated with changes in reward learning, even in healthy individuals. Studies are currently underway to examine the task in individuals undergoing treatment for opioid addiction. Given the relationship between IU and anxiety, such work should also compare addicts with and without comorbid anxiety disorders. It remains unclear if IU is an independent risk factor for both types of disorders, or if it is specific to individuals with comorbid anxiety that may have led to drug use in the first place, possibly as a form of self-medication (Khantzian, 1985). The current CPP task could also be adapted to examine a foraging scenario for further study of the effects of personality on economic decision making. This possible foraging task should include multiple rooms that the participant could explore for possible rewards. The cost for moving to other rooms could involve greater time delay that would reduce overall opportunity to forage. Thus, future computer-based behavioral tasks involving economic decision making could be used to test an individual's foraging behavior in the context of IU, as well as other personality factors, and could also be used to assess how personality affects the disorders such as substance abuse and anxiety.

## AUTHOR CONTRIBUTIONS

MR, CM, KB, AM, and MA were involved in study design. MA collected the data. MR, MA, and CM analyzed the data. MR, CM, KB, AM, and MA wrote the manuscript.

## FUNDING

Design and development of software was supported by the Clinical Science Research and Development Service of the VA Office of Research and Development (I01 CX000771). Funding for publication costs were supported by the University of

Northern Colorado. Opinions expressed herein are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs or the US Government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01175>

## REFERENCES

- Allen, M. T., Myers, C. E., and Servatius, R. J. (2014). Avoidance prone individuals self reporting behavioral inhibition exhibit facilitated acquisition and altered extinction of conditioned eyeblinks with partial reinforcement schedules. *Front. Behav. Neurosci.* 8:347. doi: 10.3389/fnbeh.2014.00347
- Allen, M. T., Myers, C. E., and Servatius, R. J. (2016). Uncertainty of trial timing enhances acquisition of conditioned eyeblinks in anxiety vulnerable individuals. *Behav. Brain Res.* 304, 86–91. doi: 10.1016/j.bbr.2016.02.007
- Astur, R. S., Carew, A. W., and Deaton, B. E. (2014). Conditioned place preferences in humans using virtual reality. *Behav. Brain Res.* 267, 173–177. doi: 10.1016/j.bbr.2014.03.018
- Balster, R. L., and Lukas, S. E. (1985). Review of self-administration. *Drug Alcohol Depend.* 14, 249–261. doi: 10.1016/0376-8716(85)90060-2
- Bardo, M. T., and Bevins, R. A. (2000). Conditioned place preference: what does it add to our preclinical understanding of drug reward? *Psychopharmacology* 153, 31–43. doi: 10.1007/s002130000569
- Bardo, M. T., Rowlett, J. K., and Harris, M. J. (1995). Conditioned place preference using opiate and stimulant drugs: a meta-analysis. *Neurosci. Biobehav. Rev.* 19, 39–51. doi: 10.1016/0149-7634(94)00021-R
- Bardo, M. T., Valone, J. M., and Bevins, R. A. (1999). Locomotion and conditioned place preference produced by acute intravenous amphetamine: role of dopamine receptors and individual differences in amphetamine self-administration. *Psychopharmacology* 143, 39–46. doi: 10.1007/s002130050917
- Birrell, J., Meares, K., Wilkinson, A., and Freeston, M. (2011). Toward a definition of intolerance of uncertainty: a review of factor analytical studies of the intolerance of uncertainty scale. *Clin. Psychol. Rev.* 31, 1198–1208. doi: 10.1016/j.cpr.2011.07.009
- Blaszczynski, A., and Nower, L. (2002). A pathways model of problem and pathological gambling. *Addiction* 97, 487–499. doi: 10.1046/j.1360-0443.2002.00015.x
- Caine, S. B., and Koob, G. F. (1994). Effects of dopamine D-1 and D-2 antagonists on cocaine self-administration under different schedules of reinforcement in the rat. *J. Pharmacol. Exp. Therap.* 270, 209–218.
- Camerer, C., and Weber, M. (1992). Recent developments in modeling preferences: uncertainty and ambiguity. *J. Risk Uncertain.* 5, 325–370. doi: 10.1007/BF00122575
- Carleton, R. N. (2012). The intolerance of uncertainty construct in the context of anxiety disorders: theoretical and practical perspectives. *Expert Rev. Neurother.* 12, 937–947. doi: 10.1586/ern.12.82
- Carleton, R. N., Mulvogue, M. K., Thibodeau, M. A., McCabe, R. E., Antony, M. M., and Asmundson, G. J. G. (2012). Increasingly certain about uncertainty: intolerance of uncertainty across anxiety and depression. *J. Anxiety Disord.* 26, 468–479. doi: 10.1016/j.janxdis.2012.01.011
- Carleton, R. N., Norton, P. J., and Asmundson, G. J. G. (2007). Fearing the unknown: a short version of the Intolerance of Uncertainty Scale. *J. Anxiety Disord.* 21, 105–117. doi: 10.1016/j.janxdis.2006.03.014
- Cervo, L., and Samanin, R. (1995). Effects of dopaminergic and glutamatergic receptor antagonists on the acquisition and expression of cocaine conditioning place preference. *Brain Res.* 673, 242–250. doi: 10.1016/0006-8993(94)01420-M
- Childs, E., and de Wit, H. (2009). Amphetamine-induced place preference in humans. *Biol. Psychiatry* 65, 900–904. doi: 10.1016/j.biopsych.2008.11.016
- Clark, L., and Robbins, T. W. (2002). Decision-making deficits in drug addiction. *Trends Cogn. Sci.* 6, 361–363. doi: 10.1016/S1364-6613(02)01960-5
- Clauss, J. A., Avery, S. N., and Blackford, J. U. (2015). The nature of individual differences in inhibited temperament and risk for psychiatric disease: a review and meta-analysis. *Prog. Neurobiol.* 127, 23–45. doi: 10.1016/j.pneurobio.2015.03.001
- Delgado, M. R., Frank, R. H., and Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat. Neurosci.* 8, 1611–1618. doi: 10.1038/nn1575
- Di Ciano, P., Blaha, C. D., and Phillips, A. G. (1998). Conditioned changes in dopamine oxidation currents in the nucleus accumbens of rats by stimuli paired with self-administration or yoked-administration of d-amphetamine. *Eur. J. Neurosci.* 10, 1121–1127. doi: 10.1046/j.1460-9568.1998.00125.x
- Dugas, M. J., Freeston, M. H., and Ladouceur, R. (1997). Intolerance of uncertainty and problem orientation in worry. *Cognit. Ther. Res.* 21, 593–606. doi: 10.1023/A:1021890322153
- Everitt, B. J., and Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* 8, 1481–1489. doi: 10.1038/nn1579
- Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G. (2015). Active inference and epistemic value. *Cogn. Neurosci.* 6, 187–214. doi: 10.1080/17588928.2015.1020053
- Goldstone, R. L., and Ashpole, B. C. (2004). Human foraging behavior in a virtual environment. *Psychon. Bull. Rev.* 11, 508–514. doi: 10.3758/BF03196603
- Grant, B. F., Stinson, F. S., Dawson, D. A., Chou, S. P., Dufour, M. C., Compton, W., et al. (2004). Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: results from the national epidemiologic survey on alcohol and related conditions. *Arch. Gen. Psychiatry* 61, 807–816. doi: 10.1001/archpsyc.61.8.807
- Green, L., Myerson, J., and McFadden, E. (1997). Rate of temporal discounting decreases with amount of reward. *Mem. Cogn.* 25, 715–723. doi: 10.3758/BF03211314
- Grupe, D. W., and Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nat. Rev. Neurosci.* 14, 488–501. doi: 10.1038/nrn3524
- Hirshfeld, D. R., Rosenbaum, J. F., Biederman, J., Bolduc, E. A., Faraone, S. V., Snidman, N., et al. (1992). Stable behavioral inhibition and its association with anxiety disorder. *J. Am. Acad. Child Adolesc. Psychiatry* 31, 103–111. doi: 10.1097/00004583-199201000-00016
- Holloway, J., Allen, M. T., Myers, C. E., and Servatius, R. J. (2014). Behaviorally inhibited individuals demonstrate significantly enhanced conditioned response acquisition under non-optimal learning conditions. *Behav. Brain Res.* 261, 49–55. doi: 10.1016/j.bbr.2013.10.041
- Jacobsen, L. K., Southwick, S. M., and Kosten, T. R. (2001). Substance use disorders in patients with posttraumatic stress disorder: a review of the literature. *Am. J. Psychiatry* 158, 1184–1190. doi: 10.1176/appi.ajp.158.8.1184

- Kagan, J., Reznick, J. S., and Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child Dev.* 58, 1459–1473. doi: 10.2307/1130685
- Khantzian, E. J. (1985). The self-medication hypothesis of addictive disorders: focus on heroin and cocaine dependence. *Am. J. Psychiatry* 142, 1259–1264. doi: 10.1176/ajp.142.11.1259
- Kolling, N., Behrens, T. E., Mars, R. B., and Rushworth, M. F. (2012). Neural mechanisms of foraging. *Science* 336, 95–98. doi: 10.1126/science.1216930
- Krain, A. L., Hefton, S., Pine, D. S., Ernst, M., Xavier Castellanos, F., Klein, R. G., et al. (2006). An fMRI examination of developmental differences in the neural correlates of uncertainty and decision-making. *J. Child Psychol. Psychiatry* 47, 1023–1030. doi: 10.1111/j.1469-7610.2006.01677.x
- Ladouceur, R., Talbot, F., and Dugas, M. J. (1997). Behavioral expressions of intolerance of uncertainty in worry: experimental findings. *Behav. Modif.* 21, 355–371. doi: 10.1177/01454455970213006
- Lett, B. T., Grant, V. L., Byrne, M. J., and Koh, M. T. (2000). Pairings of a distinctive chamber with the aftereffect of wheel running produce conditioned place preference. *Appetite* 34, 87–94. doi: 10.1006/appet.1999.0274
- Li, X., Zhang, F., Zhou, Y., Zhang, M., Wang, X., and Shen, M. (2013). Decision-making deficits are still present in heroin abusers after short-to long-term abstinence. *Drug Alcohol Depend.* 130, 61–67. doi: 10.1016/j.drugalcdep.2012.10.012
- Lorains, F. K., Cowlishaw, S., and Thomas, S. A. (2011). Prevalence of comorbid disorders in problem and pathological gambling: systematic review and meta-analysis of population surveys. *Addiction* 106, 490–498. doi: 10.1111/j.1360-0443.2010.03300.x
- Luhmann, C. C., Ishida, K., and Hajcak, G. (2011). Intolerance of uncertainty and decisions about delayed, probabilistic rewards. *Behav. Ther.* 42, 378–386. doi: 10.1016/j.beth.2010.09.002
- Madden, G. J., Begotka, A. M., Raiff, B. R., and Kasten, L. L. (2003). Delay discounting of real and hypothetical rewards. *Exp. Clin. Psychopharmacol.* 11, 139–145. doi: 10.1037/1064-1297.11.2.139
- Maner, J. K., and Schmidt, N. B. (2006). The role of risk avoidance in anxiety. *Behav. Ther.* 37, 181–189. doi: 10.1016/j.beth.2005.11.003
- Merikangas, K. R., Mehta, R. L., Molnar, B. E., Walters, E. E., Swendsen, J. D., Aguilar-Gaziola, S., et al. (1998). Comorbidity of substance use disorders with mood and anxiety disorders: results of the International Consortium in Psychiatric Epidemiology. *Addict. Behav.* 23, 893–907. doi: 10.1016/S0306-4603(98)00076-8
- Mischel, W., and Metzner, R. (1962). Preference for delayed reward as a function of age, intelligence, and length of delay interval. *J. Abnorm. Soc. Psychol.* 64, 425–431. doi: 10.1037/h0045046
- Molet, M., Billiet, G., and Bardo, M. T. (2013). Conditioned place preference and aversion for music in a virtual reality environment. *Behav. Processes* 92, 31–35. doi: 10.1016/j.beproc.2012.10.001
- Morgan, B. E. (2006). Behavioral inhibition: a neurobiological perspective. *Curr. Psychiatry Rep.* 8, 270–278. doi: 10.1007/s11920-006-0062-7
- Myers, C. E., Moustafa, A. A., Sheynin, J., VanMeenen, K. M., Gilbertson, M. W., Orr, S. P., et al. (2013). Learning to obtain reward, but not avoid punishment, is affected by presence of PTSD symptoms in male veterans: empirical data and computational model. *PLoS ONE* 8:e72508. doi: 10.1371/journal.pone.0072508
- Myers, C. E., Sheynin, J., Balsdon, T., Luzardo, A., Beck, K. D., Hogarth, L., et al. (2016). Probabilistic reward-and punishment-based learning in opioid addiction: experimental and computational data. *Behav. Brain Res.* 296, 240–248. doi: 10.1016/j.bbr.2015.09.018
- Myers, C. E., VanMeenen, K. M., and Servatius, R. J. (2012). Behavioral inhibition and PTSD symptoms in veterans. *Psychiatry Res.* 196, 271–276. doi: 10.1016/j.psychres.2011.11.015
- Nelson, B. D., Kessel, E. M., Jackson, F., and Hajcak, G. (2015). The impact of an unpredictable context and intolerance of uncertainty on the electrocortical response to monetary gains and losses. *Cogn. Affect. Behav. Neurosci.* 16, 153–163. doi: 10.3758/s13415-015-0382-3
- Oldenburger, W. P., Everitt, B. J., and De Jonge, F. H. (1992). Conditioned place preference induced by sexual interaction in female rats. *Horm. Behav.* 26, 214–228. doi: 10.1016/0018-506X(92)90043-U
- Packard, M. G. (2009). Anxiety, cognition, and habit: a multiple memory systems perspective. *Brain Res.* 1293, 121–128. doi: 10.1016/j.brainres.2009.03.029
- Packard, M. G., and Cahill, L. (2001). Affective modulation of multiple memory systems. *Curr. Opin. Neurobiol.* 11, 752–756. doi: 10.1016/S0959-4388(01)00280-X
- Petry, N. M., Bickel, W. K., and Arnett, M. (1998). Shortened time horizons and insensitivity to future consequences in heroin addicts. *Addiction* 93, 729–738. doi: 10.1046/j.1360-0443.1998.9357298.x
- Platt, M. L., and Huettel, S. A. (2008). Risky business: the neuroeconomics of decision making under uncertainty. *Nat. Neurosci.* 11, 398–403. doi: 10.1038/nn2062
- Radell, M. L., Sanchez, R., Weinfash, N., and Myers, C. E. (2016). The personality trait of behavioral inhibition modulates perceptions of moral character and performance during the trust game: behavioral results and computational modeling. *PeerJ* 4:e1631. doi: 10.7717/peerj.1631
- Robinson, T. E., and Berridge, K. C. (2001). Incentive-sensitization and addiction. *Addiction* 96, 103–114. doi: 10.1046/j.1360-0443.2001.961038.x
- Schwartz, C. E., Wright, C. I., Shin, L. M., Kagan, J., and Rauch, S. L. (2003). Inhibited and uninhibited infants “grown up”: adult amygdalar response to novelty. *Science* 300, 1952–1953. doi: 10.1126/science.10183703
- Sheynin, J., Beck, K. D., Pang, K. C., Servatius, R. J., Shikari, S., Ostovich, J., et al. (2014). Behaviourally inhibited temperament and female sex, two vulnerability factors for anxiety disorders, facilitate conditioned avoidance (also) in humans. *Behav. Processes* 103, 228–235. doi: 10.1016/j.beproc.2014.01.003
- Sheynin, J., Moustafa, A. A., Beck, K. D., Servatius, R. J., Casbolt, P. A., Haber, P., et al. (2016). Exaggerated acquisition and resistance to extinction of avoidance behavior in treated heroin-dependent men. *J. Clin. Psychiatry* 77, 386–394. doi: 10.4088/JCP.14m09284
- Sheynin, J., Shikari, S., Gluck, M. A., Moustafa, A. A., Servatius, R. J., and Myers, C. E. (2013). Enhanced avoidance learning in behaviorally inhibited young men and women. *Stress* 16, 289–299. doi: 10.3109/10253890.2012.744391
- Sinha, R. (2001). How does stress increase risk of drug abuse and relapse? *Psychopharmacology* 158, 343–359. doi: 10.1007/s002130100917
- Sprakler, C., Fibiger, H. C., and Phillips, A. G. (1982). Attenuation by haloperidol of place preference conditioning using food reinforcement. *Psychopharmacology* 77, 379–382. doi: 10.1007/BF00432775
- Tolin, D. F., Abramowitz, J. S., Brigidi, B. D., and Foa, E. B. (2003). Intolerance of uncertainty in obsessive-compulsive disorder. *J. Anxiety Dis.* 17, 233–242. doi: 10.1016/S0887-6185(02)00182-2
- Tzschentke, T. M. (2007). Review on CPP: measuring reward with the conditioned place preference (CPP) paradigm: update of the last decade. *Addict. Biol.* 12, 227–462. doi: 10.1111/j.1369-1600.2007.00070.x
- Volkow, N. D., Wang, G. J., Fowler, J. S., Tomasi, D., Telang, F., and Baler, R. (2010). Addiction: decreased reward sensitivity and increased expectation sensitivity conspire to overwhelm the brain’s control circuit. *Bioessays* 32, 748–755. doi: 10.1002/bies.201000042
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- The contents do not necessarily represent the official views of the Department of Veterans Affairs, the United States Government, or any institution with which the authors are affiliated.
- Copyright © 2016 Radell, Myers, Beck, Moustafa and Allen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Prudence, Emotional State, Personality, and Cognitive Ability

Adriana Breaban<sup>1\*</sup>, Gijs van de Kuilen<sup>1</sup> and Charles N. Noussair<sup>2</sup>

<sup>1</sup> Department of Economics, Tilburg University, Tilburg, Netherlands, <sup>2</sup> Department of Economics, University of Arizona, Tucson, AZ, USA

We report an experiment to consider the emotional correlates of prudent decision making. In the experiment, we present subjects with lotteries and measure their emotional response with facial recognition software. They then make binary choices between risky lotteries that distinguish prudent from imprudent individuals. They also perform tasks to measure their cognitive ability and a number of personality characteristics. We find that a more negative emotional state correlates with greater prudence. Higher cognitive ability and less conscientiousness is also associated with greater prudence.

**Keywords:** emotions, prudence, personality, cognitive ability

## OPEN ACCESS

**Edited by:**

Nikolaos Georgantzis,  
University of Reading, UK

**Reviewed by:**

Michalis Drouvelis,  
University of Birmingham, UK  
Sascha Behnk,  
University of Zurich, Switzerland

**\*Correspondence:**

Adriana Breaban  
a.breaban@uvt.nl

**Specialty section:**

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

**Received:** 29 June 2016

**Accepted:** 13 October 2016

**Published:** 28 October 2016

**Citation:**

Breaban A, van de Kuilen G and Noussair CN (2016) Prudence, Emotional State, Personality, and Cognitive Ability. *Front. Psychol.* 7:1688.  
doi: 10.3389/fpsyg.2016.01688

## INTRODUCTION

The study of the role of risk preferences in decision making has primarily focused on the implications of risk aversion, i.e., the preference for a certain payment to a lottery with the same expected value. If one assumes that individuals maximize expected utility (e.g., for prescriptive applications), risk aversion implies that the utility function for money is concave (i.e., that  $u''(x) < 0$ ). However, empirical work has shown that the degree of risk aversion is often affected by psychological factors not captured by the expected utility model, such as the perceived likelihood of events and the perceived domain of the outcomes (e.g., Tversky and Kahneman, 1992). Moreover, theoretical work has shown that risk aversion is not the only facet of preference governing economic decision making: it is becoming increasingly recognized that the higher order risk attitudes of *prudence* and *temperance* complement the role of risk aversion in economic decision making in important ways. For example, in the realm of saving behavior, while risk aversion drives the preference to smooth consumption over time (*consumption smoothing*; Friedman, 1957), prudence determines how saving behavior changes as future income becomes riskier (*precautionary saving*; Kimball, 1990). Other examples of areas of economics, in which higher order risk preferences have been found to play an important role in influencing behavior, include bidding in auctions (Esö and White, 2004), bargaining (White, 2008), tax compliance (Alm, 1988), and rent seeking (Treich, 2010).

Within the expected utility framework, prudence is typically defined as the convexity of marginal utility ( $u'''(x) > 0$ ), while temperance is equivalent to a negative fourth derivative of the utility function ( $u''''(x) < 0$ ). However, Eeckhoudt and Schlesinger (2006) have introduced behavioral definitions, based on observable revealed preferences, of prudence and temperance that are model-free in the sense that they retain validity if expected utility fails to accurately describe choice behavior (e.g., see Starmer, 2000). The definitions of Eeckhoudt and Schlesinger (2006) are based on risk apportionment. In particular, a decision maker (DM) is prudent if she prefers to apportion an unavoidable zero-mean risk to a relatively high rather than to a low wealth state, while a temperate DM prefers to apportion two independent zero-mean risks across different states of nature.

Several recent papers have used the behavioral definitions of Eeckhoudt and Schlesinger (2006) to quantify higher-order risk preferences empirically. The results from these studies show that the degree of prudence varies considerably among individuals within the population (Deck and Schlesinger, 2010, 2014; Ebert and Wiesen, 2011, 2014; Noussair et al., 2014), though all of these studies agree that a majority of individuals are prudent. Furthermore, Noussair et al. (2014), who study a large sample of demographically representative individuals, find that those who exhibit more prudent decision making also have greater savings, lower debt, more wealth and higher educational attainment. The results for the prevalence of temperance within the population are more mixed (e.g., Deck and Schlesinger, 2010, 2014; Noussair et al., 2014).

It is also widely recognized in behavioral economics, psychology, and management, that there is an important connection between emotional state and risk preferences. However, research in this area has focused exclusively on the link between emotional state and risk aversion. This research can be classified based on whether it considers the relationship between risk taking and overall valence (positivity or negativity of emotional state), or to specific emotions such as fear, anger, and happiness, as correlates of decision making. Johnson and Tversky (1983) propose that a positively-valenced emotional state increases risk taking, because it makes beliefs about outcomes more optimistic. This relationship is termed the *Affective Generalization Hypothesis*. On the other hand, Isen et al. (1988) have argued that a positive mood leads to less risk taking because individuals wish to preserve the positive emotional state and insulate themselves from negative outcomes. This is referred to as the *Mood Maintenance Hypothesis*.

In addition to overall valence, specific emotions have been associated with risk taking. The Appraisal Tendency Framework (Lerner and Tiedens, 2006) predicts that the emotion of fear is associated with greater risk aversion, while anger and happiness are correlated with greater risk taking. These propositions are supported by experimental studies (Lerner and Keltner, 2001; Kugler et al., 2012), in which emotions are induced prior to a risky choice task. Recent work by Nguyen and Noussair (2014), in which emotions are observed and tracked rather than induced, reports that fear, happiness, and anger all correlate positively with risk aversion, while emotional valence correlates negatively with risk aversion (negative emotions are associated with risk aversion).

Theoretical work, shows that those who are imprudent save less when their background risk increases (Kimball, 1990), behavior which may be financially hazardous for them as well as socially undesirable. Moreover, previous work has shown that imprudence correlates with poor decision-making (Noussair et al., 2014). In short, imprudent people get into financial trouble. It is, therefore, interesting and valuable to know what correlates with imprudent decision making. One factor that might get in the way of making good decisions are strong emotions. In this study, we consider which emotional states correlate with imprudent financial decisions. While research on the connection between emotions and risk aversion has established clear and important relationships, nothing is known

about the correlation between emotional state and higher order risk attitudes. In this paper, we consider the relationship between prudent decision making and emotional state. Our design is guided by the theoretical work of Eeckhoudt and Schlesinger (2006) and the experimental implementation of Deck and Schlesinger (2010, 2014). Eeckhoudt and Schlesinger (2006) show how prudent and imprudent decisions can be distinguished using risk apportionment tasks that are simple to understand and straightforward to implement in the laboratory. Just as the willingness to accept a zero-mean risk can distinguish a risk averse from a risk seeking individual, a preference for accepting an unavoidable zero-mean risk in a relatively high, rather than a low, income state can reveal prudence. Even though this behavioral definition of prudence is model-free (just like the definition of risk aversion as a preference for the expected value of a lottery over the lottery itself is), a preference for assigning unavoidable risk to relatively high income states implies convex marginal utility or  $u''(x) > 0$ , if one assumes that the DM maximizes expected utility (Eeckhoudt and Schlesinger, 2006).

We design and report an experiment that consists of two phases. In the first phase, participants are presented with a series of ten lotteries, in which two different payoff levels are equally likely. Each lottery is resolved after it is displayed. In the second phase of a session, subjects make choices between lotteries. The decisions have the feature that they offer a choice between two lotteries that are equivalent in terms of mean and variance, but that differ in skewness by varying whether they apportion risk to a high or low income state. We consider whether the emotional response to the presentation of the lotteries in the initial phase correlates with subsequent decisions. Additionally, we investigate correlations between some characteristics of individuals and their level of prudence. We measure our participants' cognitive ability using Raven's test of progressive matrices (Bors and Stokes, 1998) and personality traits as captured by the Big Five inventory (Gosling et al., 2003), and relate these to the decisions they make.

Our experiment shows that decisions depend on emotional state. The emotional state of participants in phase 1 of the experiment correlates with the level of prudence in their phase 2 decisions. More positive valence correlates with less prudent choices. Changes in arousal during the display of the prospects in the first phase of the experiment does correlate with decisions, with greater increases in arousal associated with more prudent choices. Our results as a whole indicate that stronger emotions tend to be associated with greater prudence, though all else equal, more positive emotional state correlates with less prudence. This pattern of results is similar to those observed by Nguyen and Noussair (2014) for risk aversion. They found that stronger emotions were correlated with more risk averse choices, and positive valence with less risk averse choices. We also observe that greater cognitive ability, as measured by the Raven's test score, is associated with greater prudence. This last result is in line with those reported by Noussair et al. (2014), using a different measure of cognitive ability, the Cognitive Reflection Test (Frederick, 2005). We also observe that conscientiousness correlates negatively with prudence.

## MATERIALS AND METHODS

### The Participants and the Setting

Eighty-three students from Tilburg University in the Netherlands participated in this computerized experiment, which was conducted at the CentER laboratory at Tilburg University in 2016<sup>1</sup>. There were six experimental sessions, each involving between 7 and 19 subjects. The majority of subjects studied economics. The average age was 22.5 years and 50.6% of the subjects were female.

The subjects were recruited among a pool of volunteers and were told that the experiment would last for up to 1 h. The experiment was programmed in Ztree (Fischbacher, 2007). The experiment consisted of four phases. At the start of each phase 1 to 3, separate instructions were read aloud. Instructions can be found online in the Data Sheet 1. During the experiment, facial expressions were recorded continuously by using video cameras. After completing the experiment, subjects were paid in private.

### Procedures and Data Gathered

In the first phase of the experiment, subjects were presented with 10 risky lotteries, displayed sequentially. Each lottery involved a 50/50 chance of receiving either a low or a high outcome with outcomes ranging from €1 to €13, and expected values ranging from €3.5 to €8.5. The lotteries displayed in phase 1 were unrelated to the lotteries that were presented later in the experiment.

After being presented on the screen, the lottery was resolved for each individual and the outcome of the lottery was then displayed on the screen for 10 s.<sup>2</sup> Then, the next lottery appeared on the screen. The purpose of the first phase was to observe the emotional reaction caused by merely being exposed to risk and the emotional reaction caused by experiencing the outcome of the risky option. We register the emotion data at the time of presentation of the lottery itself, which we refer to as the *exposure* emotions. We also measure emotional state at the time each lottery is resolved and we refer to these as *feedback* emotions. In addition, we also retain for analysis the emotional state before the beginning of the experiment, and designate these as *initial* emotions.

The emotions are measured in the following manner. We videotape participants for the entire session with their consent. The videotapes are then analyzed with Noldus FaceReader™ software, which tracks facial expressions and analyzes the emotions they display. FaceReader has been employed in a

<sup>1</sup>Tilburg University, where the experiment was conducted, does not have an Institutional Review Board. This is fully in line with Dutch law, which does not require IRB review for social science research. Subjects gave verbal consent to be videotaped. However, they were unaware that their facial expressions would be analyzed.

<sup>2</sup>When single emotions occur and there is no reason for them to be modified or concealed, expressions typically last between 0.5 to 4 seconds and involve the entire face (Ekman, 2003). The onset and offset of a sincere emotional response in reaction to a stimulus is generally between 2/3 of a second and 4 seconds (Hager and Ekman, 1985; Hess and Kleck, 1990). Thus, the 10 second window that we study should capture the full reaction to exposure to the lottery or to feedback from the lottery outcome. The relatively long time horizon in which we measure emotional state at the beginning of the experiment, allows us a relatively large amount of data on subjects' initial mood at the outset of the session.

number of experimental economics studies focusing on emotions (e.g., Breaban and Noussair, 2014; Nguyen and Noussair, 2014; Van Leeuwen et al., 2014; Habetinova and Noussair, 2015), but has also been used in marketing (Teixeira et al., 2012; Lewinski et al., 2014), and in psychological (Chentsova-Dutton and Tsai, 2010), research.

The FaceReader software tracks facial movements using the Facial Action Coding System, which associates specific muscle movements to the six basic universal emotions cataloged by Paul Ekman and his colleagues (e.g., Ekman et al., 1987; Ekman and Friesen, 2003). The emotions are happiness, fear, anger, disgust, surprise, and sadness. Facereader also measures how closely a facial expression conforms to a neutral state and generates an overall measure of emotional valence, as well as of arousal. The valence measure is calculated as Happiness—max{Anger, Fear, Sadness, Disgust}, that is, the value of the only positive emotion, happiness, minus the strongest of the four negative emotions. Arousal is a measure of emotional activation that varies from 0 to 1 and it is calculated as the average of the current highest five activation indicators corrected by a continuous average of activation during the last 60 s. The specific emotions are computed on a scale from 0 to 1, with one indicating complete conformity of facial movements to those associated with an emotion. It registers emotional state 30 times per second.

To compute the initial value of an emotion, we average the registered value of the emotion over the 60 s before phase 1 of the experiment began. During this period, subjects had no task to perform, and were passively waiting for the experiment to start. Exposure emotions represent the average over the 10 s during which a lottery is presented, and feedback emotions are computed as the average over the 10 s immediately following the resolution of the lottery.

The second phase of the experiment involves 10 direct pairwise choices. Each consists of a choice between one lottery that would be preferred by a prudent individual and an alternative that would be preferred by a decision maker who is imprudent. An example of a choice as presented to participants can be found in **Figure 1**. In both phases, all subjects were presented with all lotteries in the same order.

In the example of a choice shown in the figure, with 50% probability Left yields €10 and an additional 50/50 lottery yielding either a further gain or loss of €4. Otherwise, Left yields €4. Similarly, Right yields either €10 or €4 and an additional 50/50 lottery yielding either a gain of €4 or a loss of €4, both with 50% probability. Thus, the choice between left and right amounts to whether the subject prefers to apportion a zero-mean €4 risk to a state with relatively high wealth (left), or to a state with relatively low wealth (right). A choice for left (right) indicates that the decision maker can better cope with the zero-mean €4 risk when she has relatively more (less) wealth, implying that she is prudent (imprudent). The precise lotteries that were used are given in **Table 1**. In line with the existing literature (Deck and Schlesinger, 2010, 2014; Noussair et al., 2014), we use the number of prudent choices that a subject makes as a measure of the individual strength of prudence. If an individual chooses the prudent option in 6 or more of the 10 decisions she takes, we classify the individual as prudent. Analogously, if she chooses the

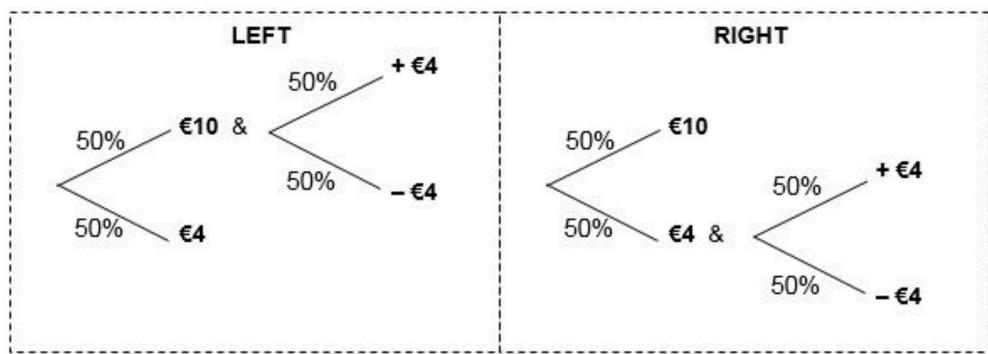


FIGURE 1 | Example of a choice from phase 2 of the experiment.

TABLE 1 | Prudent lotteries used and choice proportions.

Choice #	Lottery displayed on left	Lottery displayed on right	% of instances in which prudent choice was made
1	(10+(4_4)_4)	(10_4+(4_4))	88.0***
2	(6+(1_1)_1)	(6_1+(1_1))	79.5***
3	(12+(2_2)_3)	(12_3+(2_2))	79.5***
4	(9+(2_2)_3)	(9_3+(2_2))	74.7***
5	(8+(4_4)_4)	(8_4+(4_4))	83.1***
6	(6+(1_1)_3)	(6_3+(1_1))	73.5***
7	(7+(2_2)_2)	(7_2+(2_2))	85.5***
8	(11+(3_3)_3)	(11_3+(3_3))	88.0***
9	(13+(4_4)_4)	(13_4+(4_4))	85.5***
10	(12+(2_2)_2)	(12_2+(2_2))	86.7***

(x\_y) indicates a lottery with an equal probability of receiving either x or y; outcomes in euros; \*\*\* indicates significant difference at 1% level from random choice between left and right option, binomial test, two-sided.

prudent option in 5 or fewer instances, the individual is said to be imprudent.

In the third phase of the experiment, cognitive ability is measured using Raven's advanced progressive matrices test (Raven et al., 1998), a protocol commonly used to measure fluid intelligence. The task involves choosing the correct one out of eight possible alternatives to complete a 3-by-3 matrix of abstract symbols in a consistent pattern. Due to the limited amount of time available in our sessions, we used the short form of the test proposed by Bors and Stokes (1998) that consists of 12 tasks. Subjects were given a total of 10 min to complete the 12 tasks, and were allowed to revise previous answers if time allowed.

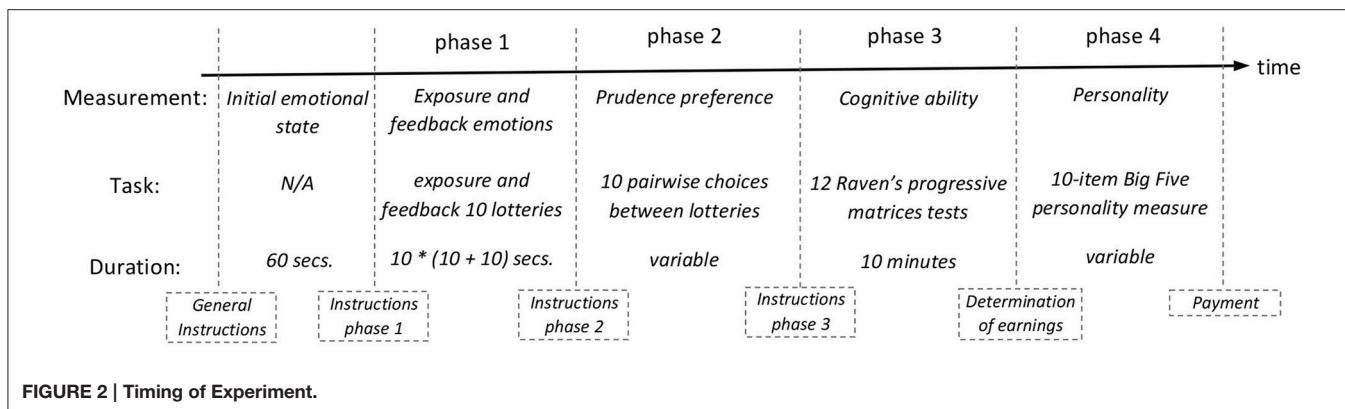
The final phase of the experiment consists of a questionnaire designed to obtain a classification of personality. More specifically, we administer the 10-item Big Five personality measure developed by Gosling et al. (2003). This measure allows one to classify individual differences in personality into five broad dimensions: extraversion, agreeableness, conscientiousness, neuroticism, and openness to new experiences, by registering applicability of 10 items regarding subject's personality on a scale from 1 (disagree strongly) to 7 (agree strongly). In addition, background information of subjects regarding age, gender, study,

year of study was gathered. There is some previous evidence that the dimensions of openness and extraversion correlate negatively with risk aversion, and neuroticism, agreeableness and conscientiousness correlating positively (Nicholson et al., 2005; Becker et al., 2012). We are unaware of any prior work correlating personality characteristics and prudence.

Thus, for each participant, we observe the emotional reaction caused by being exposed to risk and the emotional reaction caused by experiencing the outcome of a risky lottery (phase 1), as well as a measure of the degree of prudence (phase 2), of cognitive ability (phase 3), and of personality dimensions (phase 4). **Figure 2** below shows a timeline of the experiment.

To avoid potential income effects on the measure of prudence [such as Thaler and Johnson's (1990) house money effect] and to provide incentives for truthfully reporting preferences, the random incentive mechanism was used. That is, subjects were informed from the outset that at the end of the experiment, phase 1 or phase 2 would be randomly selected with equal probability. If the first phase is selected, the observed outcome of one of the ten of the lotteries (randomly selected) count toward the participant's earnings. If the second phase is selected, the computer randomly selects one of the ten pairs of lotteries. The outcome of the chosen lottery in that pair would then count toward earnings. On top of these earnings, subjects received €0.50 for each of the correct answers to the Raven test in phase 3 as well as a fixed participation fee of €2. On average, subjects earned €12.18 during the experiment.

One of our design choices merits some further comment. We have chosen to track, without attempting to influence, the emotions and arousal level that our participants exhibit during our task. An alternative would be to induce different emotional or arousal states and compare the resulting decisions, as many other authors have done. The induction of emotions is well suited to addressing questions regarding the causal effects of emotional variables on decision making, and is a powerful tool for addressing many if not most important questions in emotion research. The design we have chosen is meant to document correlates of prudent decision making, rather than causal relationships. We consider whether those who tend to exhibit particular emotions, greater or less arousal, and positive or negative emotional state after exposure to and experience with



**FIGURE 2 |** Timing of Experiment.

lotteries, exhibit more or less prudence in subsequent decisions. Identifying such correlates of prudence in decision making is the purpose of this research.

## RESULTS

A clear majority of individuals in the study were prudent. 42.17% (35 of 83) of participants made a prudent decision at every opportunity. Another 46.99% (39 of 83) made a prudent choice between 6 and 9 times, indicating that they chose prudently in a majority of instances in which they had an opportunity to do so. Thus, 89.16% of individuals are classified as prudent. 10.84% (9 of 83) of participants made fewer than 6 prudent choices are thus classified as imprudent. The fact that a majority of participants is prudent is consistent with the previous literature (Deck and Schlesinger, 2010, 2014; Ebert and Wiesen, 2011, 2014; Noussair et al., 2014).

**Figure 3** illustrates the average emotional state in phase 1 of the experiment for those who made 0–5, between 6 and 9, and who made 10 prudent decisions in phase 2. The panels on the left indicate the average value of the exposure emotions, measured at the time that the lotteries are displayed in phase 1. Those on the right are the feedback emotions, those registered at the time that each of the phase 1 lotteries is resolved. The strength of the various emotions is typically similar at the exposure as at the feedback point. The figure shows that those who exhibit more negative valence, as well as stronger anger, surprise and disgust, and lower happiness, when viewing the lotteries, make more prudent decisions. The results are similar whether exposure or feedback emotions are considered.

To make these impressions more precise and to control for other potential influences on prudence, we conduct Poisson count regressions in which the number of prudent choices is the dependent variable. The estimates for feedback emotions are reported in **Table 2**, and those for exposure emotions are in **Table 3**.<sup>3</sup>

<sup>3</sup>Subjects were told to pay attention to their screen and were asked not to touch their face during the experiment. This ensured that we were able to gather facial expression data for the vast majority of decisions. There are 60 to 69 for missing observations for the results in **Table 2** and 110 to 116 missing observations for the results reported in **Table 3**. These missing observations are instances when subjects looked away from their computer screens or covered part of their faces with their hands.

In results 1–4, we report our results concerning the correlates of prudence. The first result below indicates that there is a negative correlation between the overall valence of emotional state and prudence. Those in a more positive emotional state are less prudent.

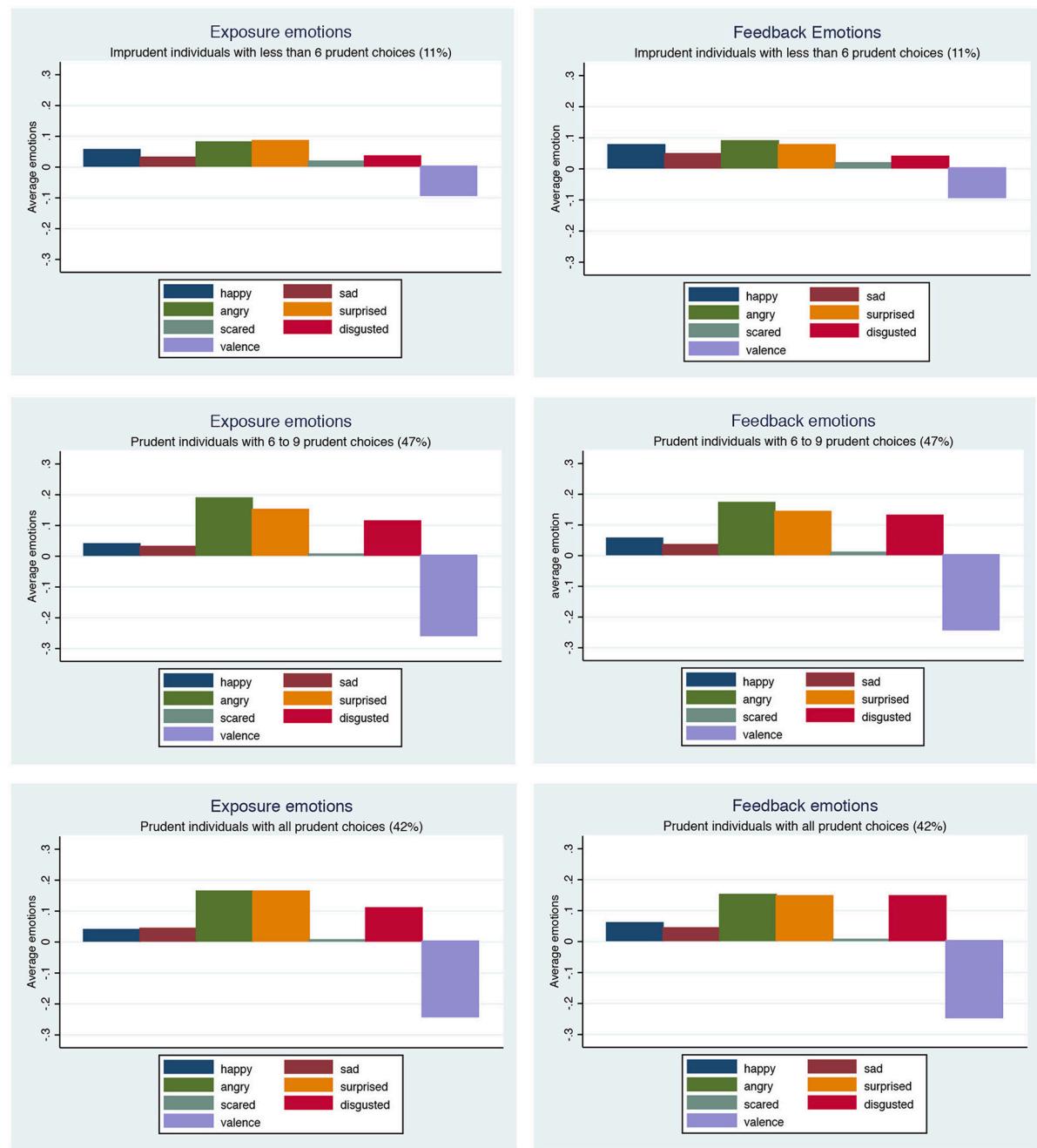
### Result 1: Positivity of Emotional State, When Facing Risky Lotteries, Correlates with Imprudence

#### Support for Result 1

**Table 2** contains estimates of Poisson count regressions in which the number of prudent choices is the dependent variable. The valence variable is evaluated at the feedback stage. The coefficients of valence in specifications (1), (2), (4), and (5) indicate that valence is a significant predictor of decisions. In all four regressions, the coefficient of valence is negative and significant at the  $p < 0.05$  level in three specifications and  $p < 0.01$  level in one specification. Those in a more positive state are more imprudent, while more negative states are associated with prudence. In **Table 3**, we report the results from similar regressions with valence measured at the exposure stage. In all four specifications in which it appears, the variable Valence is negative in sign, though it is marginally significant only in specification (5). Overall, in our view, the balance of the evidence indicates a negative relationship between positivity of emotional state and prudence.<sup>4</sup>

The second dimension of emotional state that we consider is arousal. While positive emotional state is associated with less prudence, we find that stronger arousal is associated with greater prudence. However, as we describe in the supporting argument for result 2, it is the change in arousal from the initial level that is correlated with subsequent decisions. The level of arousal at the

<sup>4</sup>We also considered whether the difference in valence at the time of feedback, between instances of positive and negative outcomes of the lottery, predict prudence in decision making in phase 2. It is presumed that individuals will tend to have more positive valence after a favorable than an unfavorable outcome. However, for those who have a relatively high value of the difference,  $\text{Valdiff} = \text{Valence(Favorable outcome)} - \text{Valence(Unfavorable outcome)}$ , might be more prudent. This is because, if a positive emotional state leads to more risk taking, and a negative emotional state leads to lower risk taking, individuals with a relatively high value of Valdiff might be more willing to apportion the unavoidable risk to the high income state. This would lead to a positive correlation between Valdiff and prudent decision making. However, no such correlation appears in the data.



**FIGURE 3 | Emotional profiles and the number of prudent decisions.**

time of exposure to or feedback from the lotteries in phase 1 is uncorrelated with the number of prudent choices in phase 2.

## Result 2: Increases in Arousal When Facing Risky Lotteries Correlates with Prudent Decision Making Support for Result 2

Specifications (2), (4), and (5) in Tables 2, 3 reveal that the absolute amount of arousal in phase 1 is not correlated with

prudence in decision making. However, as specification (3) shows, the results are different if changes in arousal from the beginning of the session to the moment of measurement are considered. In equation (3), the emotional variables are the actual value of the emotion at the moment of feedback or exposure in phase 1, minus the initial level at the beginning of the session prior to the start of phase 1. In both tables, the results show that overall arousal level does not presage more prudent decision making, but an increase in arousal when confronted with

**TABLE 2 | Number of prudent choices as a function of emotional, ability, and personality measures; feedback emotions.**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gender	0.024	0.014	0.003	0.021	0.031	0.017	0.029
Arousal		-0.258	0.283*	-0.276	-0.252		
Valence	-0.086**	-0.102**	0.039	-0.090**	-0.108***		
Raven score		0.026***	0.026***		0.026***		0.027***
Extraverted					0.012		0.019*
Agreeableness					-0.005		-0.009
Neuroticism					0.009		0.006
Conscientiousness					-0.027***		-0.028***
Openness to experiences					0.026**		0.018
Happy						0.066	-0.010
Sad						0.086	0.212
Scared						-0.319	-0.609
Angry						0.121**	0.082
Disgusted						0.327***	0.360***
Surprised						0.191***	0.153***
	Obs 770	Obs 770	Obs 761	Obs 770	Obs 770	Obs 770	Obs 770
	Groups 10						

Dependent variable is the number of prudent decisions [0, 10] made by an individual in phase 2 of the experiment. In all equations other than (3), the emotion and arousal variables are those averaged over the 10 s after the resolution of the 10 lotteries in phase 1. In Equation (3), the emotion and arousal variables are the difference between those in the 60 s before the start of phase 1 and those at the time of the resolution of the lotteries. Regressions use panel data format that adjusts the standard errors for repeated measures.

\* , \*\* , \*\*\* denotes significance at the 10%, 5%, 1% level.

**TABLE 3 | Number of prudent choices as a function of emotional, ability, and personality measures; exposure emotions.**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gender	-0.0004	-0.008	-0.015	-0.0004	0.006	-0.012	-0.003
Arousal		-0.016	0.279**	-0.006	-0.071		
Valence	-0.055	-0.054	0.026	-0.055	-0.063*		
Raven score		0.025	0.026***		0.026***		0.027***
Extraverted					0.020*		0.023**
Agreeableness					-0.005		-0.006
Neuroticism					0.006		0.004
Conscientiousness					-0.025**		-0.027***
Openness to experiences					0.019		0.015
Happy						-0.085	-0.114
Sad						0.167	0.242*
Scared						-0.150	-0.341
Angry						0.066	0.047
Disgusted						0.217***	0.238***
Surprised						0.160***	0.131***
	Obs 720	Obs 720	Obs 714	Obs 720	Obs 720	Obs 720	Obs 720
	Groups 10						

Dependent variable is the number of prudent decisions [0, 10] made by an individual in phase 2 of the experiment. In all equations other than (3), the emotion and arousal variables are those averaged over the first 10 s that the 10 lotteries in phase 1 are displayed. In Equation (3), the emotion and arousal variables are the difference between those in the 60 s before the start of phase 1 and those at the time of the display of the lotteries. Regressions use panel data format that adjusts the standard errors for repeated measures.

\* , \*\* , \*\*\* denotes significance at the 10%, 5%, 1% level.

risky lotteries does correlate with a greater number of prudent choices.

We now turn to the individual emotions as correlates of decisions. The principal pattern in the data is that more intense emotions, in particular surprise and disgust, correlate

with greater prudence. There is some evidence that greater anger and sadness also are associated with more prudence. Fear and happiness do not exhibit a significant relation with prudent decision making. Our findings are reported as result 3.

## **Result 3: Stronger Emotions Are Correlated with Greater Prudence**

### **Support for Result 3**

The results are shown in specifications (6) and (7) in **Table 2** for emotions in the feedback stage and in **Table 3** for the exposure stage. The tables reveal a significantly positive relationship between disgust and surprise with the number of prudent decisions made in all relevant equations. Sadness and anger are each significant in one of the four specifications in which they appear. In all cases, a greater value of the emotion correlates with greater prudence.

The last result considers the other correlates of prudence that our design permits us to evaluate.

## **Result 4: There Are No Gender Differences in the Average Level of prudence.**

### **Prudence is Positively Correlated with Cognitive Ability. Prudence Is Negatively Correlated with Conscientiousness**

#### **Support for Result 4**

In all of the specifications reported in **Tables 2, 3**, the variable Gender is insignificant. The variable Raven, the score of an individual on the Raven's test, is significant at the 1% level in all estimated equations in which it appears. Furthermore, none of the big 5 personality traits is significant other than conscientiousness.

## **DISCUSSION**

We observe that those who experience more positive valence at the time of the resolution of risky lotteries tend to make less prudent subsequent decisions. The same correlation obtains if valence at the time of presentation of the lotteries is considered, although this effect is only marginally significant. This result is similar in spirit to those obtained for risk aversion by a number of authors, who find that negative emotional state is associated with greater risk aversion. There are a number of possible explanations for this correlation. If a negative emotional state prompts more pessimistic beliefs, as under the Affective Generalization Hypothesis, an individual with negative valence might believe that the bad state is more likely to occur than the good state. If this is the case, and the agent is risk averse, she will apportion an unavoidable zero-mean risk to what she believes is the less likely state, i.e., the one yielding the relatively high outcome. Alternatively, it may be the case that a negative emotional state prompts individuals to behave defensively by maximizing their minimum payoff. This pattern would translate into declining to accept zero-mean risks when given an opportunity to do so (risk aversion), and apportioning unavoidable risks into relatively high income states when possible (prudence). Future research would be needed to distinguish between the hypotheses that a negative emotional state leads individuals to apply a heuristic in which they maximize their minimum payoff and the alternative that negative emotions prompt more risk averse as well as more prudent decisions.

We also observe that increases in arousal during the phase 1 task, which can be interpreted as integral arousal, is positively correlated with prudence in subsequent decisions. It may be the case that greater arousal, like more negative valence, leads to more pessimistic beliefs. The consequence would be that the high income state is viewed as less likely, and that a risk averse individual would allocate the risk to what she believes is the less likely state, and generate behavior consistent with prudence. Alternatively, arousal may lead to a focus on relatively unfavorable outcomes and choices that maximize payoff under the worst possible outcome. While some prior research associates greater arousal with risk taking (Haim, 1994), other work argues that underarousal increases risk taking as individuals seek arousing stimuli (Schmidt et al., 2013). Here, it may be the case that underaroused individuals place the risk in the low income state as stimulation to increase their level of emotional arousal.

An overall pattern emerges with respect to the relationship between individual emotions and prudence in decision making. This is that stronger emotions are associated with more prudent decision making. The result is also similar to, and might be viewed as somewhat of an extension of, those reported by Nguyen and Noussair (2014), who also find that stronger emotions correlate with risk aversion, though they observe their relationship for a different set of emotions. Explaining why there is a relationship between more intense emotions and prudence is beyond the scope of what this experiment can test, but the explanations may be similar to those proposed for the correlation between prudence and valence or arousal described above. Strong emotions might influence beliefs about the likelihood of each state or encourage the use of heuristics such as the maximization of minimum payoff.

The absence of a gender effect and the strong link between prudence and cognitive ability echoes the results of Noussair et al. (2014), who observed the same patterns in a large demographically representative sample of the Dutch population. The emerging pattern with regard to gender differences in prudence contrasts with that for risk aversion, in which gender differences are widely observed (see e.g., Eckel and Grossman, 2008). The particular relationship we observe between personality and prudence is surprising for a couple of reasons. The first reason is that the Big Five personality characteristics and risk aversion exhibit a pattern of correlation that is both strong and intuitive to interpret. Here, the relationship is relatively weak with only conscientiousness exhibiting a robust relationship. The second reason is that because prudence is associated with high cognitive ability and precautionary savings, one might think that it would also be correlated with greater conscientiousness, rather than less, as we observe here. However, the effect of conscientiousness remains in regressions (not reported here but available from the authors), in which Raven's score is left out of the specification. The effect of conscientiousness becomes insignificant when the emotional state variables of valence, arousal, and specific emotions are not included in the specification, suggesting that emotional states may affect individuals' decisions differently, depending on their personality profile. Conducting an analysis of the mediating and moderating relationships of such a large number of personality characteristics

and emotional variables on prudence would require a much larger data set than we gathered for this study, but we believe it would be worthwhile to pursue such an analysis in future work.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication. All authors contributed equally; authors names appear in alphabetical order.

## REFERENCES

- Alm, J. (1988). Uncertain tax policies, individual behavior, and welfare. *Am. Econ. Rev.* 78, 237–245.
- Becker, A., Dekkers, T., Dohmen, T., Falk, A., and Kosse, F. (2012). The relationship between economic preferences and psychological personality measures. *Annu. Rev. Econ.* 4, 453–478. doi: 10.1146/annurev-economics-080511-110922
- Bors, D. A., and Stokes, T. L. (1998). Raven's advanced progressive matrices: norms for first-year university students and the development of a short form. *Educ. Psychol. Meas.* 58, 382–398. doi: 10.1177/0013164498058003002
- Breaban, A., and Noussair, C. N. (2014). *Emotional State and Market Behavior*. Working paper, Tilburg University.
- Chentsova-Dutton, Y. E., and Tsai, J. (2010). Self-focused attention and emotional response: the role of culture. *J. Pers. Soc. Psychol.* 98, 507–519. doi: 10.1037/a0018534
- Deck, C., and Schlesinger, H. (2010). Exploring higher order risk effects. *Rev. Econ. Stud.* 77, 1403–1420. doi: 10.1111/j.1467-937X.2010.00605.x
- Deck, C., and Schlesinger, H. (2014). Consistency of higher order risk preferences. *Econometrica* 82, 1913–1943. doi: 10.3982/ECTA11396
- Ebert, S., and Wiesen, D. (2011). Testing for prudence and skewness seeking. *Manage. Sci.* 57, 1334–1349. doi: 10.1287/mnsc.1110.1354
- Ebert, S., and Wiesen, D. (2014). Joint measurement of risk aversion, prudence, and temperance. *J. Risk Uncertain.* 48, 231–252. doi: 10.1007/s11166-014-9193-0
- Eckel, C. C., and Grossman, P. J. (2008). “Men, women, and risk aversion: experimental evidence,” in *The Handbook of Experimental Economic Results*, eds C. Plott and V. Smith (Amsterdam: Elsevier), 1061–1073.
- EEckhoudt, L., and Schlesinger, H. (2006). Putting risk in its proper place. *Am. Econ. Rev.* 96, 280–289. doi: 10.1257/000282806776157777
- Ekman, P. (2003). *Emotions Revealed, 2nd Edn*. New York, NY: Times Books.
- Ekman, P., and Friesen, W. (2003). *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Los Altos, CA: Alor publishers.
- Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *J. Pers. Soc. Psychol.* 53, 712.
- Esö, P., and White, L. (2004). Precautionary bidding in auctions. *Econometrica* 72, 77–92. doi: 10.1111/j.1468-0262.2004.00478.x
- Fischbacher, U. (2007). Z-tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* 10, 171–178. doi: 10.1007/s10683-006-9159-4
- Frederick, S. (2005). Cognitive reflection and decision making. *J. Econ. Pers.* 19, 25–42. doi: 10.1257/089533005775196732
- Friedman, M. (1957). “The permanent income hypothesis,” in *A Theory of the Consumption Function* (Princeton, NJ: Princeton University Press), 20–37.
- Gosling, S. D., Rentfrow, P. J., and Swann, W. B. (2003). A very brief measure of the big-five personality domains. *J. Res. Pers.* 37, 504–528. doi: 10.1016/S0092-6566(03)00046-1
- Habetanova, L., and Noussair, C. N. (2015). *Charitable Giving, Emotions, and the Default Effect*. Working paper, Tilburg University.
- Hager, J. C., and Ekman, P. (1985). The asymmetry of facial actions is inconsistent with models of hemispheric specialization. *Psychophysiology* 22, 307–318.
- Haim, M. (1994). Risk-taking, framing effects, and affect. *Organ. Behav. Hum. Decis. Process.* 57, 38–58.
- Hess, U., and Kleck, R. (1990). Differentiating emotion elicited and deliberate emotional facial expressions. *Eur. J. Soc. Psychol.* 20, 369–385.
- Isen, A. M., Nygren, T. E., and Ashby, F. G. (1988). Influence of positive affect on the subjective utility of gains and losses: it is just not worth the risk. *J. Pers. Soc. Psychol.* 55, 710–717.
- Johnson, E. J., and Tversky, A. (1983). Affect, generalization, and the perception of risk. *J. Pers. Soc. Psychol.* 45, 20–31.
- Kimball, M. S. (1990). Precautionary saving in the small and in the large. *Econometrica* 58, 53–73.
- Kugler, T., Connolly, T., and Ordonez, L. D. (2012). Emotion, decision, and risk: betting on gambles versus betting on people. *J. Behav. Decis. Making* 25, 123–145. doi: 10.1002/bdm.724
- Lerner, J. S., and Keltner, D. (2001). Fear, anger, and risk. *J. Pers. Soc. Psychol.* 81:146. doi: 10.1037/0022-3514.81.1.146
- Lerner, J. S., and Tiedens, L. (2006). Portrait of the angry decision maker: how appraisal tendencies shape anger's influence on cognition. *J. Behav. Decis. Making* 19, 115–137. doi: 10.1002/bdm.515
- Lewinski, P., de Uyl, T., and Butler, C. (2014). Automated facial coding: validation of basic emotions and FACS AUs in facereader. *J. Neurosci. Psychol. Econ.* 7, 227–236. doi: 10.1037/npe0000028
- Nguyen, Y., and Noussair, C. N. (2014). Risk aversion and emotions. *Pacific Econ. Rev.* 19, 296–312. doi: 10.1111/1468-0106.12067
- Nicholson, N., Soane, E., Fenton-O'Creevy, M., and Willman, P. (2005). Personality and domain-specific risk taking. *J. Risk Res.* 8, 157–176. doi: 10.1080/1366987032000123856
- Noussair, C. N., Trautmann, S. T., and van de Kuilen, G. (2014). Higher order risk attitudes, demographics, and financial decisions. *Rev. Econ. Stud.* 81, 325–355. doi: 10.1093/restud/rdt032
- Raven, J., Raven, J. C., and Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 4: The Advanced Progressive Matrices*. San Antonio, TX: Harcourt Assessment.
- Schmidt, B., Mussel, P., and Hewig, J. (2013). I'm too calm—let's take a risk! on the impact of state and trait arousal on risk taking. *Psychophysiology* 50, 498–503. doi: 10.1111/psyp.12032
- Starmer, C. (2000). Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. *J. Econ. Lit.* 38, 332–382. doi: 10.1257/jel.38.2.332
- Teixeira, T., Wedel, M., and Pieters, R. (2012). Emotion-induced engagement in internet video advertisements. *J. Mark. Res.* 49, 144–159. doi: 10.1509/jmr.10.0207
- Thaler, R., and Johnson, E. (1990). Gambling with the house money and trying to break even: the effect of prior outcomes on risky choice. *Manage. Sci.* 36, 643–660.
- Treich, N. (2010). Risk-aversion and prudence in rent-seeking games. *Public Choice* 145, 339–349. doi: 10.1007/s11127-009-9569-x

## FUNDING

We thank the VIDI program of NWO for funding to support this experiment.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01688/full#supplementary-material>

- Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertain.* 5, 297–323.
- Van Leeuwen, B., Noussair, C. N., Offerman, T. J. S., Suetens, S., Van Veelen, M., and Van de Ven, J. (2014). *Predictably Angry: Facial Cues Provide a Credible Signal of Destructive Behavior*. Working paper, Tilburg University.
- White, L. (2008). Prudence in bargaining: the effect of uncertainty on bargaining outcomes. *Games Econ. Behav.* 62, 211–231. doi: 10.1016/j.geb.2006.11.006

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Breaban, van de Kuilen and Noussair. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Does the Dark Triad of Personality Predict Corrupt Intention? The Mediating Role of Belief in Good Luck

Huanhuan Zhao, Heyun Zhang and Yan Xu \*

School of Psychology, Beijing Normal University, Beijing, China

The current study is the first attempt to examine the association between the Dark Triad of personality (i.e., Machiavellianism, narcissism, and psychopathy) and corruption through a mediator—belief in good luck. Based on Ajzen's theory of planned behavior, we assumed that individuals with Dark Triad would be more likely to engage in corruption as a result of belief in good luck. In Study 1, a set of hypothetical scenarios was used to assess the bribe-offering intention and the corresponding belief in good luck. Results indicated that while the Dark Triad of personality positively predicted bribe-offering intention, it was mediated by the belief in good luck in gain-seeking. In Study 2, we presented participants with some hypothetical scenarios of bribe-taking and the corresponding belief in good luck. Findings revealed that the Dark Triad of personality was positively related to bribe-taking intention; the relationship between narcissism and bribe-taking intention, and that between psychopathy and bribe-taking intention was mediated by the belief in good luck in penalty-avoidance. However, this belief in good luck did not mediate the relationship between Machiavellianism and bribe-taking intention. These results hold while controlling for demographic variables, dispositional optimism, and self-efficacy. Taken together, this study extended previous research by providing evidence that belief in good luck may be one of the reasons explaining why people with Dark Triad are more likely to engage in corruption regardless of the potential outcomes. Theoretical and practical implications were discussed.

## OPEN ACCESS

### **Edited by:**

Nikolaos Georgantzis,  
University of Reading, UK

### **Reviewed by:**

Tarek Jaber-Lopez,  
INRA and BETA-Université  
Strasbourg, France  
Michalis Drouvelis,  
University of Birmingham, UK

### **\*Correspondence:**

Yan Xu  
[xuyan@bnu.edu.cn](mailto:xuyan@bnu.edu.cn)

### **Specialty section:**

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

**Received:** 12 March 2016

**Accepted:** 12 April 2016

**Published:** 28 April 2016

### **Citation:**

Zhao H, Zhang H and Xu Y (2016)  
Does the Dark Triad of Personality  
Predict Corrupt Intention? The  
Mediating Role of Belief in Good Luck.  
*Front. Psychol.* 7:608.  
doi: 10.3389/fpsyg.2016.00608

## INTRODUCTION

Corruption exists widely. It is commonly defined as deviant behavior that deliberately breaks legal or moral norms and abuses public authority or resources for personal gain (He, 2000; Lindgreen, 2004; Rabl and Kuhlmann, 2008). It impairs political stability, damages economic growth, misallocates public resources, hinders normal upward social mobility, increases social inequality, undermines people's trust in government, and lowers moral standards in a society (He, 2000; Lu and Gunnison, 2003; Alesina and Angeletos, 2005; Sobhani and Bechara, 2011). Since corruption harms a society tremendously, a thorough understanding of it followed by the proper counter measures becomes extremely important.

When it comes to corruption, the following important questions are often raised first: What kinds of people are more likely to act corruptly? What type of personality they possess that leads them to gain profits from corruption at the expense of others? Why do these people tend to engage

in corruption more often than others? Though previous research tried to uncover the occurrence of corruption at both macro (Treisman, 2000; Blackburn and Forques-Puccio, 2009) and micro level (Jaber-López et al., 2014), and found that corruption was a result of interactions among various variables (e.g., political, social, economic, or psychological factors), yet to date, little research has explored the wider range of personality traits potentially associated with corrupt behaviors. A growing body of evidence suggested that the Dark Triad of personality (i.e., Machiavellianism, narcissism, and psychopathy) was associated with unethical behaviors (Egan et al., 2015; Azizli et al., 2016; Roeser et al., 2016). We reckoned that this association may extend to corrupt behaviors. Thus, the first purpose of this study is to examine whether people with Dark Triad are more likely to engage in corruption.

Furthermore, despite the increasing evidence justifying the effects of the Dark Triad traits on various deviant behaviors, scant attention has been given to the underlying mechanism and processes through which this relationship occurs. Therefore, the second purpose of this study is to explore the psychological mechanism that underlies the association between the Dark Triad traits and corruption. According to Ajzen's theory of planned behavior, behavioral dispositions, such as personality traits and social attitudes, played a critical role in predicting cognitive beliefs (e.g., behavioral beliefs, normative beliefs, and control beliefs), which in turn explain human behaviors (Fishbein and Ajzen, 1975; Ajzen, 1991). Ajzen (1991) proposed that personality traits influenced behavior indirectly through cognitive beliefs. Belief in good luck, as an irrational cognitive belief, thus may be affected by one's particular personality trait. Additionally, belief in good luck has been proved to shape one's behaviors (Chiu and Storm, 2010). Accordingly, we assume that the relationship between the Dark Triad of personality and corruption is mediated by belief in good luck.

## The Dark Triad of Personality and Corruption

The Dark Triad consists of three antisocial personality traits: Machiavellianism, narcissism, and psychopathy (Paulhus and Williams, 2002). Machiavellianism is portrayed as calculated manipulation, duplicity, cunningness, and a disregard for morality (Hodson et al., 2009; Rauthmann and Kolar, 2013; Djeriouat and Tremoliere, 2014). Narcissism is characterized by optimistic egotism (Jones, 2013), and is positively correlated with self-centeredness, sense of superiority, entitlement, vanity, and grandiosity (Crysel et al., 2013; Rauthmann and Kolar, 2013; Buelow and Brunell, 2014). Narcissists often pursue immediate profits at the expense of others' interests (Lakey et al., 2008; Foster et al., 2009). Psychopathy is linked to high impulsivity, callousness, and socially aversive behaviors (Hodson et al., 2009; Rauthmann and Kolar, 2013). According to the life history theory, individuals with Dark Triad manifest the fast life strategy characterized by a disregard for social rules, short-term thinking, and extensive future-discounting related behaviors (Jonason et al., 2010, 2012). These traits are positively related to numerous deviant behaviors like gambling (Jones and Figueiredo, 2013),

lying and deception (Azizli et al., 2016), cyber-aggression (Pabian et al., 2015), and white-collar crime (Egan et al., 2015). Since corruption is a deviant behavior and can be a criminal offense (Rabl and Kuhlmann, 2008), we posit that the Dark Triad of personality may be positively associated with corrupt behaviors.

Additionally, corruption is based on the exchange between at least two partners, i.e., the bribe giver and taker strike a deal, often by putting their personal interests ahead of others' (Rabl and Kuhlmann, 2008). Research indicated that people with Dark Triad fit this description. Each trait in Dark Triad may have a unique set of features, but all three have something in common as well. First, all three traits are related to the willingness to gain profit at the expense of others (Jones, 2013). Individuals who exhibit high Dark Triad tendencies employ devious means to achieve personal goals with little concern for others' interests (Linton and Power, 2013). Second, the common features of the Dark Triad, such as manipulation, callousness, and selfishness, positively predict deliberate toxic behaviors (O'Boyle et al., 2012; Jones and Figueiredo, 2013). Machiavellians often manipulate others for personal gain, against others' welfare (Tang et al., 2008). Narcissists are relentless and toxic when they have power (Schmidt, 2008). A recent study found that individuals with psychopathy share many behavioral characteristics observed in patients suffering from ventromedial prefrontal cortex and amygdala lesions. This finding served as a neuroscientific evidence to explain why psychopaths engage in corrupt and immoral behaviors (Sobhani and Bechara, 2011). Therefore, considering the generally antisocial and socially undesirable nature of the Dark Triad traits, we made an assumption that the Dark Triad traits would predict greater intention to engage in corruption (Hypothesis 1).

## The Mediating Role of Belief in Good Luck

As pointed out previously, people with Dark Triad may tend to engage in corrupt behaviors. The question is, what motivates these "Dark Triad" people to ignore the law and involve in corruption? According to the theory of planned behavior, belief in good luck can provide a new perspective to look at the reasons behind the Dark Triad people engaging in corruption.

### Belief in Good Luck

We can easily observe in daily life the following phenomena: some people cannot stop a gambling game, for they believe that they will have a good luck to win the game, and the good luck makes them believe that their chance of winning will be high (e.g., 70%), despite the actual winning probability is very low (e.g., 5%) and they have lost many times. Or, some people believe that they will have a good luck which makes them not being caught or have low chances of being caught (e.g., 5%) if they cheat just once in an examination, despite the actual probability of being caught is very high (e.g., 60%) and many cheaters have been caught. This irrational belief, often closely related to the negative or deviant behaviors, is named "belief in good luck" in this study. Belief in good luck is an irrational cognition about luck (Day and Maltby, 2003). It can increase ones' unrealistic optimism and self-efficacy (Darke and Freedman, 1997a; Damisch et al., 2010), and affect their future expectations (Darke and Freedman, 1997b).

Belief in good luck is often manifested as blindness in making decisions on probability events, especially with reference to the events involving deviant behaviors (Chiu and Storm, 2010). It should be noted that there are two outcomes of negative or deviant behaviors: positive-valence and negative-valence outcomes. In the above examples, people are attracted to the deviant behaviors of gambling and cheating through different mechanisms. While the former involves overrating the probability of a positive-valence outcome (i.e., winning the game) despite its actual probability is low, the latter involves underrating the probability of a negative-valence outcome (i.e., being caught) despite its actual probability is high. If the actual probability of a positive-valence outcome is very low and the actual probability of a negative-valence outcome is very high, but people still irrationally believe that they will have a good luck, and the corresponding good luck makes them irrationally believe that they are more likely to experience a positive-valence outcome and less likely to experience a negative-valence outcome, then this is the effect of what we call “belief in good luck.” When good luck was thought of as a personal quality possessed by persons, it could provide a perceived ability that can be used to exert control over what otherwise may consider a chance event (Wohl and Enzle, 2002). Accordingly, belief in good luck exerts its influence on negative or deviant behavior via two mechanisms: (1) irrationally overestimating the probability of a positive-valence outcome when its actual probability is very low; and (2) irrationally underestimating the probability of a negative-valence outcome when its actual probability is very high.

Obviously, the operational definition of belief in good luck is somewhat similar to the prospect theory, which also emphasizes decision-making and probability estimation (Tversky and Kahneman, 1981, 1992). The prospect theory suggests that people tend to overrate the small probabilities and underrate the moderate and large probabilities regardless of the nature of the event (e.g., good or deviant behavior) and the valence of the event outcome (i.e., positive or negative valence; Tversky and Kahneman, 1981, 1992; Kusev et al., 2009). Nevertheless, belief in good luck takes into account of the outcome valence of deviant event as well as its actual probability. Additionally, we should point out that it is plausible to make the inference that when the probability of a positive-valence outcome is very high, people may still overrate it, and when the probability of a negative-valence outcome is very low, people may still underrate it, despite the respective degrees of overestimation and underestimation may be very small. It is reasonable and natural for people to expect that they will experience a positive-valence outcome when its probability is very high and will not suffer from a negative-valence outcome when its probability is very low. Evidently, these two cases not only contradict the prospect theory (Tversky and Kahneman, 1981, 1992), but also in line with people’s rational expectations, therefore, are not considered as an irrational belief in good luck.

Based on the prospect theory and the operational definition of belief in good luck, we used the adapted research paradigm of “*objective probability event-subjective probability estimation*” to measure one’s belief in good luck in corruption. Here the objective probability event means the actual probability of

deviant event outcomes, whereas the subjective probability estimation means people’s irrational overestimation and underestimation. We also employed two different outcome valences of the forms of corruption in this study to verify the related mechanisms of belief in good luck. As to bribe-offering, in order to gain unfair advantages over others, one may offer a bribe to someone who is in power. However, in China, since bribe-offering is legally much less penalized than bribe-taking (Wang and Wu, 2012), we focused on the likelihood of seeking gains via bribe-offering (a positive-valence outcome of deviant behavior). We contended that some people may have the lucky belief and tend to irrationally overestimate the probability of seeking personal advantages via bribe-offering (*namely belief in good luck in gain-seeking*) (Study 1). Additionally, since bribe-taking behaviors in China face much more severe penalties and involvement in bribe-taking is becoming increasingly risky (Gong, 2002; Lu and Gunnison, 2003), the factor to focus on here is the probability of being penalized in bribe-taking (a negative-valence outcome of deviant behavior). We speculated that some people may hold the lucky belief and have a tendency to irrationally underestimate the likelihood of being penalized for bribe-taking (*namely belief in good luck in penalty-avoidance*) (Study 2).

### The Dark Triad of Personality and Belief in Good Luck

The theory of planned behavior suggests that personality traits play an important role in predicting cognitive beliefs (Fishbein and Ajzen, 1975; Ajzen, 1991). People’s personality influences how they perceive and evaluate things around them (Andre, 2006; Jibeen, 2015). If a personality trait toward cognitive irrationality is rooted largely in innate or biological differences, it is more likely to result in irrational beliefs (Andre, 2006; Yang et al., 2007; Samar et al., 2013; Jibeen, 2015). For example, the Big-five personality traits have been proven to predict people’s irrational beliefs (Samar et al., 2013; Jibeen, 2015). As important personality characteristics, the Dark Triad traits were closely associated with Big-five personality traits (Lee and Ashton, 2005), thus suggesting that the latter would also predict and affect individuals’ irrational and unrealistic beliefs. A series of compelling studies support our inference (Paulhus and Williams, 2002; Lakey et al., 2008; Jones, 2014; Birkas et al., 2015).

Research has shown that high Machiavellianism was associated with greater perceived reward for engaging in deviant behavior and less perceived punishment for that activity (Birkas et al., 2015). This indicates that Machiavellians tend to consider merely the profits they want to pursue, which may result in their erroneous estimation about the odds of rewards or punishment (Rauthmann and Kolar, 2013; Birkas et al., 2015). They may form unrealistic beliefs about good luck, and irrationally overrate the gain-seeking probability and underrate or even neglect the punishment probability. Additionally, individuals with high narcissism often possess a sense of entitlement (Morf et al., 2000) and overconfidence (Campbell et al., 2004), which lead them to misjudge the chances of success. In other words, they may inappropriately raise the subjective probabilities of successes (Paulhus and Williams, 2002). Besides, the inflated self-beliefs caused narcissists to

form an unrealistic view that luck always works in their favor, which make them underrate the probability of risks or losses and arrive at irrational decisions (Judge et al., 2006; Chatterjee and Hambrick, 2007; Lakey et al., 2008). As such, people with high narcissism may hold the illusory beliefs that good luck would fall to them and they can control an event. Results from previous research also demonstrated that psychopathy was positively associated with irrational beliefs (Samar et al., 2013); people high in psychopathy tend to exhibit a biased judgment of risk perceptions, or even ignore the inherent risks related to an event (Jones, 2014). In addition, the characteristic of low self-control renders them unable to resist the temptations from unfair advantages, which cause them unable to keep a cool mind to make rational judgments (Tangney et al., 2004). Indeed, all of these would exacerbate the “Dark Triad” people’s unrealistic estimations about potential gains or risks. Taken together, it is nature to assume that when faced with deviant behaviors, individuals with high Dark Triad would be more likely to hold irrational beliefs in good luck and make irrational estimations.

### **Belief in Good Luck and Corruption**

Ajzen’s theory of planned behavior also indicated that individuals’ cognitive beliefs about a behavior are considered as the prevailing determinants of their behavioral tendencies (Fishbein and Ajzen, 1975; Ajzen, 1991). Additionally, a number of cognitive-behavioral theories suggested that the deviant behaviors are caused by inaccurate or irrational beliefs (Ellis, 1999; Andre, 2006; Jibeen, 2015). Although, a previous study has suggested that beliefs about luck can serve as a positive expectation for future events to a certain degree (Darde and Freedman, 1997a), when confronted with antisocial and unethical behaviors, such as corruption, this irrational belief will lead to serious consequences. Research has shown that belief in good luck generated the feelings of illusion of control and optimistic bias (Darde and Freedman, 1997a,b), and these unrealistic feelings were prevalent amongst gambling and risk-taking behaviors (Darde and Freedman, 1997b; Chiu and Storm, 2010). For example, the gamblers’ perception of themselves being lucky led them to continue gambling (Wohl and Enzle, 2003; Chiu and Storm, 2010); and the lower perceived likelihood of punishment lead people to have a higher perceived corrupt intention (Bai et al., 2014). Therefore, as to corruption, under the influence of lucky belief, irrationally overestimating the likelihood of seeking personal benefits via bribe-offering, and inappropriately underestimating the likelihood of being penalized for bribe-taking, would together make people have a strong tendency to engage in bribe-offering and bribe-taking behaviors. Thus, engaging in corruption is, at least to some degree, dependent upon one’s irrational beliefs in good luck in seeking gains or avoiding penalty. We then propose that the more people believe in good luck, the more likely they would be to engage in corruption.

Given that belief in good luck is closely linked to the Dark Triad of personality and corrupt behaviors, based on the theory of planned behavior, it is reasonable to hypothesize that belief in good luck may play a mediating role in the relationship between the Dark Triad traits and corrupt intention (Hypothesis 2).

## **Overview of the Current Studies**

Based on the literature review of previous theories and studies, we posited the following two hypotheses:

Hypothesis 1: The Dark Triad traits positively predict corrupt intention; and

Hypothesis 2: Belief in good luck plays a mediating role in the relationship between the Dark Triad traits and corrupt intention.

We conducted two sub-studies in China to test Hypothesis 1 and Hypothesis 2. In these sub-studies, bribe-offering and bribe-taking were used as two different outcome valences of the forms of corruption in hypothetical scenarios, and the measurement of belief in good luck was embedded in the corresponding scenario. In Study 1, we explored the correlations between each Dark Triad trait and bribe-offering intention, and constructed mediation models to verify the assumption that belief in good luck in seeking gains (irrationally overestimating the probability of a positive-valence outcome) would mediate the effect of the Dark Triad traits on bribe-offering intention. In Study 2, we further examined whether each Dark Triad trait could facilitate bribe-taking intention, and tested whether belief in good luck in avoiding penalty (irrationally underestimating the probability of a negative-valence outcome) would mediate the effect of the Dark Triad traits on bribe-taking intention.

## **STUDY 1**

The aim of Study 1 was twofold. First, we examined whether the Dark Triad traits could predict bribe-offering intention. We expected that individuals with high Dark Triad traits were more likely engage in bribe-offering behaviors. Second, we explored whether one’s belief in good luck can mediate the effects of the Dark Triad traits on bribe-offering intention. We predicted that individuals with Dark Triad tend to engage in corruption partially because they hold the lucky belief and irrationally overestimate the likelihood of seeking gains via bribe-offering.

### **Methods**

#### **Participants**

A total of 404 Chinese adults were recruited online, via the Qualtrics Survey from different enterprises in China. The final valid sample comprised 395 Chinese adults (231 female and 164 male;  $M_{age} = 29.56$  years,  $SD = 6.30$  years; age range: 18–60 years), as 9 adults were excluded because 4 of them failed to complete the questionnaires, and 5 responded with extreme values. The effective response rate was 97.77%. Participants varied considerably in terms of their education levels (18.2% with high school education or less, 31.4% with a college degree, 42.0% with a bachelor degree, and 8.4% with a postgraduate degree) and monthly income (14.4% with less than 2000 yuan, 49.1% with 2001–5000 yuan, 24.6% with 5001–8000 yuan, 9.1% with 8001–20,000 yuan, and 2.8% with more than 20,000 yuan).

#### **Procedure**

After signing a consent form, the participants completed a series of self-report questionnaires within 45 mins. These

questionnaires included the Short Dark Triad scale, the bribe-offering intention measure, the belief in good luck measure, the life orientation test revised and the new general self-efficacy scale. After they completed the questionnaires, the participants were asked to provide demographic information, including gender, age, education level and monthly income. Upon completion, participants were thanked and debriefed. Anonymity of the participants was ensured in order to reduce social desirability, and all procedures were approved by the ethics board at the School of Psychology, Beijing Normal University.

## Measures

### *The short Dark Triad (SD3)*

Following Jones and Paulhus (2014) methods, the 27-item Dark Triad scale was translated via a back-translation procedure and then was used to assess the Dark Triad personality traits. This scale is divided into three dimensions: Machiavellianism (e.g., "Make sure your plans benefit yourself, not others"), narcissism (e.g., "People see me as a natural leader") and psychopathy (e.g., "I like to get revenge on authorities"). Each sub-scale comprises 9 items. Participants were asked to rate their level of agreement with each item on a 5-point Likert scale (1 = *strongly disagree*, 5 = *strongly agree*). Higher scores on the scale indicate higher levels of Dark Triad tendencies. In this study, the Cronbach's  $\alpha$  for the total scale was 0.87, and that for Machiavellianism, narcissism and psychopathy was 0.71, 0.78, and 0.78, respectively. A confirmatory factor analysis also showed a good fit for the measurement model ( $\chi^2/df = 1.83$ , GFI = 0.91, CFI = 0.90, RMSEA = 0.05). The results of reliability and validity analyses indicated that this scale was applicable to a Chinese sample.

### *The bribe-offering scenario*

Participants were asked to read three hypothetical daily life scenarios on bribe-offering, which were generated from a panel discussion (see Supplementary Material). Each participant read the following instructions before the task: "Please vividly imagine that you are in each situation." The following is a sample of a bribe-offering scenario:

"Suppose you are a section-level employee who has a strong desire to gain a promotion. The municipal government is currently selecting and promoting one section chief. You are in a disadvantaged position in the competition compared with other section-level candidates. Before the final decision, you ask the deputy mayor to help you and plan to privately promise him a certain sum of money as a token of your thanks if you win in the competition. You are aware that winning the competition via offering a bribe is an unlawful behavior."

After each scenario, propensity to engage in bribe-offering was assessed by instructing the participants to "Please estimate the likelihood you would offer the bribe to someone who is in charge" on a 7-point Likert scale (1 = *extremely unlikely*, 7 = *extremely likely*). The index of bribe-offering intention was calculated as the average score of the three scenarios, where a higher score indicates greater intention of bribe-offering. The Cronbach's  $\alpha$  of this tool was 0.80.

### *Belief in good luck*

In this study, we adopted the research paradigm of "objective probability event-subjective probability estimation" to measure one's belief in good luck related to corruption. After each bribe-offering scenario, there was a corresponding scenario to measure one's belief in good luck in gain-seeking. The following is a sample of such a scenario related to the previously presented example of a bribe-offering scenario:

"Suppose you have offered a bribe to the deputy mayor privately. According to some recent studies related to this scenario, the probability of securing a promotion via bribe-offering is only about 5% in recent years. Please respond to the following two items: (1) Despite the low probability, however, you still definitely believe that you will have a good luck to gain promotion when you offer the bribe; (2) The good luck makes you believe that your winning probability will be significantly higher than 5% when you offer the bribe."

Participants responded to each item on a 7-point Likert scale (1 = *strongly disagree*, 7 = *strongly agree*). Three scenarios were averaged together to create an indicator of belief in good luck, where higher scores were indicative of higher levels of belief in good luck. The Cronbach's  $\alpha$  of this tool was 0.79. It is important to note that according to the statistical probability, the 5% rate that was used in the scenarios present in this study was a very low probability event.

### *Control variables*

We included gender, age, education, income, dispositional optimism, and self-efficacy as control variables that potentially influenced the findings. For example, research has suggested that gender, age, education, and income significantly affect corruption (Čábelková and Hanousek, 2004; Donchev and Ujhelyi, 2007). Moreover, optimism and self-efficacy were positively associated with belief in good luck (Darke and Freedman, 1997a; Damisch et al., 2010). Therefore, these variables were assessed and controlled to isolate the independent impacts of the Dark Triad traits and belief in good luck on corrupt intention in our following analyses.

The 10-item Life orientation test revised measure was used to evaluate the participants' level of dispositional optimism (Scheier et al., 1994). Six items were used to assess optimism (e.g. "I'm always optimistic about my future") and four items used as filler items were not scored. All items were rated on a 5-point Likert scale (1 = *strongly disagree*, 5 = *strongly agree*), with higher scores indicating a greater degree of dispositional optimism. The Cronbach's alpha was 0.83.

Self-efficacy was measured using the 8-item new general self-efficacy scale (Chen et al., 2001). One sample item is "In general, I think that I can obtain outcomes that are important to me". Participants completed these items on a 5-point Likert scale (1 = *strongly disagree*, 5 = *strongly agree*), with higher scores representing a greater degree of generalized self-efficacy. The Cronbach's alpha was 0.88.

## Results

### Discriminant Validity

To examine the discriminant validity of belief in good luck, we conducted a confirmatory factor analysis on belief in good luck, dispositional optimism, and self-efficacy. Results showed that a three-factor model provided a good fit to the data [ $\chi^2_{(116, N = 395)} = 278.45, p < 0.001$ , GFI = 0.92, CFI = 0.93, RMSEA = 0.06], all factor loadings were statistically significant, with standardized loadings ranging from 0.60 to 0.80. Model fit was significantly better for the three-factor model compared with a single-factor model [ $\Delta\chi^2_{(3, N = 395)} = 998.50, p < 0.001$ ], a two-factor model that combined belief in good luck and dispositional optimism into one factor [ $\Delta\chi^2_{(2, N = 395)} = 336.43, p < 0.001$ ], and a two-factor model that combined belief in good luck and self-efficacy into one factor [ $\Delta\chi^2_{(2, N = 395)} = 343.23, p < 0.001$ ].

### Descriptive Analyses

Means, standard deviations, and zero-order correlation coefficients among the variables have been presented in **Table 1**. As hypothesized, the results showed that there were significant correlations between each Dark Triad trait, belief in good luck, and bribe-offering intention. More specifically, Machiavellianism, narcissism, and psychopathy were positively correlated with both belief in good luck and bribe-offering intention.

### Testing the Mediating Role of Belief in Good Luck in the Relationship between the Dark Triad Traits and Bribe-Offering Intention

In order to test Hypothesis 1 that the Dark Triad traits would predict corrupt intention, we first entered the control variables and then the three Dark Triad traits in the hierarchical regression analysis. The results showed that Machiavellianism ( $\beta = 0.30, p < 0.001$ , 95% CI [0.22, 0.39]), narcissism ( $\beta = 0.15, p < 0.01$ , 95% CI [0.04, 0.25]), and psychopathy ( $\beta = 0.15, p < 0.01$ , 95% CI [0.05, 0.25]) were positive predictors of bribe-offering intention. Thus, Hypothesis 1 was supported, indicating that people with

higher levels of the Dark Triad tendencies were likely to exhibit a higher corrupt intention.

To explain the psychological process underlying the effects of the Dark Triad traits on corrupt intention, we conducted regression analyses according to the specification set out by Andrew Hayes' (2013) PROCESS for SPSS using Model 4 (a bootstrapping CI method with  $N = 5000$  bootstrap samples) to test the mediation effect of belief in good luck on the relationship between the Dark Triad traits and corrupt intention. The mediation effects were statistically significant when the 95% confidence intervals did not include zero. To yield standardized coefficients, all variables were converted to  $z$ -scores prior to analysis. As illustrated in **Tables 2–4** and **Figures 1–3**, after adjusting for the control variables, belief in good luck in gain-seeking was found to mediate the associations between each Dark Triad trait and bribe-offering intention. Thus, Hypothesis 2 was confirmed.

When Machiavellianism was the independent variable, the link between Machiavellianism and bribe-offering intention was significantly mediated by belief in good luck in seeking gains ( $\beta_{\text{indirect}} = 0.06, SE = 0.02, 95\% \text{ CI} [0.03, 0.10]$ ), as depicted in **Table 2** and **Figure 1**.

When narcissism was the independent variable, we found that the relationship between narcissism and bribe-offering intention was fully mediated by belief in good luck in seeking gains ( $\beta_{\text{indirect}} = 0.09, SE = 0.02, 95\% \text{ CI} [0.05, 0.14]$ ), as depicted in **Table 3** and **Figure 2**.

When psychopathy was the independent variable, the results indicated that belief in good luck in seeking gains partially mediated the relationship between psychopathy and bribe-offering intention ( $\beta_{\text{indirect}} = 0.04, SE = 0.02, 95\% \text{ CI} [0.01, 0.08]$ ), as depicted in **Table 4** and **Figure 3**.

### Discussion

The results supported Hypothesis 1 and 2, demonstrated that the Dark Triad traits foster corruption and this effect was mediated by belief in good luck. The association between narcissism and bribe-offering intention was fully mediated by belief in good luck,

**TABLE 1 | Descriptive statistics and correlations among the variables.**

Variables	M	SD	1	2	3	4	5	6	7	8	9	10	11
1. Gender	0.58	0.49	—										
2. Age	29.56	6.30	−0.02	—									
3. Education	2.41	0.88	0.05	−0.10*	—								
4. Income	2.37	0.93	−0.13*	0.12*	0.27***	—							
5. Optimism	3.37	0.66	−0.02	−0.04	0.05	0.004	—						
6. Self-efficacy	3.54	0.65	−0.12*	0.12*	0.03	0.07	0.13**	—					
7. Machiavellianism	3.19	0.58	−0.13*	−0.10	0.04	0.05	0.13*	0.12*	—				
8. Narcissism	2.80	0.66	−0.18***	−0.19***	0.11*	0.08	0.14**	0.16**	0.46***	—			
9. Psychopathy	2.42	0.65	−0.18***	−0.14**	0.003	−0.02	0.12*	0.07	0.34**	0.63***	—		
10. BIGL	4.14	1.24	−0.16**	0.04	0.06	0.10*	0.12*	0.10*	0.26***	0.33***	0.17**	—	
11. BOI	3.90	1.45	−0.13*	−0.07	0.04	0.07	0.16**	0.10	0.34***	0.20***	0.19***	0.33***	—

BIGL, Belief in good luck; BOI, Bribe-offering intention. Gender was dummy coded as 0 = male and 1 = female. Education was coded as 1 = high school education or less, 2 = college degree, 3 = bachelor degree and 4 = postgraduate degree. Monthly income (CNY) was coded as 1 = under 2000, 2 = 2001–5000, 3 = 5001–8000, 4 = 8001–20,000 and 5 = above 20,000. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

**TABLE 2 | Test the mediation effect of Belief in good luck on the link between Machiavellianism and Bribe-offering intention (*N* = 395).**

Predictors	Belief in good luck			Bribe-offering intention		
	$\beta$	<i>t</i>	95%CI	$\beta$	<i>t</i>	95%CI
Gender	-0.12*	-2.48	[-0.22, -0.03]	-0.04	-0.95	[-0.14, 0.05]
Age	0.06	1.15	[-0.04, 0.15]	-0.06	-1.30	[-0.15, 0.03]
Education	0.05	0.90	[-0.05, 0.15]	-0.01	-0.11	[-0.10, 0.09]
Income	0.05	1.06	[-0.05, 0.15]	0.03	0.59	[-0.07, 0.12]
Optimism	0.09	1.74	[-0.01, 0.18]	0.09	1.87	[-0.005, 0.18]
Self-efficacy	0.04	0.72	[-0.06, 0.13]	0.03	0.69	[-0.06, 0.13]
Machiavellianism	0.23***	4.60	[0.13, 0.33]	0.25***	5.11	[0.15, 0.34]
Belief in good luck				0.25***	5.16	[0.15, 0.34]
<i>R</i> <sup>2</sup>		0.10			0.20	
<i>F</i>		6.39***			11.69***	

Each column set is a regression equation that predicts the criterion at the top of the column. \**p* < 0.05; \*\**p* < 0.001.

**TABLE 3 | Test the mediation effect of Belief in good luck on the link between Narcissism and Bribe-offering intention (*N* = 395).**

Predictors	Belief in good luck			Bribe-offering intention		
	$\beta$	<i>t</i>	95%CI	$\beta$	<i>t</i>	95%CI
Gender	-0.10*	-2.03	[-0.19, -0.003]	-0.06	-1.18	[-0.15, 0.04]
Age	0.10	1.94	[-0.002, 0.19]	-0.08	-1.57	[-0.18, 0.02]
Education	0.03	0.56	[-0.07, 0.13]	-0.01	-0.13	[-0.10, 0.09]
Income	0.05	0.90	[-0.05, 0.14]	0.03	0.63	[-0.07, 0.13]
Optimism	0.08	1.60	[-0.02, 0.17]	0.10*	2.15	[0.01, 0.20]
Self-efficacy	0.02	0.30	[-0.08, 0.11]	0.05	0.96	[-0.05, 0.14]
Narcissism	0.31***	6.16	[0.21, 0.41]	0.06	1.09	[-0.05, 0.16]
Belief in good luck				0.29***	5.69	[0.19, 0.39]
<i>R</i> <sup>2</sup>		0.14			0.14	
<i>F</i>		8.92***			8.05***	

Each column set is a regression equation that predicts the criterion at the top of the column. \**p* < 0.05; \*\**p* < 0.001.

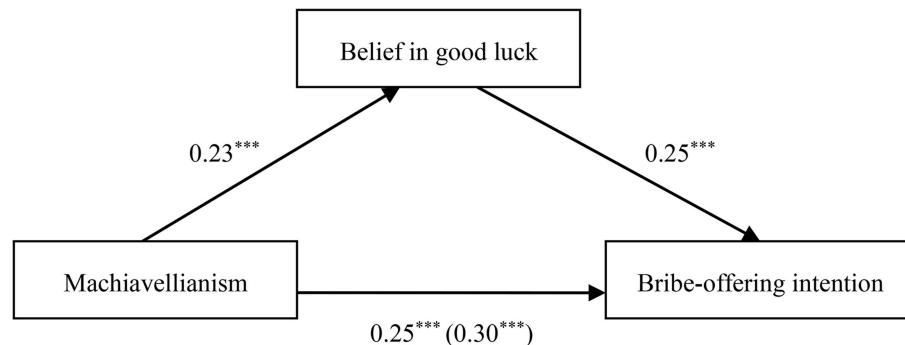
**TABLE 4 | Test the mediation effect of Belief in good luck on the link between Psychopathy and Bribe-offering intention (*N* = 395).**

Predictors	Belief in good luck			Bribe-offering intention		
	$\beta$	<i>t</i>	95%CI	$\beta$	<i>t</i>	95%CI
Gender	-0.12*	-2.41	[-0.22, -0.02]	-0.05	-0.97	[-0.14, 0.05]
Age	0.05	1.04	[-0.05, 0.15]	-0.07	-1.53	[-0.17, 0.02]
Education	0.05	0.97	[-0.05, 0.15]	-0.002	-0.05	[-0.10, 0.10]
Income	0.07	1.26	[-0.04, 0.17]	0.04	0.75	[-0.06, 0.14]
Optimism	0.10	1.91	[-0.003, 0.19]	0.10*	2.06	[0.004, 0.19]
Self-efficacy	0.05	1.02	[-0.05, 0.15]	0.05	1.02	[-0.05, 0.14]
Psychopathy	0.14**	2.85	[0.04, 0.24]	0.11*	2.15	[0.01, 0.20]
Belief in good luck				0.29***	5.95	[0.19, 0.39]
<i>R</i> <sup>2</sup>		0.07			0.15	
<i>F</i>		4.41***			8.56***	

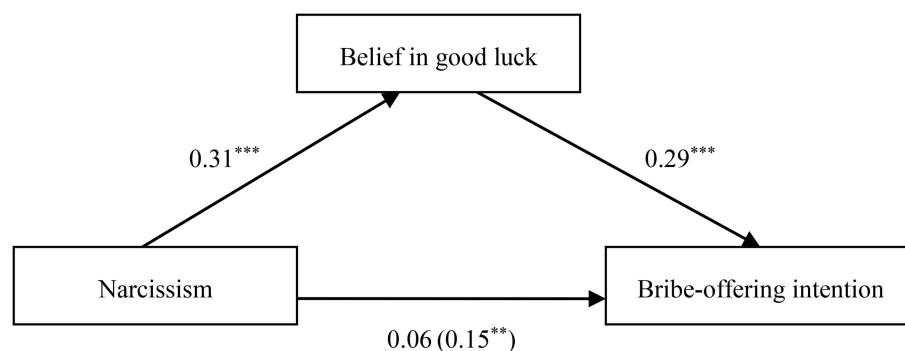
Each column set is a regression equation that predicts the criterion at the top of the column. \**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001.

while the effects of Machiavellianism and psychopathy on bribe-offering intention were both partially mediated by belief in good luck. Individuals with Dark Triad tended to be driven by personal

goals and interests (Jonason and Webster, 2012), even at a cost to other people. They held good luck beliefs and overestimated their chances of seeking unfair advantages via bribe-offering, and may



**FIGURE 1 |** Indirect effect of belief in good luck on the link between Machiavellianism and bribe-offering intention. \*\*\* $p < 0.001$ .



**FIGURE 2 |** Indirect effect of belief in good luck on the link between narcissism and bribe-offering intention. \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

judge that one cannot succeed in the competition without bribe-offering and that the benefit of winning the competition clearly outweighs the cost of bribe-offering.

Study 1 mainly focused on one mechanism related to belief in good luck, that is, people with the Dark Triad traits tended to overestimate the probability of positive-valence outcomes. However, it was unclear how they would react when faced with the negative-valence outcomes of corruption, such as penalty. To further explore the mediating role of belief in good luck, we conducted Study 2.

## STUDY 2

In Study 2, we used a different context and sample to examine the association between each of the Dark Triad traits and bribe-taking intention, and examined the mediating role of belief in good luck in this relationship. We speculated that individuals with Dark Triad tend to engage in bribe-taking behaviors partially because they hold the lucky belief and irrationally underestimate the likelihood of being penalized for bribe-taking.

## Methods

### Participants

A total of 386 Chinese adults were recruited online, via the Qualtrics Survey from different enterprises in China. The final

valid sample size was 382 Chinese adults (193 female and 189 male;  $M_{age} = 28.19$  years,  $SD = 5.66$  years; age range: 18–63 years), as 4 adults were excluded because they failed to complete the questionnaires. The effective response rate was 98.96%. Participants varied considerably in terms of their education levels (12.8% with high school education or less, 39.3% with a college degree, 43.5% with a bachelor degree, and 4.5% with a postgraduate degree) and monthly income (8.4% with less than 2000 yuan, 48.4% with 2001–5000 yuan, 30.9% with 5001–8000 yuan, 9.4% with 8001–20,000 yuan, and 2.9% with more than 20,000 yuan).

### Procedure

The study procedure was the same as that employed in Study 1.

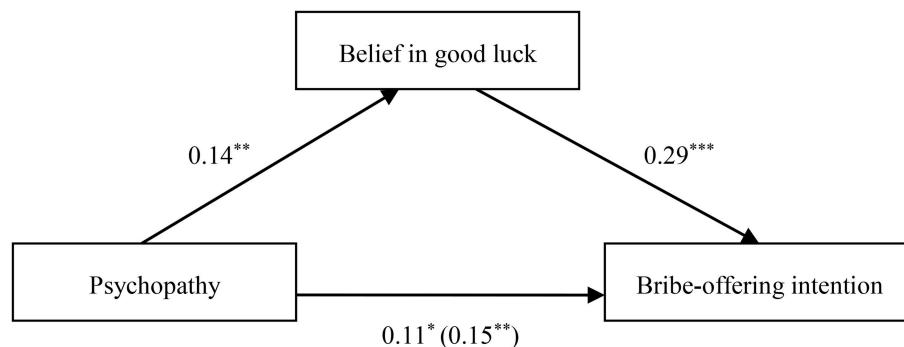
### Measures

#### *The short Dark Triad (SD3)*

The 27-item Dark Triad was used, as in Study 1. In this study, the Cronbach's  $\alpha$  for the total scale was 0.87, and that for Machiavellianism, narcissism, and psychopathy was 0.78, 0.81, and 0.77, respectively.

#### *The bribe-taking scenario*

We adapted bribery scenarios successfully used in past research (Bai et al., 2014). Participants were exposed to three hypothetical daily life scenarios about bribe-taking (see Supplementary



**FIGURE 3 | Indirect effect of belief in good luck on the link between psychopathy and bribe-offering intention.** \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

Material). The experimental procedure was the same as that employed in Study 1. The following is a sample of the bribe-taking scenario:

“Suppose you are a director who is in charge of bidding. Compared to other bidders, Company A is in an unfavorable position in the competition. In order to win the bid, the CEO of Company A asks you to help him, and also privately promises you a certain sum of money if his company wins the bid. If you help him, the probability that he will win the bid will be greatly improved. But you are aware that it is against the law to help him win the bid by accepting a bribe.”

After each scenario, propensity to engage in bribe-taking was measured by “Please estimate the likelihood that you would offer the help to Company A” on a 7-point Likert scale (1 = *extremely unlikely*, 7 = *extremely likely*). The index of bribe-taking intention was calculated as the average score of the three scenarios, where higher scores are indicative of greater bribe-taking intention. The Cronbach’s  $\alpha$  of the tool was 0.85.

#### Belief in good luck

The research paradigm of “objective probability event-subjective probability estimation” was used, same as in Study 1. After each bribe-taking scenario, there was a corresponding scenario to measure one’s belief in good luck in penalty-avoidance. The following is a sample scenario that corresponds to the example illustrated in the bribe-taking scenario:

“Suppose you have accepted the bribe that the CEO of Company A offered to you and helped him to win the bid. According to the statistics of the national department related to this scenario, the probability of penalty for bribe-taking in bidding is almost 40% in recent years. Please respond to the following two items: (1) Despite the high probability, however, you still definitely believe that you will have a good luck to avoid penalty; (2) The good luck makes you believe that your probability of being penalized will be significant lower than 40% even if you take the bribe.”

Participants responded to each item on a 7-point Likert scale (1 = *strongly disagree*, 7 = *strongly agree*). The average of three scenarios comprised the score of belief in good luck. The higher

the averaged score, the higher is the level of belief in good luck. The Cronbach’s  $\alpha$  of this tool was 0.89. Additionally, we should note that though China is currently ramping up efforts to fight corruption, the results of a pilot investigation show that the penalty rate in corruption cases is just about 10%, and the 40% rate that was presented in the scenarios used in this study is a very high penalty rate.

#### Control variables

We controlled the same variables as in Study 1. The Cronbach’s  $\alpha$  of the 10-item Life orientation test revised measure and the new general self-efficacy scale was 0.81, and 0.88, respectively.

## Results

### Discriminate Validity

The procedure to test the discriminant validity of belief in good luck was the same as that employed in Study 1. Results demonstrated that the three-factor model was better fit the data [ $\chi^2_{(116, N = 382)} = 335.88, p < 0.001$ , GFI = 0.90, CFI = 0.92, RMSEA = 0.07] than the single-factor model [ $\chi^2_{(3, N = 382)} = 1188.22, p < 0.001$ ], the two-factor model that combined belief in good luck and dispositional optimism into one factor [ $\Delta\chi^2_{(2, N = 382)} = 636.56, p < 0.001$ ], and the two-factor model that combined belief in good luck and self-efficacy into one factor [ $\Delta\chi^2_{(2, N = 395)} = 645.50, p < 0.001$ ].

### Descriptive Statistics

**Table 5** displays the descriptive statistics and zero-order correlation coefficients among the variables. As expected, narcissism, and psychopathy were positively correlated with belief in good luck and bribe-taking intention. Interestingly, Machiavellianism was positively related to bribe-taking intention, but was not significantly linked to the corresponding belief in good luck.

### Testing the Mediating Role of Belief in Good Luck in the Relationship between the Dark Triad Traits and Bribe-Taking Intention

While controlling the control variables, Machiavellianism ( $\beta = 0.24, p < 0.001$ , 95% CI [0.15, 0.34]), narcissism ( $\beta = 0.13, p < 0.05$ , 95% CI [0.03, 0.25]), and psychopathy ( $\beta = 0.25, p <$

**TABLE 5 | Descriptive statistics and correlations among the variables.**

Variables	M	SD	1	2	3	4	5	6	7	8	9	10	11
1. Gender	0.51	0.50	—										
2. Age	28.19	5.66	-0.04	—									
3. Education	2.40	0.77	-0.12*	0.05	—								
4. Income	2.50	0.88	-0.27***	0.04	0.30***	—							
5. Optimism	3.36	0.73	-0.04	-0.04	0.07	-0.06	—						
6. Self-efficacy	3.51	0.69	-0.04	0.05	0.05	0.12*	0.11*	—					
7. Machiavellianism	3.20	0.67	-0.17**	-0.08	0.11*	0.02	0.14**	0.14**	—				
8. Narcissism	2.90	0.69	-0.20**	-0.05	0.16**	0.07	0.13**	0.14**	0.44***	—			
9. Psychopathy	2.46	0.67	-0.18***	-0.11*	0.11*	0.04	0.15**	0.04	0.37***	0.54***	—		
10. BIGL	3.56	1.72	-0.18***	0.05	0.11*	0.13*	0.12*	0.10*	0.09	0.26***	0.18***	—	
11. BTI	3.09	1.56	-0.18***	-0.18**	0.01	0.08	0.14**	0.04	0.28***	0.18***	0.30***	0.32***	—

BIGL, Belief in good luck; BTI, Bribe-taking intention. Gender was dummy coded as 0 = male and 1 = female. Education was coded as 1 = high school education or less, 2 = college degree, 3 = bachelor degree and 4 = postgraduate degree. Monthly income (CNY) was coded as 1 = under 2000; 2 = 2001–5000; 3 = 5001–8000; 4 = 8001–20,000; and 5 = above 20,000. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

0.001, 95% CI [0.15, 0.34]) was positively predicted bribe-taking intention. Thus, Hypothesis 1 was verified again.

We then examined whether belief in good luck in penalty-avoidance mediated the effect of each Dark Triad trait on bribe-taking intention. Similar to that in Study 1, the Model 4 of the Hayes' PROCESS macro for SPSS was adopted (Hayes, 2013). The results have been illustrated in Tables 6–8 and Figures 4–6.

When Machiavellianism was the independent variable, the results showed that belief in good luck in penalty-avoidance did not mediate the relationship between Machiavellianism and bribe-taking intention ( $\beta_{\text{indirect}} = 0.01$ ,  $SE = 0.02$ , 95% CI [−0.02, 0.05]), as depicted in Table 6 and Figure 4.

When narcissism was the independent variable, the relationship between narcissism and bribe-taking intention was fully mediated by belief in good luck in penalty-avoidance ( $\beta_{\text{indirect}} = 0.06$ ,  $SE = 0.02$ , 95% CI [0.03, 0.11]), as depicted in Table 7 and Figure 5.

When psychopathy was the independent variable, the results demonstrated that belief in good luck in penalty-avoidance partially mediated the relationship between psychopathy and bribe-taking intention ( $\beta_{\text{indirect}} = 0.04$ ,  $SE = 0.02$ , 95% CI [0.01, 0.07]), as depicted in Table 8 and Figure 6.

## Discussion

These results reconfirmed Hypothesis 1 and showed that the Dark Triad traits were positively associated with corruption. Additionally, Hypothesis 2 was partially verified. The effect of narcissism on bribe-taking intention was fully mediated by belief in good luck, whereas the effect of psychopathy was partially mediated by belief in good luck. This indicates that “dark” individuals’ exaggerated beliefs in good luck may engender a false sense of control (Darke and Freedman, 1997a,b), which may cause them to ignore the real risks of bribe-taking and underestimate the odds of an unfavorable consequence. Interestingly, however, belief in good luck in penalty-avoidance did not mediate the relationship between Machiavellianism and bribe-taking intention. The

calculated strategy of Machiavellianism may help us explain this result.

## GENERAL DISCUSSION

It is important to note that very few studies have examined the mediating role of irrational beliefs in good luck between the Dark Triad of personality and corruption. Our results indicated that the three Dark Triad traits significantly contributed to explaining the variance in corrupt intention. More importantly, the mediating effects suggest that people with Dark Triad are more likely to engage in corruption, partially due to their belief in good luck. In other words, they tend to overestimate the likelihood of seeking gains via bribe-offering and underestimate the likelihood of being penalty for bribe-taking irrationally.

Consistent with the theory of planned behavior (Fishbein and Ajzen, 1975; Ajzen, 1991), the Dark Triad of personality predicted the irrational behavioral beliefs in good luck, which in turn, affected one’s corrupt intention. While the Chinese central government has begun prioritizing anti-corruption work and has drastically intensified anti-corruption campaigns, engagement in corruption, especially bribe-taking, has become increasingly risky (Gong, 2002). In Chinese legal sanctions of corruption, engaging in bribe-taking will be penalized severely, whereas bribe-offering may be punished leniently or may even be exempted from investigation (Lu and Gunnison, 2003; Wang and Wu, 2012). Additionally, in the daily social norms of China, people often adopt double standards toward bribe-taking and bribe-offering behaviors. For example, after bribery cases are exposed, people often express their condemnation toward bribe recipients, but show less negativity toward bribe payers. Accordingly, in Study 1, in order to gain an unfair advantage, people with a dark personality were found to tend to overestimate the probability of seeking gains via bribe-offering, which led them to have a tendency to engage in corruption. In Study 2, when faced with a severe penalty for bribe-taking in China, people with

**TABLE 6 | Test the mediation effect of Belief in good luck on the link between Machiavellianism and Bribe-taking intention (*N* = 382).**

Predictors	Belief in good luck			Bribe-taking intention		
	$\beta$	<i>t</i>	95%CI	$\beta$	<i>t</i>	95%CI
Gender	-0.14**	-2.63	[-0.24, -0.03]	-0.09	-1.89	[-0.19, 0.003]
Age	0.04	0.82	[-0.06, 0.14]	-0.17***	-3.61	[-0.26, -0.08]
Education	0.06	1.06	[-0.05, 0.16]	-0.07	-1.51	[-0.17, 0.02]
Income	0.07	1.21	[-0.04, 0.17]	0.05	1.03	[-0.05, 0.15]
Optimism	0.10*	2.01	[0.002, 0.20]	0.07	1.49	[-0.02, 0.16]
Self-efficacy	0.07	1.32	[-0.03, 0.17]	-0.02	-0.50	[-0.12, 0.07]
Machiavellianism	0.04	0.80	[-0.06, 0.14]	0.23***	4.82	[0.14, 0.32]
Belief in good luck				0.28***	5.93	[0.19, 0.37]
<i>R</i> <sup>2</sup>		0.06			0.21	
<i>F</i>		3.67***			12.62***	

Each column set is a regression equation that predicts the criterion at the top of the column. \**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001.

**TABLE 7 | Test the mediation effect of Belief in good luck on the link between Narcissism and Bribe-taking intention (*N* = 382).**

Predictors	Belief in good luck			Bribe-taking intention		
	$\beta$	<i>t</i>	95%CI	$\beta$	<i>t</i>	95%CI
Gender	-0.11*	-2.03	[-0.21, -0.004]	-0.12*	-2.33	[-0.22, -0.02]
Age	0.05	1.03	[-0.05, 0.15]	-0.18***	-3.84	[-0.28, -0.09]
Education	0.03	0.62	[-0.07, 0.13]	-0.06	-1.18	[-0.16, 0.04]
Income	0.07	1.29	[-0.04, 0.17]	0.04	0.79	[-0.06, 0.14]
Optimism	0.09	1.72	[-0.01, 0.18]	0.09	1.80	[-0.01, 0.18]
Self-efficacy	0.05	0.95	[-0.05, 0.15]	-0.002	-0.05	[-0.10, 0.09]
Narcissism	0.21***	4.16	[0.11, 0.31]	0.08	1.48	[-0.02, 0.17]
Belief in good luck				0.28***	5.52	[0.18, 0.37]
<i>R</i> <sup>2</sup>		0.10			0.17	
<i>F</i>		6.21***			9.48***	

Each column set is a regression equation that predicts the criterion at the top of the column. \**p* < 0.05; \*\*\**p* < 0.001.

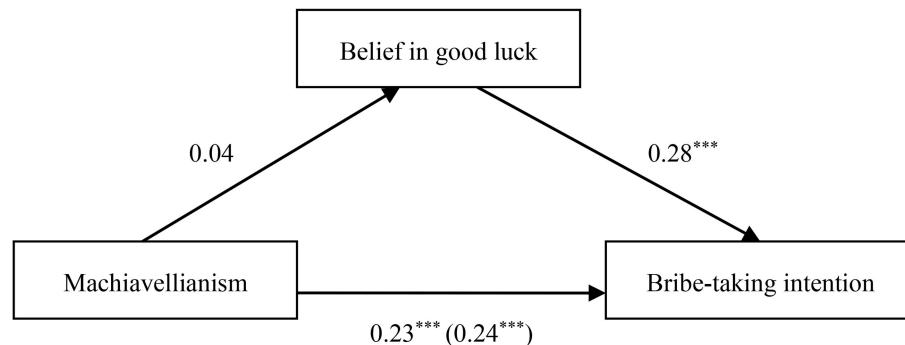
**TABLE 8 | Test the mediation effect of Belief in good luck on the link between Psychopathy and Bribe-taking intention (*N* = 382).**

Predictors	Belief in good luck			Bribe-taking intention		
	$\beta$	<i>t</i>	95%CI	$\beta$	<i>t</i>	95%CI
Gender	-0.12*	-2.31	[-0.22, -0.02]	-0.10	-1.95	[-0.19, 0.001]
Age	0.05	1.07	[-0.04, 0.15]	-0.16***	-3.46	[-0.25, -0.07]
Education	0.05	0.90	[-0.06, 0.15]	-0.07	-1.39	[-0.16, 0.03]
Income	0.07	1.24	[-0.04, 0.17]	0.04	0.88	[-0.06, 0.14]
Optimism	0.09	1.77	[-0.01, 0.19]	0.07	1.47	[-0.02, 0.16]
Self-efficacy	0.07	1.38	[-0.03, 0.17]	0.002	0.97	[-0.09, 0.09]
Psychopathy	0.14**	2.65	[0.04, 0.24]	0.21***	4.33	[0.11, 0.31]
Belief in good luck				0.26***	5.45	[0.17, 0.36]
<i>R</i> <sup>2</sup>		0.08			0.20	
<i>F</i>		4.64***			11.95***	

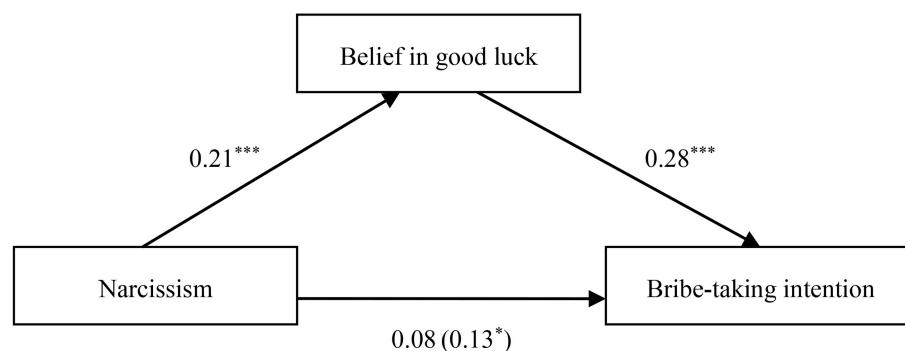
Each column set is a regression equation that predicts the criterion at the top of the column. \**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001.

high narcissism and psychopathy tended to underestimate or even ignore the penalty odds, which drove them to engage in corruption. However, the strategic nature of Machiavellians

protects them from generating an irrational belief in good luck in avoiding penalty, such that they only accept bribes when minimal or no threat of penalty exists.



**FIGURE 4 |** Indirect effect of belief in good luck on the link between Machiavellianism and bribe-taking intention. \*\*\* $p < 0.001$ .

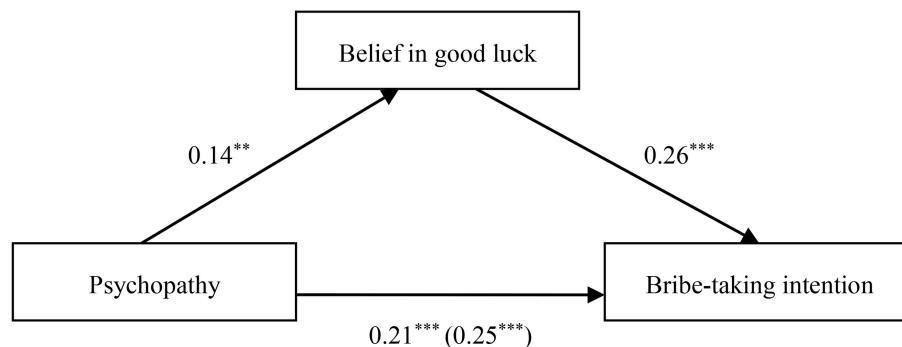


**FIGURE 5 |** Indirect effect of belief in good luck on the link between narcissism and bribe-taking intention. \* $p < 0.05$ ; \*\*\* $p < 0.001$ .

In line with Hypothesis 1, Machiavellianism positively predicts corruption. Machiavellians employ manipulative, exploitive, and devious methods to achieve private goals and make unethical choices if chances for benefit emerge (Gunnthorsdottir et al., 2002; Birkas et al., 2015). However, one inconsistency in the results is that, in Study 1, Machiavellianism was positively related to belief in good luck in seeking gains, whereas, in Study 2, Machiavellianism was less correlated with belief in good luck in avoiding penalty. In hindsight, these results appear to be consistent with previous theories presented in the Machiavellianism literature. Specifically, recent empirical evidence has shown that Machiavellians were sensitive to rewards (Birkas et al., 2015), likely making reward-oriented decisions, and thus overrating the benefits and probability of gaining an unfair advantage derived from bribe-offering, positively linking to their irrational belief in good luck, which, in part, fuels their tendency to engage in corruption. Perhaps, not surprisingly, Machiavellians only accept bribes when there is maximal benefit with minimal punishment (Jones, 2013). That is, they strategize to maximize their long-term gains (Jones, 2013), and only involve in some cautious misbehaviors. However, since bribe-taking behavior attracts a severe legal penalty in China (Lu and Gunnison, 2003), it may bring in minor benefits in a short period of time but at the expense of significant costs in the long run. With stronger detecting and evaluating abilities, Machiavellians

will carefully estimate the potential risk to their own interests (Birkas et al., 2015), and thus may not engender the irrational beliefs in good luck in avoiding penalty. These results are exactly in line with previous evidence that Machiavellianism is associated with anti-social behaviors only when there is no or little risk of being caught (Jones, 2013). Accordingly, Machiavellianism has little correlation with an irrational belief in underestimating the probability of penalty for bribe-taking. These findings about Machiavellianism may be an interesting area for future research.

Obviously, it is reasonable to argue that individuals with high narcissism tend to engage in corruption because they believe in good luck. Our results are in accordance with previous research that narcissists who believe in good luck are overconfident (Darke and Freedman, 1997b; Jones, 2013), which causes them to exhibit cognitive biases in success or penalty perceptions about corruption (Chatterjee and Hambrick, 2007; Lakey et al., 2008). In addition, narcissists possess an overly positive self-concept (Lakey et al., 2008), leading them to acquire a control illusion such that they believe that they could control their corrupt actions (Farwell and Wohlwend-Lloyd, 1998; Jones, 2013). This would further exacerbate their irrational beliefs in good luck. Research has shown that perceived behavioral control influence one's corrupt intention (Rabl and Kuhlmann, 2008). If narcissists believe that they will have a good luck and can control the whole corrupt event, they would self-aggrandize the likelihood



**FIGURE 6 | Indirect effect of belief in good luck on the link between psychopathy and bribe-taking intention.** \*\* $p < 0.05$ ; \*\*\* $p < 0.001$ .

of success and downplay the likelihood of penalty. Even when faced with opposite facts, it seems that they still hold the illusory belief that things will go as they wish. Thus, in support of our Hypothesis 2, narcissists hold good luck belief and have a tendency to overestimate their chances of winning via bribe-offering, even though the chance is very low; and underestimate the probability of being penalized for bribe-taking, even though the odds are very high; which would drive them to engage in corrupt behaviors.

Additionally, as predicted, people with high psychopathy tend to engage in corruption. These results furnished preliminary evidence that the erratic antisocial and reckless nature of psychopathy easily lead them to engage in corrupt behaviors (Jones, 2013). Psychopathic individuals with lower self-control (Tangney et al., 2004) cannot resist the temptation of corruption. Lured by potential gains, such individuals seem to be willing to involve in corruption. In addition, mediational data demonstrate that the effect of psychopathy on corrupt intention is partially explained by an irrational belief in good luck. Individuals high in psychopathy cannot regulate impulses effectively and easily create irrational beliefs in good luck in seeking gains or avoiding penalty, and even view gains or penalty as merely a by-product of corruption. These findings confirm our hypothesis that individuals' psychopathy can positively influence their irrational beliefs in good luck, which, in turn, partially leads to a higher corrupt intention.

## Implications

Our research brings significant theoretical implications for the literature on corruption. To our knowledge, this study is among the first attempts to examine the impact of the Dark Triad traits on corruption and on the mediating role of belief in good luck. Firstly, this study extends the preliminary research on corruption from the perspective of individual differences and confirms the relationship between each Dark Triad trait and corruption. Rampant corruption events have raised questions surrounding the personality traits responsible for corruption. In other words, do certain personality traits facilitate corruption? To a certain extent, the current study seems to have answered this question by revealing that

people with high Dark Triad of personality are more easily engage in corruption. Secondly, the present study furthered the research on the theory of planned behavior, and encourages researchers to understand the occurrence of corruption by providing insight into the underlying psychological mechanisms between the Dark Triad of personality and corruption. The mediating role of belief in good luck helps reveal the reason why the Dark Triad traits facilitate corruption. Thirdly, these findings also enrich the studies on the prospect theory and establish that people's decisions about the negative or deviant events are not independent of outcome valences once the probabilities are specified. In the present study, we redefined the concept of belief in good luck by using an adapted research paradigm of "objective probability event-subjective probability estimation," and confirmed its two mechanisms by examining the two different outcome valences of corruption forms.

It is noteworthy that this study was also pragmatic because it provided some anti-corruption measures. First, at the individual level, although one's personality cannot be easily changed, if individuals could be aware that their personality predisposes them to generate unrealistic beliefs in good luck and to engage in corrupt behaviors, then they could take more positive steps to deter them. Second, at the government level, the knowledge that the Dark Triad and irrational beliefs in good luck are associated with corrupt intention can help anti-corruption agencies and institutions become more effective in their actions of restraining corruption. In view of the mediating role of luck beliefs, anti-corruption policies should focus on inhabiting people's irrational beliefs in corruption. For those with high Dark Triad tendencies, the government can decrease their corrupt behaviors by discouraging their irrational belief in good luck in corruption through some ways. For example, creating a fair competition environment is the permanent solution to reduce the necessity of offering bribe and to decrease the probability of gaining profits via bribe-offering. This will encourage people to win competitions or seek benefits through appropriate means. In addition, extensively exposing people to the information about anti-corruption policies (Song and Cheng, 2012), such as the severe penalty

policy, can help them be acutely aware of the huge cost of engaging in corruption. That is, both bribe recipients and payers should be penalized heavily. Therefore, stepping up penalties for bribe-offering is also imperative for curbing corruption in China.

## Limitations and Prospects

There is no doubt that this study has several limitations. First of all, the cross-sectional data and correlational design does not allow us to detect the causal link between the Dark Triad of personality, belief in good luck, and corruption. Longitudinal studies should be conducted to replicate these findings in future. Second, it is also important to note that although self-report measures are widely used and the instruments employed in present study have good reliability, a response bias and common-method bias are still inevitable. Third, measures of corruption were based on the hypothetical scenarios, which may not reflect actual corrupt behaviors. Thus, the ecological validity of the assessment may be affected. Future research can develop alternative tools to bring research results closer to actual behaviors. Fourth, we should point out that our conclusions are based on the mediation model that we examined with each of the three Dark Triad traits as a predictor variable, belief in good luck as a mediator variable, and corrupt intention as a dependent variable. Although, our results supported the causal processes proposed in the hypothesis development sections, this study did not test the competing models, and therefore, the alternative models and alternative mediators (e.g., moral disengagement) need to be identified in future studies. Finally, the moderator variables between the Dark Triad of personality and corruption also need to be investigated to uncover the boundary conditions, which may help us understand the extent to which the Dark Triad of personality increases corruption.

## REFERENCES

- Ajzen, I. (1991). The theory of planned behavior. *Organ. Behav. Hum. Decis. Process* 50, 179–211. doi: 10.1016/0749-5978(91)90020-T
- Alesina, A., and Angeletos, G. M. (2005). Corruption, inequality, and fairness. *J. Monet. Econ.* 52, 1227–1244. doi: 10.1016/j.jmoneco.2005.05.003
- Andre, N. (2006). Good fortune, luck, opportunity and their lack: how do agents perceive them? *Pers. Individ. Dif.* 40, 1461–1472. doi: 10.1016/j.paid.2005.10.022
- Azizli, N., Atkinson, B. E., Baughman, H. M., Chin, K., Vernon, P. A., Harris, E., et al. (2016). Lies and crimes: Dark Triad, misconduct, and high-stakes deception. *Pers. Individ. Dif.* 89, 34–39. doi: 10.1016/j.paid.2015.09.034
- Bai, B. Y., Liu, X. X., and Kou, Y. (2014). Belief in a just world lowers perceived intention of corruption: the mediating role of perceived punishment. *PLoS ONE* 9:e97075. doi: 10.1371/journal.pone.0097075
- Birkas, B., Csatho, A., Gacs, B., and Bereczkei, M. (2015). Nothing ventured nothing gained: strong associations between reward sensitivity and two measures of Machiavellianism. *Pers. Individ. Dif.* 74, 112–115. doi: 10.1016/j.paid.2014.09.046
- Blackburn, K., and Forgues-Puccio, G. F. (2009). Why is corruption less harmful in some countries than in others? *J. Econ. Behav. Organ.* 72, 797–810. doi: 10.1016/j.jebo.2009.08.009
- Buelow, M. T., and Brunell, A. B. (2014). Facets of grandiose narcissism predict involvement in health-risk behaviors. *Pers. Individ. Dif.* 69, 193–198. doi: 10.1016/j.paid.2014.05.031
- Cábelková, I., and Hanousek, J. (2004). The power of negative thinking: corruption, perception and willingness to bribe in Ukraine. *Appl. Econ.* 36, 383–397. doi: 10.1080/00036840410001674303
- Campbell, W. K., Goodie, A. S., and Foster, J. D. (2004). Narcissism, confidence, and risk attitude. *J. Behav. Decis. Mak.* 17, 297–311. doi: 10.1002/bdm.475
- Chatterjee, A., and Hambrick, D. C. (2007). It's all about me: narcissistic chief executive officers and their effects on company strategy and performance. *Adm. Sci. Q.* 52, 351–386. doi: 10.2189/asqu.52.3.351
- Chen, G., Gully, S. M., and Eden, D. (2001). Validation of a new general self-efficacy scale. *Organ. Res. Methods* 4, 62–83. doi: 10.1177/109442810141004
- Chiu, J., and Storm, L. (2010). Personality, perceived luck and gambling attitudes as predictors of gambling involvement. *J. Gambl. Stud.* 26, 205–227. doi: 10.1007/s10899-009-9160-x
- Crysel, L. C., Crosier, B. S., and Webster, G. D. (2013). The Dark Triad and risk behavior. *Pers. Individ. Dif.* 54, 35–40. doi: 10.1016/j.paid.2012.07.029
- Damisch, L., Stoberock, B., and Mussweiler, T. (2010). Keep your fingers crossed! How superstition improves performance. *Psychol. Sci.* 21, 1014–1020. doi: 10.1177/0956797610372631
- Darke, P. R., and Freedman, J. L. (1997a). The belief in good luck scale. *J. Res. Pers.* 31, 486–511. doi: 10.1006/jrpe.1997.2197

## CONCLUSIONS

This study contributes to the emerging literature concerning the occurrence of corruption from the perspective of individual factors, and the findings hold substantive implications, both theoretical and practical. Using some hypothetical scenarios of corruption in the Chinese context, the two sub-studies in the current study not only present evidence that people with high Dark Triad tendencies are more likely to engage in corruption, but also support the role of their irrational beliefs in good luck as a mediator in this association. We hope that this study can provide some new insights and offer a valuable foundation for the future research on corruption.

## AUTHOR CONTRIBUTIONS

To conception and design: HZ, HZ, and YX. Collection, analysis and interpretation of data: HZ, HZ. Drafting the article: HZ. Revising the article critically: HZ, HZ, and YX.

## ACKNOWLEDGMENTS

This research was funded by grants from the National Key Technologies R&D Program of China (2012BAI36B03). We want to thank Nikolaos Georgantzis and two anonymous reviewers for comments on earlier versions of this manuscript; and a special thank you goes to Zhujiang Zhang for help in the language editing process.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00608>

- Darke, P. R., and Freedman, J. L. (1997b). Lucky events and beliefs in luck: Paradoxical effects on confidence and risk-taking. *Person. Soc. Psychol. Bull.* 23, 378–388. doi: 10.1177/0146167297234004
- Day, L., and Maltby, J. (2003). Belief in good luck and psychological well-being: the mediating role of optimism and irrational beliefs. *J. Psychol.* 137, 99–110. doi: 10.1080/00223980309600602
- Djeriouat, H., and Tremoliere, B. (2014). The Dark Triad of personality and utilitarian moral judgment: the mediating role of Honesty/Humility and Harm/Care. *Pers. Individ. Dif.* 67, 11–16. doi: 10.1016/j.paid.2013.12.026
- Donchev, D., and Ujhelyi, G. (2007). *Do Corruption Indices Measure Corruption? Unpublished Manuscript*. Cambridge, MA: Harvard University.
- Egan, V., Hughes, N., and Palmer, E. J. (2015). Moral disengagement, the dark triad, and unethical consumer attitudes. *Pers. Individ. Dif.* 76, 123–128. doi: 10.1016/j.paid.2014.11.054
- Ellis, A. (1999). Early theories and practices of rational emotive behavior therapy and how they have been augmented and revised during the last three decades. *J. Rational Emot. Cogn. Behav. Ther.* 17, 69–93. doi: 10.1023/A:1023048830350
- Farwell, L., and Wohlwend-Lloyd, R. (1998). Narcissistic processes: optimistic expectations, favorable self-evaluations, and self-enhancing attributions. *J. Pers.* 66, 65–83. doi: 10.1111/1467-6494.00003
- Fishbein, M., and Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.
- Foster, J. D., Misra, T. A., and Reidy, D. E. (2009). Narcissists are approach-oriented toward their money and their friends. *J. Res. Pers.* 43, 764–769. doi: 10.1016/j.jrp.2009.05.005
- Gong, T. (2002). Dangerous collusion: corruption as a collective venture in contemporary China. *Commun. Post Commun. Stud.* 35, 85–103. doi: 10.1016/S0967-067x(01)00026-5
- Gunnthorsdottir, A., McCabe, K., and Smith, V. (2002). Using the Machiavellianism instrument to predict trustworthiness in a bargaining game. *J. Econ. Psychol.* 23, 49–66. doi: 10.1016/S0167-4870(01)00067-8
- Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford Press.
- He, Z. K. (2000). Corruption and anti-corruption in reform China. *Commun. Post Commun. Stud.* 33, 243–270. doi: 10.1016/S0967-067x(00)00006-4
- Hodson, G., Hogg, S. M., and MacInnis, C. C. (2009). The role of “dark personalities” (narcissism, Machiavellianism, psychopathy), Big Five personality factors, and ideology in explaining prejudice. *J. Res. Pers.* 43, 686–690. doi: 10.1016/j.jrp.2009.02.005
- Jaber-López, T., García-Gallego, A., Perakakis, P., and Georgantzis, N. (2014). Physiological and behavioral patterns of corruption. *Front. Behav. Neurosci.* 8:434. doi: 10.3389/fnbeh.2014.00434
- Jibeen, T. (2015). Personality dimensions and emotional problems: the mediating role of irrational beliefs in Pakistani adult non-clinical sample. *Int. J. Psychol.* 50, 93–100. doi: 10.1002/ijop.12069
- Jonason, P. K., Koenig, B. L., and Tost, J. (2010). Living a fast life: the Dark Triad and life history theory. *Hum. Nat. Interdisc. Biosoc. Perspect.* 21, 428–442. doi: 10.1007/s12110-010-9102-4
- Jonason, P. K., and Webster, G. D. (2012). A protean approach to social influence: Dark Triad personalities and social influence tactics. *Pers. Individ. Dif.* 52, 521–526. doi: 10.1016/j.paid.2011.11.023
- Jonason, P. K., Webster, G. D., Schmitt, D. P., Li, N. P., and Crysel, L. (2012). The antihero in popular culture: life history theory and the Dark Triad personality traits. *Rev. Gen. Psychol.* 16, 192–199. doi: 10.1037/a0027914
- Jones, D. N. (2013). What's mine is mine and what's yours is mine: the Dark Triad and gambling with your neighbor's money. *J. Res. Pers.* 47, 563–571. doi: 10.1016/j.jrp.2013.04.005
- Jones, D. N. (2014). Risk in the face of retribution: psychopathic individuals persist in financial misbehavior among the Dark Triad. *Pers. Individ. Dif.* 67, 109–113. doi: 10.1016/j.paid.2014.01.030
- Jones, D. N., and Figueiredo, A. J. (2013). The core of darkness: uncovering the heart of the Dark Triad. *Eur. J. Pers.* 27, 521–531. doi: 10.1002/Per.1893
- Jones, D. N., and Paulhus, D. L. (2014). Introducing the short Dark Triad (SD3): a brief measure of dark personality traits. *Assessment* 21, 28–41. doi: 10.1177/1073191113514105
- Judge, T. A., LePine, J. A., and Rich, B. L. (2006). Loving yourself abundantly: relationship of the narcissistic personality to self- and other perceptions of workplace deviance, leadership, and task and contextual performance. *J. Appl. Psychol.* 91, 762–776. doi: 10.1037/0021-9010.91.4.762
- Kusev, P., van Schaik, P., Ayton, P., Dent, J., and Chater, N. (2009). Exaggerated risk: prospect theory and probability weighting in risky choice. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1487–1505. doi: 10.1037/a0017039
- Lakey, C. E., Rose, P., Campbell, W. K., and Goodie, A. S. (2008). Probing the link between narcissism and gambling: the mediating role of judgment and decision-making biases. *J. Behav. Decis. Mak.* 21, 113–137. doi: 10.1002/Bdm.582
- Lee, K. B., and Ashton, M. C. (2005). Psychopathy, Machiavellianism, and Narcissism in the Five-Factor Model and the HEXACO model of personality structure. *Pers. Individ. Dif.* 38, 1571–1582. doi: 10.1016/j.paid.2004.09.016
- Lindgreen, A. (2004). Corruption and unethical behavior: report on a set of Danish guidelines. *J. Busin. Eth.* 51, 31–39. doi: 10.1023/B:Busi.0000032388.68389.60
- Linton, D. K., and Power, J. L. (2013). The personality traits of workplace bullies are often shared by their victims: is there a dark side to victims? *Pers. Individ. Dif.* 54, 738–743. doi: 10.1016/j.paid.2012.11.026
- Lu, H., and Gunnison, E. (2003). Power, corruption, and the legal process in China. *Int. Crim. Just. Rev.* 13, 28–49. doi: 10.1177/105756770301300102
- Morf, C. C., Weir, C., and Davidov, M. (2000). Narcissism and intrinsic motivation: the role of goal congruence. *J. Exp. Soc. Psychol.* 36, 424–438. doi: 10.1006/jesp.1999.1421
- O'Boyle, E. H., Forsyth, D. R., Banks, G. C., and McDaniel, M. A. (2012). A meta-analysis of the Dark Triad and work behavior: a social exchange perspective. *J. Appl. Psychol.* 97, 557–579. doi: 10.1037/A0025679
- Pabian, S., De Backer, C. J. S., and Vandebosch, H. (2015). Dark Triad personality traits and adolescent cyber-aggression. *Pers. Individ. Dif.* 75, 41–46. doi: 10.1016/j.paid.2014.11.015
- Paulhus, D. L., and Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *J. Res. Pers.* 36, 556–563. doi: 10.1016/S0092-6566(02)00505-6
- Rabl, T., and Kuhlmann, T. M. (2008). Understanding corruption in organizations—Development and empirical assessment of an action model. *J. Busin. Eth.* 82, 477–495. doi: 10.1007/s10515-008-9898-6
- Rauthmann, J. F., and Kolar, G. P. (2013). Positioning the Dark Triad in the interpersonal circumplex: the friendly-dominant narcissist, hostile-submissive Machiavellian, and hostile-dominant psychopath? *Pers. Individ. Dif.* 54, 622–627. doi: 10.1016/j.paid.2012.11.021
- Roeser, K., McGregor, V. E., Stegmaier, S., Mathew, J., Kübler, A., and Meule, A. (2016). The Dark Triad of personality and unethical behavior at different times of day. *Pers. Individ. Dif.* 88, 73–77. doi: 10.1016/j.paid.2015.09.002
- Samar, S. M., Walton, K. E., and McDermut, W. (2013). Personality traits predict irrational beliefs. *J. Ration. Emot. Cogn. Behav. Ther.* 31, 231–242. doi: 10.1007/s10942-013-0172-1
- Scheier, M. F., Carver, C. S., and Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *J. Pers. Soc. Psychol.* 67, 1063–1078. doi: 10.1037/0022-3514.67.6.1063
- Schmidt, A. A. (2008). *Development and Validation of the Toxic Leadership Scale*. Doctoral dissertation. Available from ProQuest Dissertations and Theses Database. (UMI No. 1453699)
- Sobhani, M., and Bechara, A. (2011). A somatic marker perspective of immoral and corrupt behavior. *Soc. Neurosci.* 6, 640–652. doi: 10.1080/17470919.2011.605592
- Song, X. G., and Cheng, W. H. (2012). Perception of corruption in 36 major Chinese cities: based on survey of 1,642 experts. *Soc. Indic. Res.* 109, 211–221. doi: 10.1007/s11205-011-9896-4
- Tang, T. L. P., Chen, Y. J., and Sutarto, T. (2008). Bad apples in bad (business) barrels: the love of money, machiavellianism, risk tolerance, and unethical behavior. *Manag. Decis.* 46, 243–263. doi: 10.1108/00251740810854140
- Tangney, J. P., Baumeister, R. F., and Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *J. Pers.* 72, 271–324. doi: 10.1111/j.0022-3506.2004.00263.x
- Treisman, D. (2000). The causes of corruption: a cross-national study. *J. Public Econ.* 76, 399–457. doi: 10.1016/S0047-2727(99)00092-4

- Tversky, A., and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science* 211, 453–458. doi: 10.1126/science.7455683
- Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertain.* 5, 297–323. doi: 10.1007/Bf00122574
- Wang, X. X., and Wu, X. B. (2012). Dilemma of punishing bribes-giving and anti-corruption. *J. Guizhou Police Offi. Vocat. Coll.* 5, 22–48. Available online at: [http://en.cnki.com.cn/Article\\_en/CJFDTOTAL-GJGZ201205005.htm](http://en.cnki.com.cn/Article_en/CJFDTOTAL-GJGZ201205005.htm)
- Wohl, M. J. A., and Enzle, M. E. (2002). The deployment of personal luck: sympathetic magic and illusory control in games of pure chance. *Person. Soc. Psychol. Bull.* 28, 1388–1397. doi: 10.1177/014616702236870
- Wohl, M. J. A., and Enzle, M. E. (2003). The effects of near wins and near losses on self-perceived personal luck and subsequent gambling behavior. *J. Exp. Soc. Psychol.* 39, 184–191. doi: 10.1016/S0022-1031(02)00525-5
- Yang, B., Liu, X., Yang, S. Y., An, S. S., and Ying, L. H. (2007). The effect of personality, social support and irrational belief on the cravings of male drug abstainers in reeducation-through-labor institutions. *J. Psychol. Sci.* 30, 1413–1417. Available online at: [http://en.cnki.com.cn/Article\\_en/CJFDTOTAL-XLKX200706031.htm](http://en.cnki.com.cn/Article_en/CJFDTOTAL-XLKX200706031.htm)

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Zhao, Zhang and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Angels and Demons: Using Behavioral Types in a Real-Effort Moral Dilemma to Identify Expert Traits

Hernán D. Bejarano<sup>1,2\*</sup>, Ellen P. Green<sup>3</sup> and Stephen J. Rassenti<sup>2</sup>

<sup>1</sup> Department of Economics, Center of Economic Research and Teaching, Aguascalientes, Mexico, <sup>2</sup> Economic Science Institute, Chapman University, Orange, CA, USA, <sup>3</sup> School for the Science of Health Care Delivery, Arizona State University, Phoenix, AZ, USA

## OPEN ACCESS

**Edited by:**

Nikolaos Georgantzis,  
University of Reading, UK

**Reviewed by:**

Brice Corget,  
Chapman University, USA  
Jana Peliova,  
University of Economics in Bratislava,  
Slovakia

**\*Correspondence:**

Hernán D. Bejarano  
bejarano@chapman.edu

**Specialty section:**

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

**Received:** 11 July 2016

**Accepted:** 12 September 2016

**Published:** 25 October 2016

**Citation:**

Bejarano HD, Green EP and  
Rassenti SJ (2016) Angels and  
Demons: Using Behavioral Types in a  
Real-Effort Moral Dilemma to Identify  
Expert Traits. *Front. Psychol.* 7:1464.  
doi: 10.3389/fpsyg.2016.01464

In this article, we explore how independently reported measures of subjects' cognitive capabilities, preferences, and sociodemographic characteristics relate to their behavior in a real-effort moral dilemma experiment. To do this, we use a unique dataset, the Chapman Preferences and Characteristics Instrument Set (CPCIS), which contains over 30 standardized measures of preferences and characteristics. We find that simple correlation analysis provides an incomplete picture of how individual measures relate to behavior. In contrast, clustering subjects into groups based on observed behavior in the real-effort task reveals important systematic differences in individual characteristics across groups. However, while we find more differences, these differences are not systematic and difficult to interpret. These results indicate a need for more comprehensive theory explaining how combinations of different individual characteristics impact behavior is needed.

**Keywords:** cognitive capabilities, personality, preferences, real effort, abstract effort, moral dilemma, experiment, survey

## INTRODUCTION

Mainstream economic theory routinely assumes that individuals have stable, consistent preferences that at least partly determine their behavior and revealed preferences (Samuelson, 1948; Stigler and Becker, 1977). Behavioral and experimental economists have explored the validity of that assumption, and phenomena like preference reversals, endowment effects, framing, and the Ellsberg paradox imply that individuals lack stable, consistent preferences.

Most lab experiments attempt to induce consistent preferences using conditional rewards based on Smith's (1976) Induced Value Theory. In these experiments, failure to observe the behavior implied by the induced preferences leads researchers to question the narrow self-interest hypothesis and search for alternative theories. This process has contributed to a deeper understanding of preferences by examining how experimental designs and subject characteristics affect behavior (Frank and Glass, 1991; Becker, 2013). For example, experimental results imply that subjects are partially motivated by fairness (Rabin, 1993), equality (Bolton and Ockenfels, 2006), ambiguity aversion (Fox and Tversky, 1995), and identity (Akerlof and Kranton, 2000).

We argue that even with substantial improvements over the past decades in our understanding of how individual characteristics correlate with individual actions, several key questions remain: Are there systematic differences among individuals? For example, do variations in individual

characteristics matter? If so, which characteristics influence behavior? Do actions reveal more than psychological indicators of behavioral types? Furthermore, little is known about how the answer to these questions depends on the elicitation method.

There are two prevalent approaches used to try to answer these questions: (1) surveying with primary experiments; and (2) adding secondary experimental tasks. In the first approach, researchers use questionnaires either before or after the primary experimental task. For example, several authors have explored how psychological characteristics influence economic behavior using this method—e.g., personality traits (Almlund et al., 2011; Ferguson et al., 2011); emotions (Pixley, 2002); and sentiments (Smith and Wilson, 2013). Corgnet et al. (2015) found that reflective individuals, as measured by the Cognitive Reflection Test (CRT), exhibited more consistently mildly altruistic actions in a lab experiment. Frederick (2005) and Burks et al. (2009) found that cognitive capabilities related to time and risk preferences. Other researchers investigated the interaction between personality traits and risk and time preferences (Rustichini et al., 2012). Researchers have also linked experimental behavior to the results of testing such for IQ (Oechssler et al., 2009; Brañas-Garza et al., 2012, 2015), social intelligence (Takagishi et al., 2010), and personality (Almlund et al., 2011; Rustichini et al., 2012). However, the findings are not consistent with one another (Ben-Ner et al., 2007; Eckel and Grossman, 2008; Borghans et al., 2009; Hirsh and Peterson, 2009; Oechssler et al., 2009; DeAngelo et al., 2015).

The alternative approach is to add secondary experiments that are designed to measure preferences or characteristics. Researchers use these measures to determine the relationship between a subject's actions in the primary experiment and their individual preferences or characteristics. Examples of this practice are the use of the Dictator Game, the Trust Game and Risk and Time Preference experiments as complements to primary experiments. Unfortunately, correlations between behavior in the primary and secondary experiments have not been consistent. For example, while characteristics such as risk preferences have accompanied behavior in games such as repeated prisoner's dilemmas and beauty-contest games (Boone et al., 1999; Sabater-Grande and Georganzis, 2002; Goeree et al., 2003; Brocklebank et al., 2011; Lönnqvist et al., 2011; Kagel and McGee, 2014), the same characteristics sometimes failed to correlate (Aycinena et al., 2014). Another approach has found that in prisoner dilemma games, there are interesting evolutionary explanations for the existence of different types (Congleton and Vanberg, 2001).

In this article, we alter these approaches to address the inconsistencies described above. First, we utilize individual-level subject data collected on different occasions. That is, our measures of individual characteristics and preferences were collected in different experimental sessions from our primary experiment. We argue that, while difficult, using data collected from different experimental sessions implies that subjects are less likely to be influenced by portfolio and wealth effects across tasks. Secondly, we leverage a large dataset with over 30 measures of individual characteristics and preferences, the Chapman Preferences and Characteristics Instrument Set (CPCIS). These

include measures of several types such as: personality traits, preferences, strategic behavior in simple games and the socio demographics of our experimental subjects. Furthermore, the CPCIS was not designed or implemented by us, so it reduces the potential presence of any experimental demand effect. More specifically, the CPCIS not only measures characteristics that we hypothesize to influence the behavior in our primary experimental task, but also a large set of variables which a priori should not influence actions in it.

Our primary experiment, based on Green (2014), presents experimental subjects with a novel real-effort experiment with a distinct moral dilemma. Subjects in this experiment representing experts are asked to provide proofreading services to another group of subjects (customers). The quality of the expert's edits affects the customer, positively if the edits are done properly and negatively if they are done incorrectly. However, the quality of edits has no impact on the expert's personal earnings. Therefore, the experts face a moral dilemma between maximizing personal earnings and providing benefits to their customer.

Behavior in moral dilemmas is hypothesized to be influenced not only by subjects' induced payoff function and preferences for monetary rewards, but also in other-regarding preferences, subject's cognitive capabilities, values and personality traits (Bowles, 1998; Fehr and Fischbacher, 2003). Therefore, we combine observed behavior from our primary experiment, a real-effort moral dilemma task, with the individuals' measures of the CPCIS to see how individual characteristics relate to an individual's actions.

Our results provide several new insights concerning experiments with a moral dilemma. Initially, we find that simple correlational analysis provides an incomplete explanation of how individual measures relate to behavior. Both measures of preferences and other individual characteristics fail to consistently correlate with actions in our main experimental task. For example, measures of individual preferences (i.e., risk aversion, loss aversion, and time preferences) are not correlated with observed actions in the primary experiment. In contrast, some measures of strategic preferences, intelligence, and personality are significantly correlated with behavior. However, in spite of the inconsistency in correlation across individual preferences and behavior, that fact that some measures do correlate is of note. When a subject's preferences are characterized by a combination of factors such as personality, cognitive capabilities, and intelligence, as in our primary experiment, predictions of behavior become uncertain. For instance, subjects with high measures of intelligence should produce higher outcomes for their customers, whereas those same individuals may have varying levels of altruism also influencing their behavior and, thereby, theoretical predictions.

This leads us to explore individuals by behavioral groups, also known as *clusters*. Clusters are identified using the action variables "total edits" and "total incorrect edits." Cluster analysis based on these two variables allows us to distinguish between subjects who edited a lot with a high percentage of incorrect edits (the Demons) and subjects who edited sparsely with a high percentage of incorrect edits, as well as those who edited few with a high percentage of correct edits (Angels).

Behavioral group members exhibited systematic differences in their individual characteristics. We found significant differences among behavioral groups that could not be detected using simple correlation analysis, suggesting that the effect of psychological, cognitive, and demographic differences on behavior in trials with our moral dilemma experiment is nonlinear. These results indicate a need for more comprehensive theory explaining how different individual characteristics work together.

## EXPERIMENTAL DESIGN AND INDIVIDUAL DATA

### Experimental Design

The primary experimental design was introduced by Green (2014). The experimental design and data analyzed here are from Bejarano et al. (2016). Green's original experiment was designed to explore behavior between an expert and customer where the expert is presented with a moral dilemma. Experts are asked to provide proofreading services for a panel of customers. The quality of the expert's proofreading services affects the customer's wellbeing (in the form of monetary payment); however, the customer's wellbeing has no impact on the expert's personal earnings. Therefore, the experts are faced with a tradeoff between maximizing personal earnings and providing benefits for their customer.

The interaction between the expert and the customer took place in two phases with one group of subjects playing the role of the customer (Phase I) and another group playing the role of the expert (Phase II). In Phase I, customers were given 50 min to proofread 10 essays. Each essay had 10 typographical or spelling errors (e.g., misuse of "their" for "there" or "write" for "right"). Customers were initially endowed with \$25; however, for each error they were unable to find, they lost \$0.25. Phase I was designed to create customer demand for the proofreading services provided in Phase II of the experiment.

In Phase II, experts were presented with a panel of 40 customer-edited essays collected in phase I. These essays contained a total of 125 errors. To create the expert subjects, errors were highlighted when presented to the "experts." In addition to the 125 errors that were highlighted, another 250 sections of text were highlighted to create a potential for over-editing.

There were three possible payment schemes for the expert: *fee-for-service*, *capitation*, or *salary*. Under *fee-for-service*, experts were paid \$0.20 per individual field of text edited. Under *salary*, experts were paid a flat rate of \$25 to participate in the experiment. Under *capitation*, experts were paid \$0.625 for each essay in which they edited at least one highlighted section of the text. The expert's edits directly impacted the payoff of their customer. For each incorrect edit, the experts made to the text, customers lost \$0.15 and for each correct edit, customers are reimbursed \$0.05.

Each payment scheme presented a different moral dilemma; that is, strategies to maximize personal earnings or minimize effort varied across payment schemes. Under *fee-for-service*, experts faced a tradeoff between maximizing the number of edits

and the quality of each edit for their customers. Under *salary*, experts faced a tradeoff between leaving the experiments early (minimizing effort) and providing services for their customers.<sup>1</sup> Experts paid under *capitation* faced a tradeoff between the number of customers and the quality of edits for each customer.

In addition to varying the payment scheme, we also varied the expert's ability to select among the payment schemes. Our experiment included two treatments. Under the first, *self-selection*, experts could choose among the three payment schemes. Under the second, random assignment, experts were randomly assigned to one of the three payment systems: *fee-for-service*, *capitation*, or *salary*.

In Green (2014), subjects were randomly assigned to these payment schemes. Consistent with experts randomly assigned in the present analysis, experts in the *fee-for-service* treatment provided significantly more services than those in either the *capitation* or *salary* treatments. This difference was caused by a significant increase in the number of unnecessary edits to the essays provided by the experts, resulting in a much lower quality of service under the *fee-for-service* option compared to the *salary* or *capitation* payment schemes.

### The Chapman Preferences and Characteristics Instrument Set (CPCIS)

Starting in September 2015, the ESI required all subjects to complete the CPCIS prior to participating in ESI experiments. This instrument set required about 90 min of a subject's time and was run independently of any other experiment, at a time convenient to the subject. The data collected by this instrument set consisted of standardized measures of preferences and individual characteristics gleaned from a series of classic simple experiments and questionnaires.

Measures are calculated for and sorted into five characteristic categories: individual preferences, strategic preferences, intelligence, personality tasks, and demographic characteristics. Individual preferences measured in the CPCIS include time preferences, loss aversion, and risk aversion. Strategic preferences include *trust* (adapted from Berg et al., 1995), *fairness* (adapted from Güth et al., 1982), and *altruism* (adapted from Kahneman et al., 1986).

Intelligence is measured using classic psychology measures from Raven, the CRT, and Wonderlic. Additionally, subjects are asked to complete a simple adding task, once with incentives for correctness and once with none. Social intelligence is measured using The Reading the Mind in The Eyes task. Finally, subjects provided self-reported measures of intelligence via their SAT and ACT scores, as well as their GPA. Personality was measured using the Big Five personality test. Demographic variables included age, gender, volunteer hours per week, work hours per week, number of siblings, number of older siblings, and finally, religiosity.

Although, the tests used are somewhat arbitrary and controversial, the results predict behavior in traditional experimental games and are consistent with several behavioral

<sup>1</sup>Subjects who completed their task before the time was up were asked to raise their hand and were then given a short survey to complete silently. Once finished with the survey, subjects quietly exited the room and were paid outside of the laboratory. We found no session effects.

and experimental-economics studies that attempt to elicit relevant preferences. The goal of the CPCIS is to provide a panel dataset that includes the personality indicators most used by experimental economists, with indicators used by psychologists, sociologists, anthropologists, and other social scientists.

In order to integrate several traditional tasks within the same instrument set, tasks within CPCIS such as Raven, The Reading the Mind in The Eyes task and Wonderlic (Test, 1992) were truncated. Specifically, the CPCIS contained the odd-numbered questions from the last three series of matrices within the Raven test (Jaeggi et al., 2010), one that has also been used by Corgnet et al. (2015). Our Big Five questionnaire is based on the 44 items described by John et al. (2008). Conversely, we used an extended version of the CRT (Frederick, 2005). While the original task from Frederick (2005) has three questions, our task has seven questions (Toplak et al., 2011).

In addition to traditional games that elicit several types of other-regarding preferences, the CPCIS includes an instrument that elicits social preferences a la Bartling et al. (2009), hereafter referred to as the *BFMS task*. This task has been used to study preferences of subjects who self-select into competitive tasks (Bartling et al., 2009), as well as the relationship between cognitive capabilities and other-regarding preferences (Corgnet et al., 2015). In our experiment, we combine features of these two applications. Selection into a payment scheme is not based on competitiveness but tradeoffs between the desire to reimburse others and to maximize personal earnings. Therefore, we argue that selection into the different treatments could be related to social preferences elicited by the BFMS. In the following paragraphs, we briefly describe the BFMS that the students in the CPCIS faced<sup>2</sup>.

The BFMS instrument is a series of binary choices with different allocations for the decision maker and a randomly matched partner (**Table 1**). Each choice presents an egalitarian alternative and a non-egalitarian alternative. In our modified BFMS instrument, subjects have to make six choices. Of these six choices, three present subjects with a choice between an egalitarian alternative and another non-egalitarian division earnings, which is at least as good or favorable for herself but detrimental for the matched partner (choices BFMS1, BFMS2, and BFMS5). In contrast, two of the other three binary choices presented to the subject ask her to choose between the egalitarian alternative and a division that is as least as favorable for the matched partner but less than or equal for the decision maker (BFMS3, BFMS6). Finally, BFMS4 is welfare-improving or increases overall earnings but by a greater amount for the matched partner.

In the CPCIS, after all of the subjects made their decisions, two of the individuals were randomly selected to have their choices determine the earnings for this task. Models describing behavior observed in the BFMS task vary across publications. Fehr and Schmidt (1999) presented a two-parameter  $\alpha, \beta$  model,

where  $\alpha$  represents aversion to disadvantageous inequality, Behindness Aversion, and  $\beta$  aversion to advantageous inequality, Aheadness Aversion. Fehr and Schmidt (1999) assumed that  $\alpha > \beta > 0$ . In contrast, Corgnet et al. (2015) related these parameters to envy and compassion and did not impose any assumption on them. The authors summarized five motivations that could make subjects select one alternative over the other. These include self-interest, altruism, egalitarianism, spitefulness, and inequality-seeking. The authors also said that individuals could have a combination of these motives while choosing among alternatives. In order to organize BFMS choices in a way useful for our analysis, we further simplified the choices within three types of preferences. Decision makers who chose alternative A more often across all six choices demonstrated egalitarian preferences. Decision makers who chose to allocate larger earnings to their matched partner than to themselves (alternative A in BFMS3, BFMS4, and BFMS5) at no cost or a small cost to their own earnings, were considered altruistic or averse to being ahead of their partner. Finally, decision makers who were more likely to choose option A in BFMS1, BFMS2, and BFMS5 were considered Spiteful. These individuals could also be considered as having demonstrated aversion to being behind their partner.

Based on these notions, we constructed three variables based on the BFMS choices for each individual. Each individual could choose between zero and six egalitarian alternatives (Egalitarianism). Also, they could choose between zero and three beneficial alternatives (Selfishness) or detrimental alternatives (Altruism). These three variables elaborate on the theory of other-regarding preferences and improve our understanding of how a subject's choices under this instrument relate to their actions in our moral dilemma experiment.

We do not claim that the measures obtained by these truncated tasks mirror those obtained by the original tests, but for the purpose of our analysis, we determine the extent to which these measures are correlated with the experimental actions.

## THEORETICAL RELATIONSHIP AMONG CPCIS VARIABLES

In this section, we analyze the theoretical implications of expert preferences and characteristics. Two experimental-design features are important for our analysis. First, an expert in the self-selection treatment likely reveals something about her personal preferences in her selection of payment systems. Experts who are randomly assigned to their payment scheme will be the average of the general student population, rather than the conditional averages for the subject types that prefer a particular payment scheme. We will distinguish between these two groups in our predictions.

Second, the quality of the expert's proofreading directly impacted the customer's payment. But it had no impact on the expert's personal earnings. In the choice of a payment scheme, all experts in the self-selection treatment faced the same tradeoff, or moral dilemma, between choosing the payment scheme that

<sup>2</sup>It is not within the scope of this article to describe each task in the CPCIS in detail. Most of the tasks included in the CPCIS have been used in several experiments. In this case, we make an exception for the BFMS, assuming that it is not as well-known as the other tasks. Still, we encourage the reader to read Bartling et al. (2009) and Corgnet et al. (2015) for more detailed descriptions of this type of instrument.

**TABLE 1 | Bartling binary choice task.**

Binary choice variable name	Egalitarian alternative A	Non-egalitarian alternative B
BFMS1	Both subjects earn \$10	Decision-Maker earns \$10, Matched Partner \$6
BFMS2	Both subjects earn \$10	Decision-Maker earns \$16, Matched Partner \$4
BFMS3	Both subjects earn \$10	Decision-Maker earns \$10, Matched Partner \$18
BFMS4	Both subjects earn \$10	Decision-Maker earns, \$11 Matched Partner \$19
BFMS5	Both subjects earn \$10	Decision-Maker earns, \$12 Matched Partner \$4
BFMS6	Both subjects earn \$10	Decision-Maker earns \$8, Matched Partner \$16

would maximize personal earnings or one that would limit their maximum earnings. Therefore, selecting a payment scheme may reveal something about subjects' characteristics.

The following *ceteris paribus* predictions highlight the expected relationship between each individual characteristics and behavior in the primary experiment. However, we note that individuals do not differ from each other in *ceteris paribus* ways; therefore theoretical implications are unlikely to describe the expected differences in behavior among any two given subjects.

### Predicted Behavior with Homo Economicus Preferences

The predicted behavior varies with assumptions about expert preferences that are not induced. However, there are simple predictions for the outcomes of these experiments if we assume subjects prefer to be purely self-interested (*homo economicus*). If careful editing requires bearing a real-effort or cognitive cost, a *homo economicus* expert assigned to the *salary* scheme will exert no effort and conduct no edits. A *homo economicus* expert randomly assigned to the *capitation* scheme should exert the minimum effort and only conduct one edit per essay. A *homo economicus* expert assigned to *fee-for-service* should maximize the number of edits with minimum effort and make both necessary and unnecessary edits. Furthermore, in the *selection* treatment, *homo economicus* would select *fee-for-service* 100% of the time, because under that scheme, experts can earn three times more than the maximum earnings possible under *salary* or *capitation*.

However, the experimental evidence presented in Green (2014) and Bejarano et al. (2016) demonstrates that subjects deviated from income-maximizing strategies. These results suggest that subject preferences were more complex than those of *homo economicus*. This leads us to investigate what role additional preferences might be in play in order to modify our assumptions regarding the effects of the payoff schemes on actions.

### Predicted Behavior with Other Preferences and Choice-Relevant Characteristics

The experimental design has some implications concerning the relevance of other personal characteristics as well. For example, risk aversion, loss aversion, and time preferences should not affect behavior. Subjects earnings do not depend on the correctness of their editing but only on their payment system and their decision to edit or not. Payments are deterministic. Therefore, subjects do not face risks of the usual kind. Similarly, the effect of choice on

earnings is almost immediate; hence, time preferences should not influence choices.

On the other hand, a subject's actions in Phase II have an impact on the earnings of subjects who participated on Phase I. Therefore, we expect that measures of what might be regarded as social preferences should affect behavior. For example, differences in the extent of altruism is likely to affect behavior, as has been found in Dictator, Trust, Ultimatum Game, and Prisoner's Dilemma experiments. We expect measures of altruism to be positively correlated with efforts to help subjects in Phase I. Error rates should fall under *fee-for-service*, and more time (and care) should be spent editing under salary and capitation.

The three variables described above (Egalitarianism, Selfishness, Altruism) have an intrinsic relationship with what we expect to uncover with the selection and related actions in our experiment. We expect that those demonstrating Selfishness through these measures will prioritize their earnings over their customers'. Hence, these subjects will likely select *fee-for-service* and perform a larger number of edits rather than maximize their incomes, even at the expense of their customer. In contrast, those individuals that prioritize the earnings of their matched partners will likely choose *salary* and only attempt to conduct beneficial edits for the customers, even at a cognitive and time cost to themselves.

In contrast to the preference measures, predictions regarding Intelligence and demographic variables are not clear. Little is known regarding how actions in our experiment will be influenced by a subject's demographic characteristics. We also have variables that reflect Numeracy, Academic, and IQ Intelligence. To the best of our knowledge, this is the first time that researchers aimed to explore how these measures correlate with performance on incentivized linguistic tasks that affect third parties.

### Cognitive Capabilities and Personality

In this section, we clarify the implications that cognitive capabilities and personality traits could have on the behavior observed in our primary experiment given their indirect relationship with strategic preferences. In a novel study, Corgnet et al. (2015) found that Chapman students with a more reflective nature were less likely than intuitive individuals to be associated with egalitarian and spiteful motives. The authors named the behavior of those with scores above median CRT as mildly altruistic. Given that we have access to the same subject database

with the same measures of cognitive capability (CRT) and preferences for egalitarianism or spitefulness (Bartling et al., 2009), we might expect also that subjects with higher CRTs would show some type of characteristic behavior. However, it is not clear what exactly would comprise mildly altruistic behavior in our experiment. The moral dilemma at hand implies that for each treatment, experts face a different tradeoff between self-interest and customer welfare. We expect subjects with higher CRT scores to be more likely to balance this tradeoff differently in the various treatments examined because they are more likely to reflect on the cost of the tradeoff at stake.

In an attempt to relate personality traits to preferences measures, Rustichini et al. (2012) used a dataset with 1000 truck drivers. They measured the truck drivers' Big Five traits, time preference, risk aversion, truck accidents, job persistence, credit score, and body mass index (BMI). The authors found that personality traits had stronger predictive power than time preferences or risk aversion for truck accidents, job persistence, credit score, and BMI. However, the authors argue that both economic and psychological theories are needed to understand truck-driver behavior.

Big Five personality traits are also likely to help explain differences in the behavior of experts among treatments and payment systems. Unfortunately, the Big Five factors are not orthogonal. Although, qualitative predictions can often be made for individual factors, a person's particular vector of factors often includes factors with the opposite effects on the behavior of interest. For example, *openness* is associated with curiosity and a higher willingness to explore. Therefore, relatively open individuals might be more likely to conduct a larger number of edits and to spend more time on them.

Conscientiousness is associated with being dependable and disciplined. In our experiment, experts have a mission. In their mission, they know that they could affect the earnings of their customers. Higher conscientiousness is likely to be correlated positively with measures of correct edits. Agreeableness is associated with higher cooperation against the exploitation of others (Andersen et al., 2006). We expect that subjects with higher agreeableness should conduct more correct edits to increase the earnings of customers. These three dispositions, therefore, tend to induce better outcomes for the customers.

Higher extroversion is associated with higher sensitivity to rewards. In this case, the perceived nature of the reward matters. Subjects with a higher extroversion measure (maintaining the degree of preferences for others' welfare) may be driven by monetary rewards. In that case, they will be more likely to choose *fee-for-service* and to conduct unnecessary edits. However, if they perceive their reward to be correlated with the benefits of their customers, extroverts will take greater account of such effects than introverts.

Finally, neuroticism appears to be the factor that is not likely to influence the behavior of subjects in a clearly predictive way. Because the experimental environment is set up to isolate subjects from situations where moods, anxiety, and depression

play a significant role, we do not expect to find any significant correlation between neuroticism and behavior.

## EXPERIMENT, DATA, AND ANALYSIS

The experiments were conducted in the ESI laboratory and conference rooms at Chapman University between May 2014 and May 2016. Experimental subjects were recruited from the ESI database of more than 2000 students. Phase I was conducted either in the ESI laboratory or the ESI conference room. Phase II was conducted in the ESI's computer laboratories. Printed instructions were provided for the students to read on their own for 10 min. At the end of the 10 min, the experimental coordinator read the instructions out loud. Subjects were not able to start the experiment until they satisfactorily completed a quiz.

Many of these subjects were also recruited to participate in the CPCIS by a different recruitment email on a previous date convenient to the subject's schedule. The CPCIS sessions were implemented in the same laboratory but had no formal connection to any other experiments being conducted at ESI. The local Institutional Review Board (IRB) approved both studies. In both studies, participants received a show-up fee of 7 USD plus additional incentive payments earned by their behavior in the session.

In the primary experiment, there was a total of 20 undergraduates (customers) recruited in Phase I and 228 undergraduates (experts) recruited in Phase II. In Phase II, which was dedicated to experts performing editing services, 105 subjects were randomly assigned to their payment scheme, and 125 selected their payment scheme. Of the subjects in Phase II, 161 had completed the CPCIS; 115 of those were in the *self-selection* treatment and the other 46 were randomly assigned to one of the three payment schemes. We focus our analysis below on the behavior of those 161 subjects who participated in the primary experiment and had undertaken the CPCIS. The primary experiment lasted an average of 1 h and 15 min, and completion of the CPCIC instrument required an average of 1 h and 35 min.

In the primary experiment, expert subjects could edit correctly or incorrectly. We will focus our analysis on six experimental actions: total edits, total incorrect edits, percentage wrong, net impact on the customer earnings, expert earnings, and total editing time taken. Total edits (total incorrect) is the sum of all (incorrect) edits made by the expert over four rounds of editing. Percentage wrong was calculated by dividing total incorrect by total edited. Cumulative impact on the customer earnings, or impact, was calculated as the customer payoff generated by the expert's behavior over all four rounds. As subjects were given the opportunity to leave the experiment early, total time taken is the amount of time the experts spent editing the essays across all four rounds.

**Table 2** provides a summary of the actions taken in the different treatments. As discussed in Bejarano et al. (2016), experts preferred either *fee-for-service* or *salary* over *capitation*. Those subjects who self-selected *fee-for-service* provided significantly more edits than those randomly assigned,

resulting in more earnings for themselves and less help for their customers. The observed behavior between the randomly assigned *salary* treatment and those who self-selected *salary* did not significantly differ.

We begin with a correlational analysis of the relationship between subjects' actions and CPCIS measures. The correlation analysis only captures the way in which actions correlated with specific individual's characteristics. In the second part of this section, we report the results of a cluster analysis that groups subjects acting in similar ways. These clusters were most salient when subjects could self-select into one of the payment schemes. We analyze whether particular subjects' behavior or action-strategy types are revealed by actions in the experiment, and whether we observe differences across types in the *self-selected* treatment. Finally, we analyze how the observed relationships between experimental actions and CPCIS measures relate to our theoretical hypotheses.

## Correlation Analysis

We start this section by exploring the individual characteristics across the six experimental subject types: self-selected and three randomly assigned into either *fee-for-service*, *capitation*, or *salary* types. When comparing across experimental subject types, we do not expect to see much difference between individual characteristics of those subjects that were randomly assigned individuals to the different payment schemes, because they were randomly selected from the general subject population. In contrast, we would expect to see differences in the individual characteristics of those that self-selected different payment schemes.

We proceed as follows: First, we study the correlation between experimental actions and individual characteristics for all those subjects for whom we have the CPCIS data (A summary of each of the CPCIS data measure can be found in the Appendix). This analysis, which includes the pooled set of randomly assigned and self-selected individuals, should reveal if *ceteris paribus* measures within a characteristic category are strongly correlated with actions in a particular way. Second, we use the fact that self-selecting into different payment schemes might reveal something about a subject's type to better understand behavior. Here, we analyze the correlation between each one of the payment schemes disaggregated by self-selection and randomly assigned with each of the individual characteristic measures in the CPCIS data. In both cases, we estimated the Spearman correlation coefficient<sup>3</sup> and test significance correcting for the multiple hypothesis effects via the Bonferroni adjustment.

In the analysis of the pooled set of subjects, there are two main findings: First and not surprisingly, variables within a characteristic category are typically highly correlated with one another. Second, we did not find any significant correlation between any of the preference measures and subject actions in the experimental treatments. The lack of correlation is consistent with our predictions of individual preferences but surprising for

those measures of strategic preferences, which were hypothesized to play a role in behavior in our primary experiment.

One exception is the correlation between all BFMS variables, measures of strategic preference, and action variables in our primary experiments. Particularly, we observed that when evaluating the correlation between the pooled data, i.e., all subjects in all treatments, selfishness correlates positively with total edited ( $r_s = 0.185$ ,  $p < 0.10$ ). Furthermore, in all three cases, the three variables, egalitarianism, altruism and selfishness, have significant positive correlation with the amount experts earned with  $r_s = 0.158$ ,  $r_s = 0.221$ , and  $r_s = 0.179$ , and  $p < 0.10$ , respectively. In contrast, altruism is not correlated with the number of wrong edits or its percentage. Furthermore, both egalitarianism and selfishness have a positive correlation with the number of wrong edits (and its percentage) with these respective statistics,  $r_s = 0.1779$ ,  $r_s = 0.214$ , and  $p < 0.05$  in both cases. Accounting for self-selection in general or self-selection into a particular payment scheme, all these correlations hold their significance except the correlation between the number of total edits, which now is not statistically significantly related to egalitarianism.

We found CRT measures correlated with total earnings in two dimensions: The number of correct CRT answers is positively correlated with total earnings ( $r_s = 0.2348$ ,  $r_s = 0.3030$ ,  $p < 0.05$ ), and CRT impulsiveness is negatively correlated with total earnings ( $r_s = -0.2270$ ,  $p < 0.10$ ). This result is consistent with the findings of Corgnet et al. (2015) given that CRT relates to how compulsive/deliberative subjects are. However, these results should not be generalized since these traits could affect both the self-selection and the actions taken by subjects after this choice. Therefore, the outcome could be either driven by the self-selected portion of the subjects or not.

The lack of significant correlation between most of our measures of individual characteristics and subject actions conflicts with the theoretical hypotheses that we discussed in the previous section. None of the preference measures were correlated with any of the experimental action variables. Several explanations for this result are feasible. One possible explanation for the lack of correlations is that the CPCIS instrument and the primary experiment were conducted at different times by different researchers. This might imply that subjects are less likely to act in a manner consistent with the behavior characterized by their responses to the CPCIS tasks while performing in the primary experiment. Differences in the timing and circumstances of the CPCIS tasks and the primary experiments imply that their behavior in the primary experiment is less likely to reflect any implicit experimenter demand effect.

We continue our analysis by examining only correlations among those who self-selected the same treatment. This is an important step in our analysis, as the act of choosing a treatment might reveal differences in individual characteristics. To analyze this possibility, we break down the correlation analysis into two steps. First, we conduct the same correlation analysis as above but only for those subjects in the *self-selection* treatment.

Not surprisingly, there is no correlation between experimental actions and the individual characteristic measures of the CPCIS

<sup>3</sup>The Spearman's rho is a nonparametric measure of rank correlation. The assumption of monotonicity in Spearman's rho test is satisfied.

**TABLE 2 | Actions summary by treatment.**

Variable	Randomly assigned			Selection		
	Fee-for-service	Salary	Capitation	Fee-for-service	Salary	Capitation
Total edited	175.8	81.7	89.2	250.3	85.8	73.8
Total wrong	98.1	7.6	17.6	180.6	13.8	7.0
Total correct	77.8	74.1	71.5	69.7	72.0	66.8
Percentage correct	62%	90%	83%	40%	85%	89%
Cumulative impact	6.8	10.8	9.8	1.1	10.4	9.7
Total earnings	\$35	\$25	\$21	\$51	\$25	\$23
Number of subjects	39	41	25	49	70	4

for the subjects that self-selected the two most popular payment schemes<sup>4</sup>. The next step in our analysis is to break down the correlation analysis, controlling for self-selection into a particular payment scheme, *salary* or *fee-for-service*.

The analysis of correlation between subjects that self-selected a similar moral dilemma presents two main findings. First, almost all the finding of the analysis of the pooled set of a subject's data persists. This means that those characteristics that were not found significantly correlated persisted and presented a lack of relationship with actions and were still not correlated when disaggregating by payment scheme. In contrast, we found that the selection choice may work as a screening device of subjects with different values for those that were found significant for all the self-selected subjects. This is reflected by the fact that accounting for the particular payment schemes eliminates the significance for those relationships that were significant for the pooled set of subjects into both payment schemes. This result holds for all the correlations between actions and individual characteristics reflected by variables such as egalitarianism, altruism and selfishness, as well as CRT correct and CRT. This result could be explained if values for these variables and actions are similar among those that self-selected salary but very different for those that self-selected fee for service.

## Cluster Analysis

The results of the previous section lead us to believe that there may be different types of experimental subjects. More specifically, we argue that the inconsistencies in correlations between measures of individual characteristics and observed actions are due to the fact that in our primary experiment multiple characteristics, i.e., cognitive capabilities, individual preferences, social preferences and personality traits, might affect behavior. That is, given the moral dilemma and real effort features of our primary experiment, we expect that certain individual characteristics will pull the subject's behavior in opposite directions. For example, experts with high measures of intelligence would be more likely to provide better outcomes for their customers, whereas low levels of altruism imply worse outcomes for their customers. Therefore, a subject's the combination of the individual characteristics each subject

possesses may have uncertain implications for theoretical predictions.

For this reason, we next explore if expert actions reveal behavioral types and whether behavioral groups correspond to differences in preference, cognitive, and demographic characteristics. To do this, we use cluster analysis to build behavioral groups from the actions of subjects in the selection treatment of our primary experiment.

Clusters (behavioral groups) are based on a subject's actions. Specifically, behavioral groups are created using the action variables "total edits" and "total incorrect edits." Cluster analysis based on these two variables allows us to distinguish between subjects who edited a lot with a high percentage of incorrect edits (the Demons) and subjects who edited sparsely with a high percentage of incorrect edits, as well as those who edited few with a high percentage of correct edits (the Angels).

Behavioral groups were created using the *k*-mean algorithm with Euclidian distances. We clustered on values of *k* from 2 to 6 and maximized the Calinski and Harabasz (CH) pseudo f-statistics to find the optimal clustering (Calinski and Harabasz, 1974). In order to control for the robustness of the *k*-mean algorithm, we ran it in a loop with 50 repetitions for each value of *k*. From these repetitions, we selected the cluster with the highest CH pseudo F-statistic for each value of *k*. Then, comparing across the *k* values, we selected the clustering with the highest CH pseudo f-statistic.

## Behavioral Groups

**Figure 1** and **Table 3** provide summaries of the cluster groupings. **Table 3** summarizes the experimental actions taken by the typical member of the five behavioral groups created by our cluster analysis. The results displayed in **Table 3** reveal three things. First, they reveal that various subjects in our primary experiment behaved in very different ways. Second, the significant differences on actions across behavioral groups imply that our cluster methodology identified different types of subjects. Lastly, a large part of subject behavior is captured by the subjects' choices of payment scheme. The payment scheme selection action completely and consistently separates the five groups into two subsets, {A, B, D} and {C, E}. No significant differences exist between any pair of groups from within either subset, but significant differences do exist between any pair of groups across subsets.

<sup>4</sup>Total time taken showed a positive correlation with self-reported GPA ( $r_s = 0.1981, p < 0.10$ ).



FIGURE 1 | K mean cluster analysis.

TABLE 3 | Actions summary by group.

Variable	Group A	Group B	Group C	Group D	Group E
Total edited	370	328	37	182	101
Total wrong	347	241	6	99	13
Total correct	23	87	31	83	88
Percentage correct	6%	26%	80%	46%	87%
Cumulative impact	-14	2	5	7	12
Total earnings	\$74	\$63	\$24	\$36	\$25
Percentage fee-for-service	100%	93%	5%	90%	21%
Number of subjects	13	14	19	10	63

However, payment choice does not capture all the dimensions of subject behavior. **Figure 1** reveals that even for those groups with a large percentage of subjects choosing the *fee-for-service* payment scheme (Groups A, B, and D), behavior varied significantly. And although a much smaller percentage of subjects chose the *fee-for-service* payment scheme, Groups C and E also displayed dissimilar behavior in other dimensions. For example, although Group E has a large number of subjects choosing salary rather than the *fee-for-service*, the *fee-for-service* subjects of Group E (the Angels) behaved very different than *fee-for-service* subjects in Groups A, B and D and, in particular, most different from those in Group A (the Demons).

The experimental actions from our primary experiment show strong support for the existence of behavioral types as revealed in **Table 3** and **Figure 1**. In the spirit of the ongoing claims in various fields of behavioral science, we seek to determine whether the differences in primary experimental behavior relate to individual characteristics that may be captured independently

TABLE 4 | Summary of FPRANK comparisons across groups and individual preferences.

	Group A	Group B	Group C	Group D	Group E
Group A	–	6	5	4	3
Group B	6	–	7	5	7
Group C	5	7	–	6	4
Group D	4	5	6	–	2
Group E	3	7	4	2	–
Totals	18	25	22	17	16
% of Total	13%	18%	16%	13%	12%

by the CPCIS database. Understanding this question is of great importance to experimental research.

In order to test the hypotheses that there are no differences in the individual characteristics of students who have been clustered into different groups, we perform a binary comparison of the aggregate experimental actions taken by subjects in each pair of groups for each CPCIS characteristic. The complete results for the Two Sample Fligner-Policello Rank Test are displayed in **Table 5**. **Table 4** provides a summary of these results by reporting the count of the number of CPCIS characteristics in which each pair of behavioral groups differed significantly.

Although, all the groups were formed by Chapman students, each group displayed at least 2 and up to 9 significant differences in the characteristics of its membership. There were 26 characteristic differences that reinforced the basic subdivision ( $\{A, B, D\}$ ,  $\{C, E\}$ ) that was revealed by payment selection, but there were 23 characteristic differences between groups within

**TABLE 5 | Summary statistics of action and individual characteristics.**

Variable	Group A		Group B		Group C		Group D		Group E	
	Mean	Ranking	Mean	Ranking	Mean	Ranking	Mean	Ranking	Mean	Ranking
<b>INDIVIDUAL PREFERENCES</b>										
Risk aversion	4.00	3	–	4.15	2	>3	4.00	3	<2	3.90
Loss aversion	3.38	3	–	3.15	5	<1	3.28	4	–	3.90
Time preference	6.00	2	–	6.23	1	–	4.94	5	–	5.30
<b>STRATEGIC PREFERENCES</b>										
<b>Trust</b>										
Trust sent	7.69	2	–	8.46	1	–	6.11	3	–	6.00
Trust return	13.08	1	>4, 3	10.38	4	<1	13.06	2	>3	9.00
<b>Ultimatum</b>										
Offer	4.85	5	–	5.69	2	–	5.11	4	–	5.70
First accepted offer	3.92	1	–	3.62	4	–	3.67	3	–	3.80
Advantageous offers	0.00	4	–	0.31	2	–	0.50	1	–	0.50
<b>Dictator</b>										
Sent	4.31	4	<2	5.62	1	–	5.44	2	>1, 3	4.30
<b>Prisoners' dilemma</b>										
Cooperative action	1.54	1	>4	1.23	4	<1	1.50	2	–	1.39
<b>Bartling</b>										
Egalitarianism	3.54	1	>5	3.07	4	–	3.21	2	>5	2.70
Altruism	0.92	5	–	1.50	1	–	1.26	3	–	1.40
Selfishness	1.54	2	>5	1.00	5	<2, 1	1.21	3	<1	1.90
<b>INTELLIGENCE</b>										
<b>Psychology</b>										
Raven	13.69	1	–	13.15	3	–	12.17	4	–	12.60
CRT	3.23	2	–	4.08	1	>5	2.67	5	<1	3.10
Wonderlic	19.77	3	–	20.23	1	>5	18.94	5	<1, 2	20.20
<b>Numeracy</b>										
Adding task (Incentivized)	15.38	4	–	17.54	1	>5	14.78	5	<1	16.60
Adding task (Not Incentivized)	13.85	4	<1	17.08	1	>4, 5	12.83	5	<1, 2, 3	16.80
<b>Academic</b>										
SAT	5.15	4	–	5.38	2	>5	5.29	3	–	5.40
ACT	6.42	1	–	6.23	2	–	5.82	5	–	5.89
GPA	3.61	2	>5	3.72	1	>5, 3	3.41	5	<2, 1, 3	3.45
<b>Social</b>										
Theory of mind	27.69	1	–	27.30	2	–	26.41	5	–	26.50
<b>PERSONALITY-BIG 5</b>										
Openness	36.54	4	–	39.08	1	–	36.94	3	–	36.40
Conscientiousness	30.77	5	–	32.23	2	–	32.06	3	–	31.10
Extroversion	28.08	1	>4	26.85	3	–	27.28	2	–	26.60
Agreeableness	31.77	4	<3, 2, 1	31.54	5	<1, 2	33.72	3	>4	35.20
Neuroticism	25.85	1	>4, 5	24.08	2	–	21.50	4	<1	20.00
<b>DEMOGRAPHICS</b>										
Age	18.85	4	–	18.62	5	<1	18.89	3	–	18.90
Gender	1.54	2	–	1.43	5	–	1.53	3	–	1.60
Volunteer hours	1.15	5	<1	1.67	1	>3, 4, 5	1.47	3	<1	1.22
Work hours	1.46	5	<1	2.25	1	>3, 4, 5	1.76	3	<1	1.89
Number of siblings	1.23	4	<1	1.15	5	–	1.56	2	–	1.50
Older siblings	0.77	4	–	0.46	5	<3	0.83	3	>5, <1	1.10
Religiosity	1.67	3	–	1.54	4	<1	1.41	5	<2	1.90

Subjects who chose the capitation payment were not able to be clustered into a group. Their data is not summarized here.

Significance in ranking reported by < or > signs next to ranking.

the same subset; this allows us to differentiate between groups that have a similar predilection for payment scheme.

**Table 5** displays the differences among those individual characteristics for each group. For each variable in the CPCIS, we provide the information of the mean at the group level and the number of subjects in the group. We also rank them from the highest (1) to the lowest (5) value and the sign for those differences that were statistically significant according to the results of the binary Two Sample Fligner-Policello Rank Test. We next describe how the results in **Table 5** relate to the theoretical implications discussed in Section Theoretical Relationship among CPCIS Variables.

### Individual Preferences

In contrast to the correlation analysis where no measures of individual preferences were significantly different, risk aversion, and loss aversion were each significantly different between two groups (Groups B > C and Group D > Group B, respectively). This is surprising as experts' risk aversion and loss aversion profiles should not affect their behavior as the subjects are in control of their actions and thereby, their earnings. However, these results cannot be rationalized by either a non-egocentric egocentric view of preferences over other's risk and loss (Hsee and Weber, 1997). If we consider each action as a choice between an uncertain outcome (i.e., edit is potentially right or wrong) and a certain outcome (i.e., no edit means no risk) for their counterpart, the number of edits conducted would reflect one's risk aversion. However, Group C behaved more conservatively in editing than Group B, whereas Group C is less risk averse than Group B. In contrast, Group B had lower levels of loss aversion than Group D and conducted significantly more edits than Group D. This demonstrates Group B's willingness to act carelessly in decisions that negatively impact others more so than the behavior of Group D. These results provide a first indication of how difficult it is to relate measures of individual preference to behavior in a real-effort moral dilemma.

### Strategic Preference

In contrast to the correlation analysis, more measures of strategic preferences were found to be significantly different. First, we predicted that subjects with higher levels of reciprocity would act more benevolently than others in our primary task; that is, these subjects would provide a higher cumulative impact (income) for their customers. However, we found that when comparing the actions in the Trust Game (reciprocity), groups with high levels of reciprocity were less benevolent to their customers. For example, groups that provided a large number of incorrect edits, such as Group A (the Demons), or relatively few edits, such as Group C, had higher rankings of reciprocity, i.e., returned more money on the trust game. However, it is important to note that in the Trust Game, reciprocity from the recipient is conditional whereas, in our experiment, expert actions toward customers are not. In our primary experiment, the benefits experts confer to their customers do not affect their own earnings. This key distinction may explain the unexpected behavior.

BFMS measures of egalitarianism, selfishness or altruism also partly contradict our theoretical predictions. Groups A and C

had significantly larger measures of egalitarianism in the BFMS relative to the other groups; however, in our experiment Group A's actions reflect those of *homo economicus* and Group C's reflected those of an egalitarian. Group C has the lowest overall personal earnings and the 3rd highest cumulative impact for their customer. A similar relationship appears with measures of selfishness reported by the BFMS. Groups A and C are among those with higher levels of selfishness. The inconsistency in behavior and similarity of BFMS measures in these two Groups leads us to question the usual interpretation of the BFMS measure.

### Intelligence

We found that 6 of our 9 measures of intelligence differed across groups. In contrast to the simple correlation analysis, the different behavioral groups were not drawn from the same population with respect to the CRT and Wonderlic test results. For the CRT we found that Group B, the group with the highest CRT values, also behaved in a way that could be described as mildly altruistic and selfish. Group B mostly opted for *fee-for-service* and conducted a large number of edits, thereby increasing their earnings. However, relative to Group A, who also provided a large number of edits, Group B provided more accurate edits. This behavior we characterize as mildly altruistic and selfish, and it is consistent with our theoretical predictions. Similar results were observed by Corgnet et al. (2015). Corgnet et al. observed in their experiments that individuals with high CRT scores behave in a more altruistic way.

Our final measure of intelligence that had significant differences across groups is by numeracy both in an incentivized task and not incentivized. In our theoretical predictions, we argue that there is no relationship between numeracy and the task in our primary experiment. However, upon reflection, statistically significant differences on the not incentivized Adding Task would not contradict our predictions. Notice, that this CPCIS task measures more than numeracy skills, as subjects are not presented with any monetary incentive to add correctly. Therefore, the correct additions performed in this task is also a measure of intrinsic motivation. Hence, finding that those who groups with the highest number of correct edits (Groups B, D, and E) also have the highest scores on the not incentivized Adding Task is as one would expect.

### Personality

**Table 5** reveals that the behavioral groups differ with regard to at least three personality measures (Extroversion, Agreeableness, and Neuroticism). Two measures, however, Openness and Conscientiousness, are not statistically different among groups. The finding that Conscientiousness does not differ amongst groups regardless of choice of payment contradicts our theoretical discussion. However, we do find two supporting results. First, Group A (the Demons) has higher levels of extroversion. This group is only composed of subjects that chose *fee-for-service*; it conducted the most edits on average and had the lowest percentage of correct edits, all of which are consistent with an extrovert's attitude toward rewards. Second, Groups D and E showed the highest values of Agreeableness. These

two groups also had higher cumulative impact (the return of dollars to customers as a result of their actions). This result is consistent with the compassionate attitude associated with this trait. Finally, we also found that Neuroticism differs significantly among groups. Particularly, Groups A, B, and D, the three groups with the largest numbers of edits, presented higher values of Neuroticism than the Groups with lower numbers of edits, Groups C and E.

### Demographics

Several demographic characteristics presented significant differences amongst groups. Of particular interest to our analysis are self-reported numbers of volunteer and work hours. Again, Group B ranked the highest for these two variables. We have already described the behavior of Group B as mildly altruistic and selfish, so it is encouraging that the results are consistent with our previous finding.

## DISCUSSION

In general, understanding how individual characteristics influence behavior is a fundamental task of the economist, psychologist, and scientist. While crucial, scientists rarely have independent datasets that combine both an individual subject's characteristics and behavior (Caplan, 2003). In this article, we leverage a uniquely large dataset containing the individual characteristics of a subset of our experimental subjects to shed light on the relationship between subject's individual characteristics and their behavior in a real-effort moral dilemma with self-selection by payment scheme. Due to the unique nature of our primary experiment, we use two statistical approaches, correlation analysis and cluster analysis, to better understand these dynamics. Different scholars collected our two datasets at different times for different reasons. This allowed us to avoid issues associated with a sequence of primary and secondary experiments conducted by the same experimental team and setting. The following points summarize our results.

First, there is no clear majority of individual characteristics that correlate with behavior in our primary experiment. A set of a few, but interesting, significant correlation relationships were found across experimental actions. We found that no measure of individual preference, i.e., time discounting, risk and loss aversion, was significant. Furthermore, our measures of strategic preferences, which include variables such as Trust, Trustworthiness, and Altruism captured from implementation of canonical Trust and Ultimatum Games also failed to show any significant correlation with actions in a real-effort moral dilemma. These results highlight the importance of conducting reliability tests for simple statistical analyses exploring these social preferences (Charness and Rabin, 2002). Other measures of social preferences, such as those derived from Bartling et al. (2009), were significantly correlated with actions, but the results of these correlations were contradictory to our theoretical predictions.

Measures of intelligence and personality traits also often failed to correlate with observed behavior in our primary experiment. This result is consistent with previous findings (Becker et al.,

2012) and presents an additional call to a better development in the study of the relationship between personality traits and economic behavior (Almlund et al., 2011; Rustichini et al., 2012). Hence, there is a need for replication of this investigation to develop better theoretical models (Benjamin et al., 2013).

There are several arguments that may justify these inconsistencies. First, ordering of tasks has been shown to impact outcomes in experiments. For instance, Healy et al. (2016) demonstrates that by going from a single shot of a game to a repeated game, subjects' payment functions change and thereby, so do behaviors. Similarly, implementing a sequence of tasks, primary and secondary, or a battery of tasks and surveys, as with the CPCIS, could induce different behavior through wealth and portfolio effects. Here, we analyze the correlation between a single task (our primary experiment) and a battery of instruments (the CPCIS), collected on separate occasions. Therefore, behavior in our primary experiment should be less affected by the behavior of the CPCIS than if both datasets had been collected in the same session, producing less consistent correlations than otherwise.

Secondly, it is possible that the joint implementation of tasks in the CPCIS dataset generate spurious correlations. These spurious correlations could be generated from the bundling of experimental tasks, experimenter demand effects (Zizzo, 2010), or idiosyncratic effects of experimenter teams or their lab set up. For instance, researchers often only conduct secondary experiments that they believe will reveal something about their subjects. This potentially introduces an experimenter demand effect instead of the desired elicitation of additional characteristics, resulting in spurious correlations. We argue that the difficulty of finding significant correlations in our analysis, when both of the datasets were collected by separate research teams, is a call for attention to the interpretation of correlations found between primary experiments and secondary measures that are jointly collected. Furthermore, these findings open several research questions regarding how to implement and analyze the results of several experimental tasks, which *a priori* are correlated.

Following the correlation analysis, we found that actions in a primary experiment could be used to categorize subjects into groups based on their observed actions using cluster analysis (i.e., behavioral groups). Furthermore, because of the availability of the CPCIS data we could proceed one step further than several experiments, which have already utilized cluster analysis with the investigation of individual characteristics (Houser et al., 2004; Rong and Houser, 2015).

The cluster analysis reveals that individual characteristics are a distinguishing factor across behavioral groups. Individual measures of preferences (Risk and Loss aversion), strategic preferences (Trust Game, Dictator Game, Prisoners' Dilemma, and Bartling) and Intelligence (CRT, Wonderlic, and numeracy) all varied across behavioral groups. However, like the correlation analysis, the results often contradicted our theoretical predictions. Regardless, it is important to note that these behavioral groups revealed systematic differences in behavior regardless of inconsistencies with theoretical predictions. That is, due to the tension between some of our theoretical analyses

of the influence of personal characteristics on behavior in our moral dilemma and the observed behavior, either our theory or our measures are still far from perfect.

Our results suggest that the effects of psychological, cognitive, and demographic differences on behavior in experiments are more complex than those implied by *ceteris paribus* hypothesis. Subjects are endowed with mixtures of individual characteristics that could present contradictory theoretical interpretations. Despite this difficulty, characteristics of subjects that chose and act similarly (i.e., belong to the same behavioral group) are more likely to be similar between each other and different from those that chose and act differently in individual characteristics. This finding could not be detected using a simple correlation analysis.

## REFERENCES

- Akerlof, G. A., and Kranton, R. E. (2000). Economics and identity. *Q. J. Econ.* 115, 715–753. doi: 10.1162/003355300554881
- Almlund, M., Duckworth, A. L., Heckman, J. J., and Kautz, T. D. (2011). *Personality Psychology and Economics*. No. w16822. National Bureau of Economic Research. Available online at: <http://www.nber.org/papers/w16822> (Accessed July 4, 2016).
- Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2006). Elicitation using multiple price list formats. *Exp. Econ.* 9, 383–405. doi: 10.1007/s10683-006-7055-6
- Aycinena, D., Bejarano, H., and Rentschler, L. (2014). *Informed Entry in Auction*. Working Paper. Available online at: [http://fce.ufm.edu/wp-content/uploads/2014/12/AEP2\\_2015.01\\_WP.pdf](http://fce.ufm.edu/wp-content/uploads/2014/12/AEP2_2015.01_WP.pdf)
- Bartling, B., Fehr, E., André Maréchal, M., and Schunk, D. (2009). Egalitarianism and competitiveness. *Am. Econ. Rev.* 99, 93–98. doi: 10.1257/aer.99.2.93
- Becker, A., Deckers, T., Dohmen, T., Falk, A., and Kosse, F. (2012). *The Relationship between Economic Preferences and Psychological Personality Measures*. Institute for the Study of Labor (IZA), Discussion Paper No. 6470. Available online at: <http://ftp.iza.org/dp6470.pdf> (accessed July 4, 2016).
- Becker, G. S. (2013). *The Economic Approach to Human Behavior*. University of Chicago Press.
- Bejarano, H., Green, E. P., and Rassenti, S. (2016). *Payment Scheme Self-Selection in the Credence Goods Market: An Experimental Study*. No. 16-04. Economic Science Institute Working Paper No. 16-04. Available online at: [http://www.chapman.edu/research-and-institutions/economic-science-institute/\\_files/WorkingPapers/Payment-Scheme-Self-Selection-in-the-Credence-Goods-Market1.pdf](http://www.chapman.edu/research-and-institutions/economic-science-institute/_files/WorkingPapers/Payment-Scheme-Self-Selection-in-the-Credence-Goods-Market1.pdf) (Accessed July 4, 2016).
- Benjamin, D. J., Brown, S. A., and Shapiro, J. M. (2013). Who is ‘behavioral’? Cognitive ability and anomalous preferences. *J. Eur. Econ. Assoc.* 11, 1231–1255. doi: 10.1111/jeea.12055
- Ben-Ner, A., Puterman, L., and Ren, T. (2007). *Lavish Returns on Cheap Talk: Non-Binding Communication in a Trust Experiment*. Available online at SSRN: <http://ssrn.com/abstract=1013582>
- Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games Econ. Behav.* 10, 122–142. doi: 10.1006/game.1995.1027
- Bolton, G. E., and Ockenfels, A. (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: comment. *Am. Econ. Rev.* 96, 1906–1911. doi: 10.1257/aer.96.5.1906
- Boone, C., De Bert, B., and Van Witteloostuijn, A. (1999). The impact of personality on behavior in five Prisoner’s Dilemma games. *J. Econ. Psychol.* 20, 343–377. doi: 10.1016/S0167-4870(99)00012-4
- Borghans, L., Heckman, J. J., Golsteyn, B. H., and Meijers, H. (2009). Gender differences in risk aversion and ambiguity aversion. *J. Eur. Econ. Assoc.* 7, 649–658. doi: 10.1162/JEEA.2009.7.2-3.649
- Bowles, S. (1998). Endogenous preferences: the cultural consequences of markets and other economic institutions. *J. Econ. Lit.* 36, 75–111.
- We believe that the results of our analysis shed light on the strength of the links between individual characteristics, behavior in simple strategic games, behavior in real-effort moral dilemmas.

## AUTHOR CONTRIBUTIONS

Each author contributed to the design, implementation, analysis, and writing of the document.

## ACKNOWLEDGMENTS

We would especially like to thank Roger Congleton for his helpful comments and suggestions.

- Frederick, S. (2005). Cognitive reflection and decision making. *J. Econ. Perspect.* 19, 25–42. doi: 10.1257/089533005775196732
- Goeree, J. K., Holt, C. A., and Palfrey, T. R. (2003). Risk averse behavior in generalized matching pennies games. *Games Econ. Behav.* 45, 97–113. doi: 10.1016/S0899-8256(03)00052-6
- Green, E. P. (2014). Payment systems in the healthcare industry: an experimental study of physician incentives. *J. Econ. Behav. Organ.* 106, 367–378. doi: 10.1016/j.jebo.2014.05.009
- Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* 3, 367–388. doi: 10.1016/0167-2681(82)90011-7
- Healy, P. J., Azrieli, Y., and Chambers, C. P. (2016). *Incentives in Experiments: A Theoretical Analysis*. Economics Department Working paper, Ohio State University.
- Hirsh, J. B., and Peterson J. B. (2009). Extraversion, neuroticism, and the prisoner's dilemma. *Person. Indiv. Dif.* 46, 254–256. doi: 10.1016/j.paid.2008.10.006
- Houser, D., Keane, M., and McCabe, K. (2004). Behavior in a dynamic decision problem: an analysis of experimental evidence using a Bayesian-type classification algorithm. *Econometrica* 72, 781–822. doi: 10.1111/j.1468-0262.2004.00512.x
- Hsee, C. K., and Weber, E. U. (1997). A fundamental prediction error: Self-others discrepancies in risk preference. *J. Exp. Psychol.* 126, 45–53. doi: 10.1037/0096-3445.126.1.45
- Jaeggi, S. M., Buschkuhl, M., Perrig, W. J., and Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory* 18, 394–412. doi: 10.1080/09658211003702171
- John, O. P., Naumann, L. P., and Soto, C. J. (2008). “Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues,” in *Handbook of Personality: Theory and Research*, Vol. 3, eds O. P. John, R. W. Robins, and L. A. Pervin (New York, NY: Guilford Press), 114–158.
- Kagel, J., and McGee, P. (2014). Personality and cooperation in finitely repeated prisoner's dilemma games. *Econ. Lett.* 124, 274–277.
- Kahneman, D., Knetsch, J. L., and Thaler, R. (1986). Fairness as a constraint on profit seeking: entitlements in the market. *Am. Econ. Rev.* 76, 728–741.
- Lönnqvist, J.-E., Verkasalo, M., and Walkowitz, G. (2011). It pays to pay—Big Five personality influences on co-operative behaviour in an incentivized and hypothetical prisoner's dilemma game. *Pers. Individ. Dif.* 50, 300–304. doi: 10.1016/j.paid.2010.10.009
- Oechssler, J., Andreas, R., and Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *J. Econ. Behav. Organ.* 72, 147–152. doi: 10.1016/j.jebo.2009.04.018
- Pixel, J. (2002). Emotions and economics. *Sociol. Rev.* 50, 69–89. doi: 10.1111/j.1467-954x.2002.tb03592.x
- Rabin, M. (1993). Incorporating fairness into game theory and economics. *Am. Econ. Rev.* 83, 1281–1302.
- Rong, R., and Houser, D. (2015). Growing stars: a laboratory analysis of network formation. *J. Econ. Behav. Organ.* 117, 380–394.
- Rustichini, A., DeYoung, C. G. D., Anderson, J. A., and Burks, S. V. (2012). *Toward the Integration of Personality and Decision Theory in the Explanation of Economic and Health Behavior*. Institute for the Study of Labor (IZA), Discussion Paper No. 6750 (IZA, 6570). Available online at: <http://ftp.iza.org/dp6750dp6570.pdf> (Accessed July 4, 2016).
- Sabater-Grande, G., and Georgantzis, N. (2002). Accounting for risk aversion in repeated prisoners' dilemma games: an experimental test. *J. Econ. Behav. Organ.* 48, 37–50. doi: 10.1016/S0167-2681(01)00223-2
- Samuelson, P. A. (1948). Consumption theory in terms of revealed preference. *Economica* 15, 243–253. doi: 10.2307/2549561
- Smith, V. L. (1976). Experimental economics: induced value theory. *Am. Econ. Rev.* 66, 274–279.
- Smith, V. L., and Wilson, B. J. (2013). Sentiments, Conduct, and Trust in the Laboratory. Economic Science Institute Working Paper, Chapman University.
- Stigler, G. J., and Becker, G. S. (1977). *De gustibus non est disputandum*. *Am. Econ. Rev.* 67, 76–90.
- Takagishi, H., Kameshima, S., Schug, J., Koizumi, M., and Yamagishi, T. (2010). Theory of mind enhances preference for fairness. *J. Exp. Child Psychol.* 105, 130–137. doi: 10.1016/j.jecp.2009.09.005
- Test, W. P. (1992). *Wonderlic Personnel Test and Scholastic Level Exam User's Manual*. Libertyville, IL: Wonderlic Personnel Test, Incorporated.
- Toplak, M. E., West, R. F., and Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Mem. Cogn.* 39, 1275–1289. doi: 10.3758/s13421-011-0104-1
- Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Exp. Econ.* 13, 75–98. doi: 10.1007/s10683-009-9230-z

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Bejarano, Green and Rassenti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## APPENDIX

**TABLE A1 | CPCIS Taxonomy.**

Measure	Method–Test (Reference)	Related elicited characteristic
<b>INDIVIDUAL PREFERENCES</b>		
Risk aversion	Multiple listing method (Andersen et al., 2006)	Preferences over:
Loss aversion	Multiple listing method (Andersen et al., 2006)	Risk over lotteries
Time	Multiple listing method (Andersen et al., 2006)	Lotteries with losses
		Temporarily based payments
<b>STRATEGIC PREFERENCES</b>		
<b>Trust</b>	Trust Game (Berg et al., 1995)	
Trust sent	Trustor can send only 0 or 10	Trust
Trust SM return	Trustee respond to each possibility (Strategy Method SM)	Reciprocity
<b>Ultimatum</b>	Ultimatum (Güth et al., 1982)	
Offer	Strategy method, first player can send any even number between (0–20, 11 choices)	Altruism
First accepted offer	When playing as a second player, player can reject any proposal	Fairness
Number of advantageous offers rejected		Fairness
<b>Dictator game</b>	Dictator game (Kahneman et al., 1986)	
	Strategy method, player can send any even number between (0–20, 11 choices)	
<b>Egalitarianism, Altruism, and Selfishness</b>	Bartling, Fehr, Marechal, and Schunk (BFMS) Task from Bartling et al. (2009)	
Egalitarianism	Zero to six choices of an equitative alternative over non-equitative	Preferences for equitatives distribution
Altruism	Zero to three choices of a detrimental alternative over equitative alternative	Preferences for distribution that benefit others
Selfishness	Zero to three choices of a beneficial alternative over equitative alternative	Preferences for distribution that benefit herself
<b>INTELLIGENCE</b>		
<b>Psychology</b>		
Raven	Reduced version of the Raven Test in this case subjects have to choose only 18 questions	Fluid intelligence
Cognitive reflection test (CRT)	Extended version of the CRT described by Toplak et al. (2011)	Reflection and impulsiveness
Wonderlic	Reduced version of the Wonderlic Personnel Test, 24 questions with a maximum of 6 min	
<b>Numeracy</b>		
Adding task correct (incentivized)	Individual has to add 10 sequences of summations and its pay for each correct addition	Numeracy capabilities with extrinsic motivation
Adding task (no incentivized)	Similar to the incentivized task but individuals are not paid by correctness	
<b>Academic</b>		
SAT	Self-reported	
ACT	Self-reported	
GPA	Self-reported	
<b>Other</b>		
The reading the mind in the eyes test	A sample test in which individuals were requested to guess the most likely emotion of 36 pictures of eyes	Theory of mind
<b>PERSONALITY—BIG FIVE</b>		
Openness	Big five questionnaire with 44 items, John et al., 2008	Include traits of appreciation for unusual ideas, curiosity, and variety of experience
Conscientiousness	Big five questionnaire with 44 items, John et al., 2008	Tendency to be organized, disciplined, dependable, and to prefer planned behavior
Extraversion	Big five questionnaire with 44 items, John et al., 2008	Tendency to seek stimulation by the company of others, to be talkative, energetic, and assertive

(Continued)

**TABLE A1 | Continued**

<b>Measure</b>	<b>Method–Test (Reference)</b>	<b>Related elicited characteristic</b>
Agreeableness	Big five questionnaire with 44 items, John et al., 2008	Tendency to be sympathetic, compassionate and cooperative, kind, and affectionated
Neuroticism	Big five questionnaire with 44 items, John et al., 2008	Tendency to be moody, and to experience easily emotions such as anger, anxiety, and depression
<b>DEMOGRAPHICS</b>		
Age	Self-reported	Numeric
Gender	Self-reported	Male or female
Volunteer hours	Self-reported	Range of number of hours allocated to voluntary work or N/A
Work hours	Self-reported	Range of hours allocated to remunerated work
Number of sibling	Self-reported	Range of number of hours allocated to remunerated work or NA
Older sibling	Self-reported	Number of siblings older than the subject
Religiosity	Self-reported	Range of the frequency of service attendance



# Motivational Hierarchy in the Chinese Brain: Primacy of the Individual Self, Relational Self, or Collective Self?

Xiangru Zhu<sup>1\*</sup>, Haiyan Wu<sup>2</sup>, Suyong Yang<sup>3</sup> and Ruolei Gu<sup>2</sup>

<sup>1</sup> Institute of Cognition and Behavior, Henan University, Kaifeng, China, <sup>2</sup> Institute of Psychology, Chinese Academy of Sciences, Beijing, China, <sup>3</sup> Department of Psychology, Shanghai University of Sport, Shanghai, China

According to the three-tier hierarchy of motivational potency in the self system, the self can be divided into individual self, relational self, and collective self, and individual self is at the top of the motivational hierarchy in Western culture. However, the motivational primacy of the individual self is challenged in Chinese culture, which raises the question about whether the three-tier hierarchy of motivational potency in the self system can be differentiated in the collectivist brain. The present study recorded the event-related potentials (ERPs) to evaluate brain responses when participants gambled for individual self, for a close friend (relational self), or for the class (collective self). The ERP results showed that when outcome feedback was positive, gambling for individual self evoked a larger reward positivity compared with gambling for a friend or for the class, while there is no difference between the latter two conditions. In contrast, when outcome feedback was negative, no significant effect was found between conditions. The present findings provide direct electrophysiological evidence that individual self is at the top of the three-tier hierarchy of the motivational system in the collectivist brain, which supports the classical pancultural view that individual self has motivational primacy.

**Keywords:** self, motivation, decision making, event-related potential (ERP), feedback-related negativity (FRN)

## OPEN ACCESS

### Edited by:

Nikolaos Georgantzas,  
University of Reading, UK

### Reviewed by:

Michael E. W. Varnum,  
Arizona State University, USA  
Yanhong Wu,  
Peking University, China

### \*Correspondence:

Xiangru Zhu  
zhuxiangru@gmail.com

### Specialty section:

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

**Received:** 23 February 2016

**Accepted:** 27 May 2016

**Published:** 13 June 2016

### Citation:

Zhu X, Wu H, Yang S and Gu R  
(2016) Motivational Hierarchy  
in the Chinese Brain: Primacy of the  
Individual Self, Relational Self, or  
Collective Self? *Front. Psychol.* 7:877.  
doi: 10.3389/fpsyg.2016.00877

## HIGHLIGHTS

- A three-tier hierarchy of motivational potency exists in the self system.
- Evidence of the feedback-related negativity indicated that the individual self is at the top of the three-tier hierarchy in the collectivist brain.

## INTRODUCTION

The concept of the self occupies a central role in psychological theory, partly because of its relevance to cognitive, motivational, affective, and behavioral processes (Leary, 2007). The concept of the self is not a unitary phenomenon. Indeed, researchers have generally divided the self into individual self, relational self, and collective self (Greenwald and Pratkanis, 1984; Breckler and Greenwald, 1986; Triandis, 1989; Brewer and Gardner, 1996; Brewer and Chen, 2007). The individual self reflects cognitions that are related to traits, states, and behaviors that are stored in memory (e.g., "I am honest"). The relational self reflects cognitions that are related to one's relationships

(e.g., "I am a son"). The collective self reflects cognitions that are related to one's groups (e.g., "I am Chinese"). The three kinds of selves are all necessary and are associated with psychological and physical health benefits. However, they are not equally important or meaningful. That is to say, one of them might be closer to the motivational core of the self-concept than the others. To provide a comprehensive understanding of the motivational hierarchy among the three kinds of selves, the present study evaluated the event-related potentials (ERPs) technique, combined with a gambling task to investigate the hierarchy of the self-motivation system in the collectivistic brain.

According to the three-tier hierarchy of motivational potency in the self-system, a series of experiments showed that the individual self is at the top of the motivational hierarchy, followed by the relational self and collective self (Sedikides et al., 2013). This idea has been confirmed by many studies (Gaertner et al., 1999, 2012). Gaertner et al. (2012) used the money allocation task and instructed the subjects to list goals for each self, they further employed groups of Chinese participants and found that the three-tier hierarchy applied to both Western (United States) and Eastern (Chinese) subjects. Consistent with this view, Abdulkarim et al. (2015) used the method of the Twenty Statements Test and found the individual self is prominent compared with the relational self and collective self. These studies indicate that the primacy of the individual self is a universal phenomenon across cultural groups.

Nevertheless, some studies found that the motivational hierarchy systems are modulated by culture (Han et al., 2013; Kitayama and Park, 2014). Research on independent vs. interdependent self-construals is a prominent topic in social psychology. According to Markus and Kitayama (1991), the Western independent self is characterized as a self-contained and autonomous entity that is context independent and possesses salient internal attributes. The Eastern interdependent self, however, is treated as a member in a group and highlights personal belonging and dependence upon a context. Chinese self, but not the Western self, may include significant others. Indeed, other research revealed a different motivational hierarchy in Chinese people. For instance, by comparing the importance that Han ethnic groups placed on the three types of self, two studies found that relational self and private self in Han participants were ranked similarly, and both were more important than collective self (Huang et al., 2014; Mamat et al., 2014). The motivational hierarchy manifests itself not only in behavioral patterns but also in neural and electrocortical activities. Our previous study used a gambling paradigm and ERP technique. The feedback related negativity (FRN) results showed that the self and mother have the same motivational hierarchy in the Chinese brain (Zhu et al., 2015b). Another study found that friends also gain the same status in a self-motivation system (Kitayama and Park, 2014).

Given these inconsistent behavioral findings and the collectivist characteristics of Chinese culture, the role of the cultural factor deserves to be further explored when investigating motivational hierarchy in the Chinese brain. First, we aimed to explore whether friend has the same motivational hierarchy. According to Cai et al. (2013), the relational self can be subdivided into the familial self (involving family bonds) and the

close other self (involving connections with a friend or romantic partner). Previous behavioral studies found that Chinese were closer to their parents, but friends were less important than their parents (Li, 2002; Cai et al., 2013). So we think that the status of friend is likely different from individual self and that of a family member. Second, previous behavioral studies found collective self is less important than relational self, but close other are confounded with family members in these studies. The present study aimed to compare the motivational hierarchy between close other and collective self.

The present study aims to explore potential electrocortical markers of the motivational hierarchy by examining the FRN. Feedback-related negativity is a key component of outcome evaluation, which is a medial frontal negative-going component that peaks approximately 250 ms following feedback presentation (Gehring and Willoughby, 2002). Localization studies suggest that the FRN is generated at the mPFC (Cohen et al., 2011). The FRN is an effective neural marker to explore the self motivational hierarchy because it is sensitive to the motivational factor. Specifically, the FRN amplitude is widely considered as an index of the motivational significance of the current event (Gehring and Willoughby, 2002; Yeung and Sanfey, 2004; Yeung et al., 2005; Leng and Zhou, 2010). In addition, the FRN reflects a semi-automatic outcome evaluation process which is immune to social desirability bias and test anxiety that might either exaggerate or obscure cultural differences. Hence, the present study adopted the FRN to investigate the self motivational hierarchy in Chinese college students.

The FRN has typically been viewed as a negative deflection in the ERP waveform that increases for monetary loss and is either reduced or absent for monetary gain (Holroyd and Coles, 2002). However, an accumulating body of recent evidence suggests the opposite viewpoint, in which the FRN amplitude is largely modulated by neural activity in gain trials (for a review, see Proudfit, 2015). One proposal is that monetary gain feedback elicits a distinct positive-going deflection (Holroyd et al., 2008; Baker and Holroyd, 2011). This reward positivity directly reflects activity of the mesencephalic dopamine system (Baker and Holroyd, 2011), a neural network that is critically involved in reward processing (Schultz, 2002). Reframing FRN as a response to monetary gain (i.e., a neurobiological index of hedonic capacity) makes it well-suited for studying the motivational hierarchy in the motivational system. Indeed, in the loss domain, there is little room to be "worse than expected" because losses are already the worst outcome. A previous study found that participants were more sensitive to the win condition than to the loss condition (Yu and Zhang, 2014). Pathological gamblers manifest insensitivity to losses but hypersensitivity to wins (Hewig et al., 2010). In another study, a group of depressed individuals presented blunted responses to gain feedback compared with the control group, whereas no significant group difference emerged for loss feedback (Liu et al., 2014). Based on these data, we predicted that the influence of the motivational hierarchy on FRN would be significant in the win domain (feedback related positivity or reward positivity) but not in the loss domain.

To sum up, the present study examined the motivational hierarchy among the individual self, close other, and collective self. We compared the FRN associated with outcome evaluation using a simple gambling task. In each trial, the beneficiary could be the individual self, relational self, or the collective self. Our hypothesis was that if the individual self, relational self, and collective self have different motivational hierarchies, then the FRN amplitude should reflect the hierarchical structure, such that a larger reward positivity indicates a higher motivational hierarchy.

## MATERIALS AND METHODS

### Participants

Twenty one college students (all are Han people;  $21.4 \pm 0.8$  years of age; range, 20–24 years; 10 females) participated in the study. Informed consent was obtained prior to the study. The experiment was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Department of Psychology, Henan University, China. All of the participants had normal vision (with correction), and none had a history of neurological disease or brain injury. All of the participants were right-handed.

### Procedure

Before the simple gambling task, the participants selected a good friend (same sex but not romantic partner) to play for. In China, generally, dozens of students form a class, a class generally taking the same courses in 4 years. Each student affords a fixed amount money to establish the class fee. For the present study, participants come from different classes. Playing for class means that the money would be given to the class monitor and let all the class mates know this fact. The money should be used for class activities.

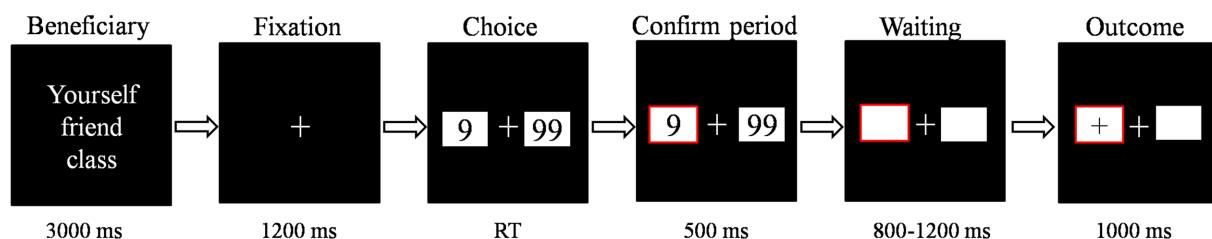
For the gambling task, the stimulus display and behavioral data acquisition were performed using E-Prime 1.1 software (Psychology Software Tools, Pittsburgh, PA, USA). During the task, the participants sat comfortably in an electrically shielded room approximately 80 cm from a computer screen. Each trial began with a 3000 ms presentation of the person for whom

the participant was playing (i.e., “for yourself,” “for your friend” and “for your class”). Two white rectangles ( $2.5^\circ \times 2.5^\circ$  of visual angle) were then presented that contained two Arabic numerals (9 and 99) to indicate two alternative options on the left and right sides of a fixation point on the computer screen. The positions of the two numbers were counterbalanced across trials. The participants were asked to make a selection by pressing the “F” or “J” key on the keyboard with the left or right index finger, respectively. The alternatives remained on the screen until the participant chose one of the rectangles, which was then highlighted by a thick red outline for 500 ms. After a subsequent interval of 800–1200 ms, the participants received feedback, lasting 1000 ms, which indicated whether he/she gained (when the valence of the outcome was “+”) or lost (when the valence of the outcome was “−”) in that particular trial (see Figure 1). The formal task consisted of six blocks of 64 trials each. Unbeknownst to the participants, the outcomes were provided according to a predetermined pseudorandom sequence, and each participant received exactly 64 of each kind of outcome for each beneficiary. Each participant was paid 15 CNY for their participation in the study. In the gambling task, each beneficiary had 15 CNY in his/her account. Based on the points gained for each beneficiary, the final gain or loss was added to the separate account (every additional 500 points gained increase payment 5 CNY). The total payment for each participant was approximately 60.6 CNY (range, 4075 CNY;  $SD = 5.6$  CNY).

Before the experiment, each participant was instructed about the rules and meaning of the symbols in the task. The participants were instructed that the money would be put on the friend’s cell phone or served as class fee. The participants were also encouraged to respond in such a way to maximize the total amount for each beneficiary. The participants were told that the higher the amount earned for each beneficiary, the more bonus money the beneficiary would receive at the end of the study. After the participant finished the task, he/she was told that the task had no optimal strategy.

### Electrophysiological Recording and Measures

Electroencephalographic (EEG) activity was recorded from 63 scalp sites using tin electrodes mounted in an elastic cap



**FIGURE 1 | The sequence of events within a single trial in the monetary gambling task.** In each trial, the beneficiary information lasted for 3000 ms then the fixation point lasted for 1200 ms. The participant was then presented with a choice of two alternatives, and the participant responded using the left or right index finger. The alternatives remained until the participant made his/her choice. Afterward, his/her choice was highlighted for 500 ms. After a subsequent interval of 800–1200 ms, the participant received feedback, lasting 1000 ms, which indicated whether he/she gained or lost in that trial.

(Brain Products, Gilching, Germany) with an online reference to the middle at FCz at the standard locations according to the international 10–20 system and off-line re-referenced to the average reference. The horizontal electrooculogram (HEOG) was recorded from an electrode placed at the outer canthi of the right eye. The vertical electrooculogram (VEOG) was recorded from an electrode placed above the left eye. All inter-electrode impedance was maintained at <10 kΩ. The EEG and EOG signals were amplified with a bandpass filter from 0.05 to 100 Hz and continuously sampled at 500 Hz/channel.

Off-line analysis of the EEG was performed using Brain Vision Analyzer software (Brain Products). The first step in data preprocessing was the correction of ocular artifacts using Independent Component Analysis (ICA) of the continuous data using Brain Vision Analyzer 2.0 software. The ocular artifact-free EEG data were low-pass-filtered below 30 Hz (12 dB/oct) and high-pass-filtered above 0.1 Hz (12 dB/oct). Separate EEG epochs of 1000ms (200 ms baseline) were extracted offline for the stimuli. All of the trials in which EEG voltages exceeded a threshold of  $\pm 75 \mu\text{V}$  during the recording epoch were excluded from the analysis ( $\sim 7$  trials per individual were excluded).

Through visual detection on the grand-averaged waveform, the FRN amplitude was measured for each participant as the average amplitude within the 220320 ms window (Boksem et al., 2012; Zhu et al., 2015a). The time window was extracted in a window extending 50 ms before and 50 ms after the peak latency. The electrodes at the mid-frontal region were selected for detecting the FRN (Frömer et al., 2016). Accordingly, the FRN amplitudes were entered into a 2 (feedback valence: win and loss)  $\times$  3 (beneficiary: individual self, friend and class)  $\times$  8 (electrodes: Fz, F1, F2, FC1, FC2, C1, C2, and Cz) repeated-measures analysis of variance (ANOVA).

## RESULTS

### Behavioral Results

We defined the choice of '9' to be the risk-avoidant choice in our experiment, predicting that participants would make this choice to avoid the possibility of a large loss ('−99'). However, by making this choice, they also gave up the opportunity to receive the larger reward ('+99'). In contrast, the choice of '99' was defined as the risky choice (high-risk or high-return).

For the number of risky choice, the one-way repeated-measures ANOVA revealed no significant main effect of beneficiary (individual self, friend, and class), [ $F(2,40) = 2.44$ ,  $P = 0.11$ ,  $\eta^2 = 0.13$ ]. For the RT (response time) data, the one way ANOVA revealed neither significant main effect nor and interaction effect,  $Ps > 0.10$ .

### ERP Results

The main effect of feedback valence was significant [ $F(1,20) = 136.70$ ,  $P < 0.001$ ,  $\eta^2 = 0.87$ ], such that the FRN was more negative after losses ( $M = 2.09 \mu\text{V}$ ,  $SE = 0.43$ ) than after gains ( $M = 4.66 \mu\text{V}$ ,  $SE = 0.54$ ). The main effect of electrode on the FRN amplitude was also significant [ $F(7,140) = 22.89$ ,  $P < 0.001$ ,  $\eta^2 = 0.53$ ], with a largest amplitude at Cz site.

The interaction between feedback valence and beneficiary was significant [ $F(2,40) = 4.09$ ,  $P = 0.03$ ,  $\eta^2 = 0.17$ ]. Simple effect analysis indicated that only in the win condition the effect of beneficiary was significant. Pairwise comparison revealed that winning for individual self ( $M = 5.40 \mu\text{V}$ ,  $SE = 0.56$ ) was larger than winning for friend ( $M = 4.23 \mu\text{V}$ ,  $SE = 0.55$ ) and winning for class ( $M = 4.36 \mu\text{V}$ ,  $SE = 0.59$ ) ( $P = 0.01$ ,  $P = 0.009$ ) (Figure 2). No significant difference existed between the latter two conditions. Neither the main effect of beneficiary nor other interactions were significant (all  $Ps > 0.05$ ).

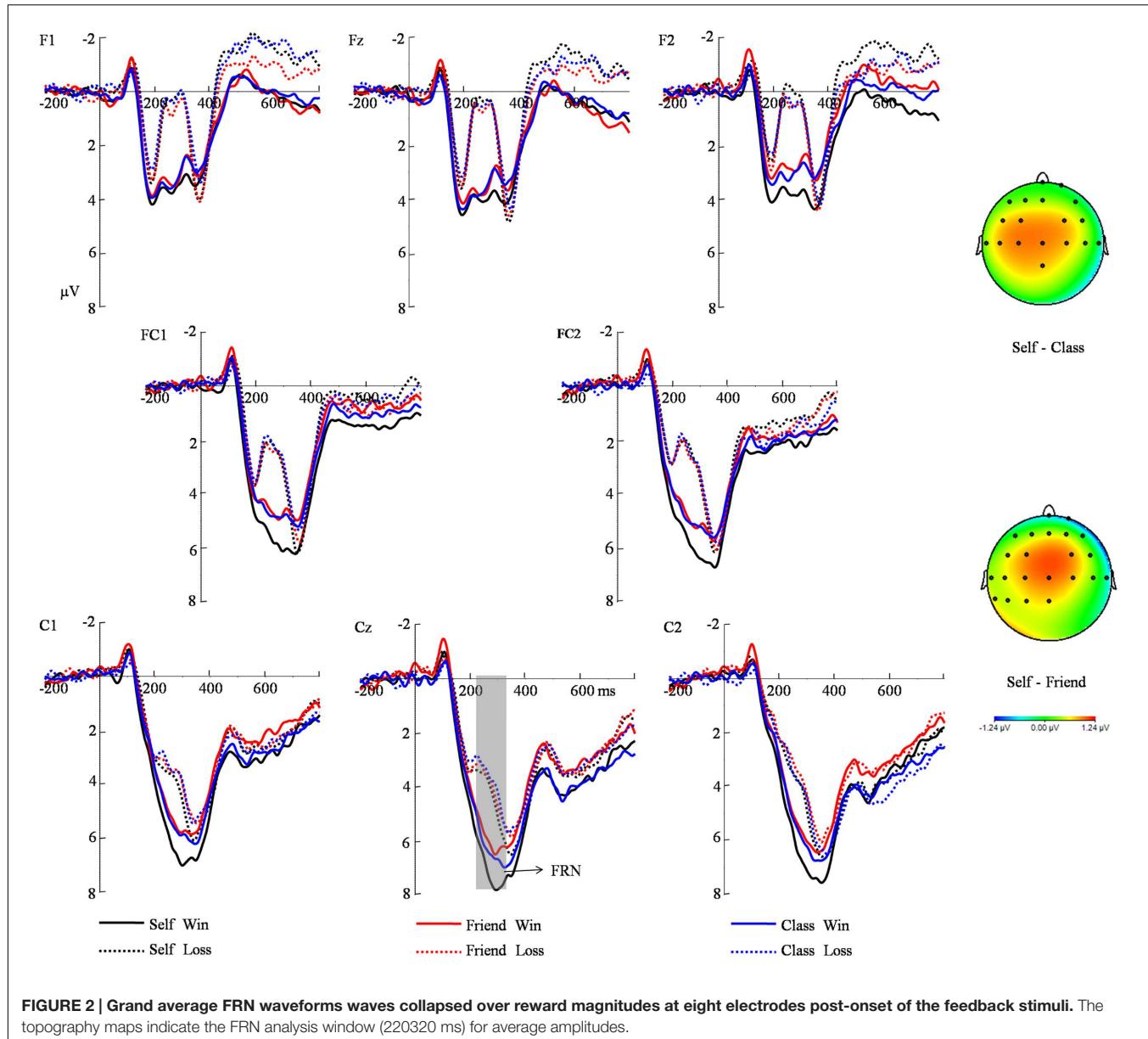
## DISCUSSION

The present study investigated ERP responses to reward in a social context, in which the individual self, relational self, and collective self were the beneficiaries. Our main findings were threefold. First, behaviorally, no differences existed among the three kinds of selves. Second, the results replicated the well-established ERP patterns whereby win evoked larger reward positivity than loss in the gambling task. Third and most importantly, reward positivity was the largest when gambling for the individual self than for the relational or collective self, with no difference between the relational self and collective self. The present FRN results clearly support the pancultural view that the individual self is at the top of the motivational hierarchy.

The present results are consistent with the findings of previous studies (Gaertner et al., 2012; Abdukeram et al., 2015). Gaertner et al. (2012) reported that participants from China allocated more money to the individual self than to the relational self and collective self, indicating that the individual self was rated as most important in the self motivational system. Abdukeram et al. (2015) found relational aspect of an individual's self became increasingly important with age in the Han cultural groups, but individual self still top the motivational hierarchy in 1024 years old participants.

However, the present results are inconsistent with Huang et al. (2014). In their study, participants were asked to write down five personal characteristics, five personal relationships, and five group memberships and then evaluate the importance they tie to each of them. As we pointed out in the introduction, the personal relationship may include family member and close others (friend or romantic partner). Given the important status of family member (Zhu et al., 2015b), it is likely to find no significant different between individual self and relational self.

Although Gaertner et al. (2012) proposed that the collective self is at the bottom of the motivational hierarchy, considerable uncertainty remains in the relative positioning of the relational and collective selves in Eastern cultures. One view posits that both selves rely on norms of interdependence, connectedness, and the importance of others and therefore might have equivalent motivational potency (Brewer and Chen, 2007). According to another view, collective behavior indicates that Eastern culture is more represented by interpersonal relationships that are internalized as the relational self than by in-group-associations that are internalized as the collective self, thus implying the relative primacy of the relational self (Yuki, 2003). In the present



**FIGURE 2 |** Grand average FRN waveforms waves collapsed over reward magnitudes at eight electrodes post-onset of the feedback stimuli. The topography maps indicate the FRN analysis window (220–320 ms) for average amplitudes.

study, the relational self and collective self did not have different FRN. One potential reason is that friendship can be fleeting and depends largely on reciprocal exchange, therefore friend is not one of the key embeddedness in relational self. Another possible reason is that we used the participant's class to represent the collective self. Participant may involve considerable dyadic relationships between the self and class, lead to the boundaries are not so obvious. Remaining unclear is whether differences between the relational self and collectivist self would become evident if we use a more abstract and important collective self.

In the present study, the motivational hierarchy of a friend was lower than the individual self. Notably, however, this motivational hierarchy is not absolute. Generally, the union with a close other, such as a friend, in Chinese culture is thought to be tight, and friends are also deeply ingrained in

the self motivational system. For example, Kitayama and Park (2014) used error-related negativity (ERN) as a motivational neurological marker and found that it differentiated the self and friends in Western culture but not in East Asian culture. Two methodological differences that may account for this discrepancy. First, the beneficiary effect only manifested in the win condition but not in the loss condition, this result reflects dopaminergic signals response to positive outcomes (Baker and Holroyd, 2011), whereas ERN is thought to index the negative reward prediction errors that are based on a computation of an incorrect response as being worse than a correct response. Another reason is the speeded conflict task (flanker task) may be particularly likely to produce anxiety for Asians because this task is akin to an intelligence test. This anxiety may eliminate the difference between self and friend. Whereas the participants in the present

study were presumed to feel safe while performing the gambling task (Hitokoto et al., 2016).

It should be noted that the self motivational hierarchy is not immune to the transient effect of temporal priming. For example, one previous recent fMRI study found that Chinese participants primed with independent self-construal showed stronger activations in the ventral striatum in response to winning money for the self than for a close friend, while those primed with interdependence self-construal showed comparable activations in two conditions (Varnum et al., 2014). This fMRI result indicates that self-construal could shapes self motivational hierarchy in a highly dynamic fashion.

In the present study, the ERP results indicated that individual self is on top of the motivational hierarchy, but the behavioral results revealed no motivational hierarchy. To explain this discrepancy, it is worth noting that behavioral research on the motivational hierarchy, which provides most of what we know about the three-tier hierarchy, are not immune to social desirability bias, because respondents are tend to answer in a socially acceptable way (van de Mortel, 2008). This social desirability bias may threaten the validity of the behavioral measures of motivational hierarchy accordingly. In contrast, neural measurements may provide more insights than behavioral methods. For example, in the study of Wang et al. (2012), behavioral questionnaires showed that the intimacy level of self-mother relationship and that of self-father relationship were not significantly different, but different neural representations of mother and father in the medial prefrontal cortex (mPFC) have been observed. Future studies that recruit alternative behavioral measures and neural markers should be conducted to examine our hypothesis.

One limitation is that we only included Han people in the present study. Although Chinese culture has been characterized as an interdependent culture, it has a certain degree of heterogeneity. Three recent studies considered intra-cultural variability in the self motivational hierarchy in China (Huang

et al., 2014; Mamat et al., 2014; Abdukeram et al., 2015). Mamat et al. (2014) found that Uyghur Chinese rated the collective self as more important than the individual self and relational self. This was likely because the Uyghur culture is based on Islam, which emphasizes the solidarity of all Muslims. Their shared religion facilitates group integration, unity, and cohesiveness within the Uyghur ethnic group (Abdukern et al., 2015). Future research that is devoted to exploring the motivational hierarchy should consider the intra-cultural variability of interdependent self-construal in Chinese populations. Another limitation is that we only employed Chinese participants in the present study, it would be advantageous if future research compares Chinese with western cultures to further explore how culture factor modulates motivational hierarchy.

## CONCLUSION

The FRN response to losses and gains in the gambling task provided electrocortical evidence that the individual self is at the top of the self motivational hierarchy in the Chinese brain, which supports the pancultural view that the individual self is more important than close other and collective self in the human motivational system.

## AUTHOR CONTRIBUTIONS

XZ and RG designed research; XZ performed research; XZ analyzed data; HW and SY contributed analytic tools; XZ wrote the paper.

## ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation of China (31300846).

## REFERENCES

- Abdukeram, Z., Mamat, M., Luo, W., and Wu, Y. (2015). Influence of culture on tripartite self-concept development in adolescence: a comparison between han and uyghur cultures. *Psychol. Rep.* 116, 292–310. doi: 10.2466/17.07.PR0.116k12w8
- Baker, T., and Holroyd, C. (2011). Dissociated roles of the anterior cingulate cortex in reward and conflict processing as revealed by the feedback error-related negativity and N200. *Biol. Psychol.* 87, 25–34. doi: 10.1016/j.biopsych.2011.01.010
- Boksem, M. A. S., Kostermans, E., Milivojevic, B., and De Cremer, D. (2012). Social status determines how we monitor and evaluate our performance. *Soc. Cogn. Affect. Neurosci.* 7, 304–313. doi: 10.1093/scan/nsr010
- Breckler, S. J., and Greenwald, A. G. (1986). “Motivational facets of the self,” in *Handbook of Motivation and Cognition*, eds R. M. Sorrentino and E. T. Higgins (New York, NY: Guilford Press), 145–164.
- Brewer, M. B., and Chen, Y. (2007). Where (who) are collectives in collectivism? Toward conceptual clarification of individualism and collectivism. *Psychol. Rev.* 114, 133–151. doi: 10.1037/0033-295X.114.1.133
- Brewer, M. B., and Gardner, W. (1996). Who is this “we”? Levels of collective identity and self representations. *J. Pers. Soc. Psychol.* 71, 83–93. doi: 10.1037/0022-3514.71.1.83
- Cai, H., Sedikides, C., and Jiang, L. (2013). Familial self as a potent source of affirmation: evidence from China. *Soc. Psychol. Pers. Sci.* 4, 529–537. doi: 10.1177/1948550612469039
- Cohen, M. X., Wilmes, K., and van de Vijver, I. (2011). Cortical electrophysiological network dynamics of feedback learning. *Trends Cogn. Sci.* 15, 558–566. doi: 10.1016/j.tics.2011.10.004
- Frömer, R., Stürmer, B., and Sommer, W. (2016). The better, the bigger: the effect of graded positive performance feedback on the reward positivity. *Biol. Psychol.* 114, 61–68. doi: 10.1016/j.biopsych.2015.12.011
- Gaertner, L., Sedikides, C., and Graetz, K. (1999). In search of self-definition: motivational primacy of the individual self, motivational primacy of the collective self, or contextual primacy? *J. Pers. Soc. Psychol.* 76, 5–18. doi: 10.1037/0022-3514.76.1.5
- Gaertner, L., Sedikides, C., Luke, M., O’Marac, E. M., Iuzzinio, J., Jackson, L. E., et al. (2012). A motivational hierarchy within: primacy of the individual self, relational self, or collective self? *J. Exp. Soc. Psychol.* 48, 997–1013. doi: 10.1016/j.jesp.2012.03.009

- Gehring, W. J., and Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 295, 2279–2282. doi: 10.1126/science.1066893
- Greenwald, A. G., and Pratkanis, A. R. (1984). “The self,” in *Handbook of Social Cognition*, eds R. S. Wyer and T. K. Srull (Hillsdale, NJ: Erlbaum), 129–178.
- Han, S., Northoff, G., Vogeley, K., Wexler, B. E., Kitayama, S., and Varnum, M. E. W. (2013). A cultural neuroscience approach to the biosocial nature of the human brain. *Annu. Rev. Psychol.* 64, 335–359. doi: 10.1146/annurev-psych-071112-054629
- Hewig, J., Kretschmer, N., Trippe, R., Hecht, H., Coles, M., Holroyd, C., et al. (2010). Hypersensitivity to reward in problem gamblers. *Biol. Psychiatry* 67, 781–783. doi: 10.1016/j.biopsych.2009.11.009
- Hitokoto, H., Glaser, J., and Kitayama, S. (2016). Cultural shaping of neural responses: feedback-related potentials vary with self-construal and face priming. *Psychophysiology* 53, 52–63. doi: 10.1111/psyp.12554
- Holroyd, C., Pakzad-Vaezi, K., and Krigolson, O. (2008). The feedback correct-related positivity: sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology* 45, 688–697. doi: 10.1111/j.1469-8986.2008.00668.x
- Holroyd, C. B., and Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* 109, 679–709. doi: 10.1037/0033-295X.109.4.679
- Huang, W., Mamat, M., Shang, R., Zhang, T., Li, H., Wang, Y., et al. (2014). Analysis of the private, collective, and relational self-cognitions among Han and Tibetan Chinese. *Psychol. Rep.* 115, 179–198. doi: 10.2466/07.17.21.PR0.115c12z9
- Kitayama, S., and Park, J. (2014). Error-related brain activity reveals self-centric motivation: culture matters. *J. Exp. Psychol. Gen.* 143, 62–70. doi: 10.1037/a0031696
- Leary, M. R. (2007). Motivational and emotional aspects of the self. *Annu. Rev. Psychol.* 58, 317–344. doi: 10.1146/annurev.psych.58.110405.085658
- Leng, Y., and Zhou, X. (2010). Modulation of the brain activity in outcome evaluation by interpersonal relationship: an ERP study. *Neuropsychologia* 48, 448–455. doi: 10.1016/j.neuropsychologia.2009.10.002
- Li, H. Z. (2002). Culture, gender and self-close-other(s) connectedness in Canadian and Chinese samples. *Eur. J. Soc. Psychol.* 32, 93–104. doi: 10.1002/ejsp.63
- Liu, W., Wang, L., Shang, H., Shen, Y., Li, Z., Cheung, E. F. C., et al. (2014). The influence of anhedonia on feedback negativity in major depressive disorder. *Neuropsychologia* 53, 213–220. doi: 10.1016/j.neuropsychologia.2013.11.023
- Mamat, M., Huang, W., Shang, R., Zhang, T., Li, H., Wang, Y., et al. (2014). Relational self versus collective self: a cross-cultural study in interdependent self-construal between Han and Uyghur in China. *J. Cross Cult. Psychol.* 45, 959–970. doi: 10.1177/0022022114530558
- Markus, H. R., and Kitayama, S. (1991). Culture and the self: implications for cognition, emotion, and motivation. *Psychol. Rev.* 98, 224–253. doi: 10.1037/0033-295X.98.2.224
- Proudfoot, G. H. (2015). The reward positivity: from basic research on reward to a biomarker for depression. *Psychophysiology* 52, 449–459. doi: 10.1111/psyp.12370
- Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron* 36, 241–263. doi: 10.1016/S0896-6273(02)00967-4
- Sedikides, C., Gaertner, L., Luke, M. A., O’Mara, E. M., and Gebauer, J. E. (2013). Chapter five—a three-tier hierarchy of self-potency: individual self, relational self, collective self. *Adv. Exp. Soc. Psychol.* 48, 235–295. doi: 10.1016/B978-0-12-407188-9.00005-3
- Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychol. Rev.* 96, 506–520. doi: 10.1037/0033-295X.96.3.506
- van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report research. *Aust. J. Adv. Nurs.* 25, 40–48.
- Varnum, M. E. W., Shi, Z., Chen, A., Qiu, J., and Han, S. (2014). When “your” reward is the same as “my” reward: self-construal priming shifts neural responses to own vs. friends’ rewards. *Neuroimage* 87, 164–169. doi: 10.1016/j.neuroimage.2013.10.042
- Wang, G., Mao, L., Ma, Y., Yang, X., Cao, J., Liu, X., et al. (2012). Neural representations of close others in collectivistic brains. *Soc. Cogn. Affect. Neurosci.* 7, 222–229. doi: 10.1093/scan/nsr002
- Yeung, N., Holroyd, C. B., and Cohen, J. D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cereb. Cortex* 15, 535–544. doi: 10.1093/cercor/bhh153
- Yeung, N., and Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *J. Neurosci.* 24, 6258–6264. doi: 10.1523/JNEUROSCI.4537-03.2004
- Yu, R., and Zhang, P. (2014). Neural evidence for description dependent reward processing in the framing effect. *Front. Neurosci.* 8:56. doi: 10.3389/fnins.2014.00056
- Yuki, M. (2003). Intergroup comparison versus intragroup relationships: a cross-cultural examination of social identity theory in North American and East Asian cultural contexts. *Soc. Psychol. Q.* 66, 166–183. doi: 10.2307/1519846
- Zhu, X., Gu, R., Wu, H., and Luo, Y. (2015a). Self-reflection modulates the outcome evaluation process: evidence from an ERP study. *Int. J. Psychophysiol.* 98, 389–393. doi: 10.1016/j.ijpsycho.2015.08.001
- Zhu, X., Zhang, Y., Yang, S., Wu, H., Wang, L., and Gu, R. (2015b). The Motivational Hierarchy between Self and Mother: evidence from the feedback-related negativity. *Acta Psychol. Sin.* 47, 807–813.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Zhu, Wu, Yang and Gu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Prosocial Personality Traits Differentially Predict Egalitarianism, Generosity, and Reciprocity in Economic Games

Kun Zhao<sup>1\*</sup>, Eamonn Ferguson<sup>2</sup> and Luke D. Smillie<sup>1</sup>

<sup>1</sup> Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, VIC, Australia, <sup>2</sup> School of Psychology, University of Nottingham, Nottingham, UK

## OPEN ACCESS

**Edited by:**

Manuel Ignacio Ibáñez,  
Jaume I University, Spain

**Reviewed by:**

Isabel Thielmann,  
University of Koblenz-Landau,  
Germany  
Joachim Israel Krueger,  
Brown University, USA

**\*Correspondence:**

Kun Zhao  
kun.zhao@unimelb.edu.au

**Specialty section:**

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

**Received:** 28 May 2016

**Accepted:** 18 July 2016

**Published:** 09 August 2016

**Citation:**

Zhao K, Ferguson E and Smillie LD  
(2016) Prosocial Personality Traits  
Differentially Predict Egalitarianism,  
Generosity, and Reciprocity  
in Economic Games.  
*Front. Psychol.* 7:1137.

doi: 10.3389/fpsyg.2016.01137

Recent research has highlighted the role of prosocial personality traits—agreeableness and honesty-humility—in egalitarian distributions of wealth in the dictator game. Expanding on these findings, we ran two studies to examine individual differences in two other forms of prosociality—generosity and reciprocity—with respect to two major models of personality, the Big Five and the HEXACO. Participants (combined  $N = 560$ ) completed a series of economic games in which allocations in the dictator game were compared with those in the generosity game, a non-constant-sum wealth distribution task where proposers with fixed payoffs selected the size of their partner's payoff ("generosity"). We further examined positive and negative reciprocity by manipulating a partner's previous move ("reciprocity"). Results showed clear evidence of both generosity and positive reciprocity in social preferences, with allocations to a partner greater in the generosity game than in the dictator game, and greater still when a player had been previously assisted by their partner. There was also a consistent interaction with gender, whereby men were more generous when this was costless and women were more egalitarian overall. Furthermore, these distinct forms of prosociality were differentially predicted by personality traits, in line with the core features of these traits and the theoretical distinctions between them. HEXACO honesty-humility predicted dictator, but not generosity allocations, while traits capturing tendencies toward irritability and anger predicted lower generosity, but not dictator allocations. In contrast, the politeness—but not compassion—aspect of Big Five agreeableness was uniquely and broadly associated with prosociality across all games. These findings support the discriminant validity between related prosocial constructs, and have important implications for understanding the motives and mechanisms taking place within economic games.

**Keywords:** dictator game, social preferences, honesty-humility, agreeableness, politeness, compassion, big five, HEXACO

## INTRODUCTION

One of the major themes in the literature on economic games is that humans care about and are motivated by the interests of others. These other-regarding or social preferences are the building blocks of prosocial behavior and have been incorporated into various economic models (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002; Falk and Fischbacher, 2006). A second major theme to emerge from this literature is the substantial heterogeneity in people's social preferences and behaviors despite being exposed to the same experimental conditions (Fischbacher et al., 2001; Camerer, 2003; Fehr and Schmidt, 2006). Measures of social value orientation, which capture motivational differences in the distribution of resources, reveal a variety of archetypes, including altruistic, prosocial, individualistic, and competitive (Murphy and Ackermann, 2014). Recent studies have also documented stable patterns of prosocial behavior correlated over time and across different games (Yamagishi et al., 2013; Peysakhovich et al., 2014).

One potential source of this heterogeneity rests in broad dispositions capturing consistent and enduring patterns in behavior and experience. Specifically, *personality traits* are "probabilistic descriptions of relatively stable patterns of emotion, motivation, cognition, and behavior, in response to classes of stimuli that have been present in human cultures over evolutionary time" (DeYoung, 2015, p. 35). A long line of research has documented how basic prosocial personality traits—known as agreeableness and honesty-humility—contribute to experimental and real-world instances of prosociality, including helping, volunteering, charitable giving, and ethical decision making (e.g., Elshaug and Metzer, 2001; Carlo et al., 2005; Penner et al., 2005; Ozer and Benet-Martínez, 2006; Graziano et al., 2007; Ashton and Lee, 2008; Aghababaei et al., 2014). It is not surprising, then, that the same prosocial traits have begun to emerge as significant predictors of inequality aversion, egalitarianism, and fairness in economic games (Hilbig et al., 2014; Zhao et al., 2016; for a review, see Zhao and Smillie, 2015).

In the current paper, we extend this nascent literature by applying a framework of distinct prosocial traits to a broader range of social preferences beyond egalitarianism. We first present an overview of the prosocial domains of major personality models and discuss their relevance for distributive and reciprocal preferences in economic games. Building on the design of the traditional dictator game, we develop a novel paradigm that simultaneously tests for two other forms of social preference beyond egalitarianism: generosity and reciprocity.

## Prosocial Domains of Major Personality Models

Prosociality is a general term referring to a variety of positive emotions, attitudes, and behaviors directed toward others, which may be manifested through acts of sharing, helping, and cooperating (Knafo-Noam et al., 2015). There is increasing recognition that neither prosociality nor its underlying motivations are unitary constructs (Batson and Powell, 2003;

Singer and Steinbeis, 2009; Böckler et al., 2016). Likewise, there are multiple prosocial tendencies, which are classified differently according to two major taxonomic models of personality, the Big Five (Goldberg, 1981; Digman, 1990; John et al., 2008; DeYoung, 2015) and the HEXACO (Honesty-Humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, Openness to Experience; Lee and Ashton, 2004).

### Prosocial Domains of the Big Five: Agreeableness and Its Aspects of Politeness and Compassion

The Five-Factor Model or "Big Five" is a robust hierarchical taxonomy of personality dimensions recovered from a number of measures of trait descriptors (John et al., 2008) and replicable across languages and cultures (Digman, 1990). Each factor represents a major dimension of covariation among traits, subsuming a number of narrower personality characteristics at intermediate (known as *aspects*) and lower (known as *facets*) levels (DeYoung et al., 2007; DeYoung, 2015).

Within the Big Five model, agreeableness captures tendencies toward altruism and cooperation, and has a core underlying motivation of maintaining interpersonal harmony (Graziano and Eisenberg, 1997). Consistent with this, agreeableness is the Big Five dimension most frequently associated with prosocial behaviors in a variety of economic games, including allocations of wealth in the dictator game (Ben-Ner et al., 2004; Becker et al., 2012; Baumert et al., 2014), acceptance of unfair offers in the ultimatum game (Mehta, 2007), cooperation in the prisoner's dilemma (Kagel and McGee, 2014), contributions in the public goods game (Volk et al., 2011), and amounts invested and returned to others in the trust game (Evans and Revelle, 2008; Becker et al., 2012; for a review, see Zhao and Smillie, 2015).

However, Big Five agreeableness is a broad domain of personality which can be divided into two distinct aspects: *politeness*, the tendency to respect others, adhere to social norms, and suppress aggressive impulses, and *compassion*, the tendency to be emotionally concerned about others (DeYoung et al., 2007; DeYoung, 2015). Though correlated, the two often show diverging associations with other individual differences. For instance, while politeness is associated with the moral foundation of authority/respect and political conservatism, compassion is more strongly linked with the moral foundations of harm/care and fairness/reciprocity, as well as political liberalism (Hirsh et al., 2010; Osborne et al., 2013). This distinction between politeness and compassion also has important implications for the study of heterogeneity in economic games, where prosocial behaviors in different games may stem from different motivations, such as adhering to normative rules around sharing and cooperating (e.g., the public goods game), or helping needy others (e.g., third party punishment and recompensation).

### Prosocial Domains of the HEXACO: Honesty-Humility and Agreeableness

A major alternative to the Big Five is the HEXACO model, a six-factor model of personality developed from psycholinguistic studies in European and Asian languages (Lee and Ashton, 2004; Ashton and Lee, 2007; Ashton et al., 2014). The most salient difference between the HEXACO and the Big Five is the addition of a

sixth dimension, honesty-humility, or the tendency to be sincere, modest, and fair, which is believed to capture trait variance beyond the Big Five. Moreover, the HEXACO representation of agreeableness and emotionality (neuroticism) are rotational variants of their Big Five counterparts (Ashton and Lee, 2007). Specifically, HEXACO agreeableness reflects the tendency to be patient, forgiving, and tolerant, and is thus non-interchangeable with Big Five agreeableness, which reflects broad tendencies toward altruism.

Together, HEXACO honesty-humility and HEXACO agreeableness span the prosocial domain typically captured by Big Five agreeableness and make up two forms of individual variation in reciprocal altruism. Honesty-Humility represents *active cooperation*, the tendency to cooperate with others despite the opportunity for exploitation, while HEXACO agreeableness represents *reactive cooperation*, the tendency to cooperate with others despite their misgivings (Hilbig et al., 2013; Ashton et al., 2014). The two diverge in studies of workplace delinquency (Lee et al., 2005), criminality (Rolison et al., 2013), dishonesty and cheating (Hilbig and Zettler, 2015), and forgiveness and revenge (Lee and Ashton, 2012). This discriminant validity is also relevant to behavior within economic games, where there is evidence of a “cooperative phenotype,” characterized by within-individual correlations across cooperative games (i.e., fair and cooperative tendencies corresponding to honesty-humility), which is independent from norm-enforcing punishment (i.e., retaliatory tendencies corresponding to HEXACO agreeableness; Peysakhovich et al., 2014).

In summary, the Big Five and HEXACO models provide an array of distinct prosocial traits which reflect different motivations and mechanisms, and which show divergent validity with respect to interpersonal and socio-political variables (see **Table 1**). We now turn to the experimental economics literature, where similar distinctions may exist between different facets of prosociality and which are expressed through multiple social preferences in games.

## Multiple Social Preferences in Economic Games

### Inequality Aversion and Egalitarianism

One basic way in which social preferences deviate from narrow self-interest is the desire for equality. Egalitarianism is a basic motivation that can be traced back to small-scale societies in human evolutionary history (Boehm, 1999) and is the cornerstone of economic theories of social preferences (Loewenstein et al., 1989; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). The tension between self-interest and equality is best captured in the dictator game, in which one player decides how to split a fixed amount of money with a second player, who must accept this unconditionally (Kahneman et al., 1986; Forsythe et al., 1994). Featuring in more than a hundred studies, the popularity of the dictator game owes to the fact that it is a simple yet powerful paradigm which yields considerable behavioral variation (Engel, 2011). While average allocations to a partner range between 20% and 30% of the pie, up to half of participants keep all the money, a quarter split it equally, and the

remainder select distributions in between (Tisserand et al., 2015). This heterogeneity thus makes the dictator game an ideal hunting ground for examining the influence of personality and for teasing apart the roles of similar but distinct personality constructs.

For example, Big Five agreeableness is a consistent predictor of egalitarian dictator allocations (for a review, see Zhao and Smillie, 2015). However, recent research indicates that this is driven by its aspect of politeness—or tendencies toward good manners and etiquette—rather than compassion (Zhao et al., 2016), in keeping with the economics literature on the importance of social norms for prosociality (Camerer and Thaler, 1995). Another kind of dissociation has emerged within the HEXACO model, with several studies showing that honesty-humility (or the tendency for active cooperation)—but not HEXACO agreeableness—is a strong, consistent, and robust predictor of egalitarian dictator allocations, and even more so than Big Five agreeableness (Hilbig and Zettler, 2009; Thielmann and Hilbig, 2014; Hilbig et al., 2015a, 2013; Zhao et al., 2016; for a review, see Zhao and Smillie, 2015).

### Costless Prosociality and Generosity

Despite the wealth of findings it has generated, the dictator game is limited when drawing inferences about a wider array of social preferences. Notably, the constant-sum structure of the game means that decisions to benefit one’s partner are always at a cost to self-interest by the same magnitude. However, many instances of real-world prosociality involve decisions which benefit others at minimal personal cost, such as giving pre-loved belongings to charity and posthumous organ donation (Saunders, 2012; Moorlock et al., 2014; Shepherd et al., 2014). In this paper, we use the term *generosity* to describe the willingness to accept a relative disadvantage when this makes others better off (either at a personal cost or at no cost), but it should not be confused with other usages in the literature (e.g., Haley and Fessler, 2005).

Acts of generosity are typically obscured by dominant norms of equality in constant-sum games, such as the dictator game, where fewer than 5% of individuals allocate more than half the endowment to their partner (Tisserand et al., 2015). However, acts of generosity emerge in tasks of costless prosociality where they may reflect concerns for efficiency and social welfare (Charness and Rabin, 2002; Fehr et al., 2008; Bartling et al., 2009; Güth, 2010). In their study of egalitarianism in children, Fehr et al. (2008) used an envy game in which participants chose between one unit each (1,1) or one for themselves and two for their partner (1,2), finding that although egalitarian preferences dominated at ages 7–8, they were gradually replaced by generosity in older ages (Fehr et al., 2013).

In adults, costless prosociality has been incorporated into modified dictator games consisting of simple allocation tasks, such as selecting an efficient but personally disadvantageous (400,750) choice over an egalitarian (400,400) one (Charness and Rabin, 2002; Engelmann and Strobel, 2004). The generosity game has been specifically designed to examine efficiency concerns, in which individuals choose the size of the overall pie when their own share is fixed (Güth, 2010; Güth et al., 2012). When there is no trade-off between self- and other-interests, most individuals maximize their partner’s payoff, with

**TABLE 1 | Prosocial domains of the Big Five and HEXACO models of personality.**

Personality dimension	Defining characteristic	Known roles in relevant games
<b>Big Five model</b>		
Agreeableness	Broad tendencies toward altruism and cooperation	<ul style="list-style-type: none"> <li>• Fair allocations of wealth</li> <li>• Amount returned in trust game</li> <li>• Acceptance in ultimatum game</li> </ul>
Politeness	Tendency to adhere to social norms; alignment with the group	<ul style="list-style-type: none"> <li>• Fair allocations of wealth</li> </ul>
Compassion	Tendency to be emotionally concerned about others; alignment with another individual	
<b>HEXACO model</b>		
Honesty-Humility	Tendency to cooperate despite opportunities for exploitation; active cooperation	<ul style="list-style-type: none"> <li>• Fair allocations of wealth</li> <li>• Amount returned in trust game</li> </ul>
Agreeableness	Tendency to cooperate despite the misgivings of others; reactive cooperation	<ul style="list-style-type: none"> <li>• Acceptance in ultimatum game</li> </ul>

For a discussion of the role of empathic concern (compassion) as alignment with other individuals and social norms (politeness) as alignment to one's group, see Jensen et al. (2014). Within the HEXACO model, honesty-humility and agreeableness are thought to represent two complementary aspects of reciprocal altruism. In addition, HEXACO emotionality, the tendency to be sentimental and oversensitive, is believed to relate to the construct of kin altruism (Ashton and Lee, 2007; Ashton et al., 2014). However, this dimension is beyond the scope of the current research, which focuses on prosocial behavior among non-kin.

a substantial portion preferring equal shares and a minority minimizing their partner's payoffs (Güth et al., 2012). At the other end of the spectrum, choosing to hurt another or refusing to help them when there is little personal gain may represent purer forms of spite or envy (Abibink and Sadrieh, 2009). Studies using joy-of-destruction games show that some individuals—almost 40% of concealed game decisions—are willing to reduce the payoffs of others even when they do not benefit directly (Abibink and Sadrieh, 2009; Zhang and Ortmann, 2016). Clearly there is much individual variation in costless prosocial and antisocial behaviors—perhaps more so than when decisions are costly and self-interest is a strong driver of uniform responding—and these differences may be reconciled by examining the role of relevant personality constructs, including tendencies toward benevolence, lenience, and spite.

### Positive and Negative Reciprocity

In addition to distributive preferences that govern egalitarianism and generosity, another major influence deeply embedded within social interactions are reciprocal preferences (Fehr and Gächter, 2000; Charness and Rabin, 2002; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006). Reciprocity is the tendency to return others' favors and to retaliate against others' wrongdoing (Gouldner, 1960) and is believed to underlie the evolution and maintenance of human cooperation (Axelrod and Hamilton, 1981; Komorita and Parks, 1999; Bowles and Gintis, 2004). In economics, behavioral signatures of positive and negative reciprocity are often studied in the second player roles of the trust and ultimatum games, respectively (e.g., Fehr et al., 2002; Becker et al., 2012).

Individual differences in the tendency to reciprocate are well documented (Gallucci and Perugini, 2000; Ackermann et al.,

2014), and self-reported reciprocity is associated with major life and economic outcomes (Dohmen et al., 2009). However, the exact relations between positive and negative reciprocity and narrower personality traits are less clear, particularly given the highly conditional nature of reciprocity. For example, positive reciprocators not only need to be sensitive to positive gestures from others, but also have a behavioral propensity to respond to these positively (Perugini et al., 2003).

Within the Big Five model, self-reported positive reciprocity is positively correlated with agreeableness and conscientiousness, while negative reciprocity is negatively correlated with the same two traits, and positively with neuroticism (Perugini et al., 2003; Dohmen et al., 2008). Interestingly, all three traits predict the same outcomes—work effort, unemployment, and subjective wellbeing—associated with individual differences in negative and positive reciprocity, providing further evidence of their overlap (Ozer and Benet-Martínez, 2006; Dohmen et al., 2009). Consistent with these self-reported findings, agreeableness is the Big Five trait most frequently associated with reciprocal behavior in economic games, where it predicts the acceptance of unfair offers in the ultimatum game (Mehta, 2007; Li and Chen, 2012) and greater amounts returned to a sender in the trust game (Evans and Revelle, 2008; Ben-Ner and Halldorsson, 2010; Becker et al., 2012; Müller and Schwieren, 2012; but see Thielmann and Hilbig, 2015).

Furthermore, the HEXACO model and its partitioning of the prosocial domain into active (i.e., honesty-humility) and reactive (i.e., agreeableness) forms of reciprocal altruism is ideally suited to the finer-grained analysis of positive and negative reciprocity in economic games. HEXACO agreeableness has been negatively associated with self-reported negative reciprocity (Perugini et al., 2003) and shown to predict acceptance of unfair offers in ultimatum games (Hilbig et al., 2013; Thielmann

et al., 2014). Meanwhile, honesty-humility has been found to predict trustworthiness, measured by the amount returned in the trust game (Thielmann and Hilbig, 2015). However, this was independent of prior trust, suggesting that the relation is likely driven by a mechanism of “unconditional kindness” (i.e., giving in the absence of any previous or future interaction with one’s partner, such as in a one-shot dictator game), rather than positive reciprocity *per se* (Thielmann and Hilbig, 2015). Other research using wealth redistribution paradigms similarly found that the behavioral expression of honesty-humility is less conditional on fairness norms overall and instead resembles an overall pattern of benevolence (Hilbig et al., 2015b).

## The Current Research

Social preferences represent a number of channels through which humans deviate from narrow self-interest and engage in prosocial behaviors. Distributive preferences capture concerns for egalitarianism and generosity, while preferences for reciprocity promote favorable or unfavorable treatment conditional on the previous acts or intentions of others. Emerging research has demonstrated considerable heterogeneity in these preferences, which may be partially underpinned by prosocial personality traits. However, most of this research has focused on the trade-off between self- and other-regarding interests in the dictator game. Detailed relations between prosocial personality traits and other forms of social preferences are less well understood, and inferences are often cobbled together from a mixture of different games and personality measures. As a result, it is difficult to disentangle trait effects from the influence of contextual factors across variable game environments (given that traits too are contextualized; DeYoung, 2015) and to interpret the findings when certain game decisions are used to approximate social preferences (e.g., the trust game, which may not capture positive reciprocity; Ben-Ner and Halldorsson, 2010; Thielmann and Hilbig, 2015).

The aims of the current research were threefold: (1) To identify a richer set of social preferences beyond egalitarianism and inequality aversion, (2) to examine the source of individual differences in these preferences using theoretical models of distinct prosocial traits, and in doing so, (3) address some of the major limitations of the existing literature (e.g., fragmented games and traits).

We developed a novel paradigm using six simple modifications of the dictator game to test multiple social preferences. This design was inspired by Charness and Rabin (2002), who incorporated reciprocity and efficiency concerns into a series of binary-choice tasks. We first manipulated the costliness of decisions by setting half the games as constant-sum (i.e., costly dictator games) and half with a fixed personal payoff but variable partner payoff (i.e., costless generosity games). Second, we manipulated the conditions for reciprocity by positioning these games after a prior decision by a partner that hurt or helped the participant, vs. a baseline condition where there was no history with a partner.

The benefit of this design is that it provided a suite of tightly controlled and manipulable conditions ideal for localizing specific prosocial constructs. For example, comparing costly vs. costless game decisions allowed us to identify different patterns of behavior after controlling for the influence of self-interest. Similarly, reciprocal tendencies can be teased apart from overall altruistic motivations. Existing studies suggest that Big Five agreeableness and HEXACO honesty-humility are associated positively with positive reciprocity and negatively with negative reciprocity, but these preferences have been largely considered in isolation. Given that these traits are already associated with greater dictator allocations, the current design will reveal whether they produce an *additional effect* for reciprocity, above and beyond unconditional kindness (Ben-Ner and Halldorsson, 2010; Thielmann and Hilbig, 2015).

We then examined the sources of heterogeneity within this paradigm with respect to the theoretically relevant prosocial domain of personality: agreeableness and its aspects of politeness and compassion within the Big Five model, and honesty-humility and agreeableness within the HEXACO model. In particular, we focused on the discriminant validity between similar prosocial personality constructs and identified unique trait effects to help shed light on the specific mechanisms and motivations taking place within economic games (for a recent example, see Zhao et al., 2016). We sought to address some of the limitations and expand on the existing research by bringing together two large and relatively diverse community samples. Sample sizes in both studies ( $N_s = 304, 256$ ) were well above the recommended minimum provided in the wake of the replicability crisis, including total  $N > 150$  for individual differences research (Mar et al., 2013) and total  $N > 180$  as a general requirement for personality and social psychological research (Vazire, 2016). As the design was within-subjects, the *per condition* sample sizes provided 80% power to identify effect sizes of approximately  $rs = 0.16\text{--}0.18$  (Faul et al., 2009), which is reasonably sensitive given the average effect sizes in the field ( $r = 0.21$ ; Richard et al., 2003; Fraley and Marks, 2007).

In line with previous research, we expected politeness from the Big Five model to be uniquely associated with costly prosocial allocations (i.e., dictator games) but expected compassion to play a relatively stronger role in costless prosocial allocations (i.e., generosity games), where allocations are less norm-driven and capture motivations of improving the wellbeing of others. Within the HEXACO model, we predicted that honesty-humility would have a unique role in both costly and costless prosociality, given its core characteristic of benevolence. Furthermore, we hypothesized that HEXACO agreeableness—which captures tendencies toward forgiveness and non-retaliation—would be negatively associated with negative reciprocity. Finally, in light of the evidence demonstrating the role of Big Five agreeableness and HEXACO honesty-humility in *both* positive reciprocal game behaviors and dictator game allocations, we were interested in examining whether any prosocial traits could explain positive reciprocity beyond their established role in unconditional kindness.

## STUDY 1

### Materials and Methods

#### Ethics Statement

This study was approved by the Human Ethics Advisory Group of the Melbourne School of Psychological Sciences, The University of Melbourne. All participants provided informed consent via an electronic survey according to the established guidelines of the Group.

#### Participants

The final sample consisted of 304 North American participants (aged 18–65 years,  $M_{\text{age}} = 30.90$ ,  $SD = 9.89$ ; 55% female) recruited from Amazon Mechanical Turk (MTurk). Only workers with fewer than 50 Human Intelligence Tasks were selected to avoid recruiting those who were familiar with economic game paradigms.

#### Personality Measures

##### *Big Five Aspect Scales (BFAS; DeYoung et al., 2007)*

Participants completed the 100-item BFAS, a measure of the five broad domains of personality (neuroticism, agreeableness, conscientiousness, extraversion, and openness/intellect) and their lower-level aspects. Of particular interest was the prosocial domain of agreeableness, including its aspects of politeness (e.g., “insult people”) and compassion (e.g., “inquire about others’ wellbeing”). These were each measured with 10 items on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree). The BFAS is a well-validated measure of the Big Five and has good internal consistency and test-retest reliability (DeYoung et al., 2007).

##### *HEXACO Personality Inventory—Revised (HEXACO-PI-R; Lee and Ashton, 2004)*

Participants also completed the 100-item HEXACO-PI-R, an alternative measure of personality comprising six broad trait domains. Of particular interest were the prosocial domains of honesty-humility (e.g., “I am an ordinary person who is no better than others”) and agreeableness (e.g., “I rarely hold a grudge, even against people who have badly wronged me”). Each trait is measured with 16 items on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree), and has good internal consistency (Lee and Ashton, 2004)<sup>1</sup>.

#### Procedure

Participants completed demographic questions, personality measures, and economic games on a survey programmed using Qualtrics Survey Software and administered through the MTurk requester interface. The BFAS and the HEXACO-PI-R were presented one after the other in a randomized order. The survey consisted of additional questionnaires and economic games

<sup>1</sup>In addition, an interstitial scale, altruism, represents a blend of HEXACO honesty-humility, agreeableness, and emotionality (e.g., “I have sympathy for people who are less fortunate than I am”; Ashton et al., 2014). Given its extensive overlap with prosocial domains from both personality models (and the focus of the current study on *distinct* prosocial traits) and its relatively lower reliability (Cronbach’s  $\alpha = 0.62, 0.71$ ), data for this scale were not included in the main analysis but can be found in the Supplementary Material (see Supplementary Table S2).

beyond the scope of the current research, including a hypothetical real-world economic decision-making task. The 200 items of the personality questionnaires served as a filler task between this and the current games of interest, and thus were expected to prevent any carryover effects.

All economic games were hypothetical, that is, participants were asked to imagine that they were playing the games with an anonymous partner who was described as another participant that they would not knowingly meet. To check the validity of responses, participants also completed two attention checks embedded in the personality measures (e.g., “Please select Strongly Agree”). Thirty-six (11%) participants were excluded for failing at least one of these attention checks. Participants were paid US\$2.00 and the median time spent on the study was 30 min.

#### *Economic games*

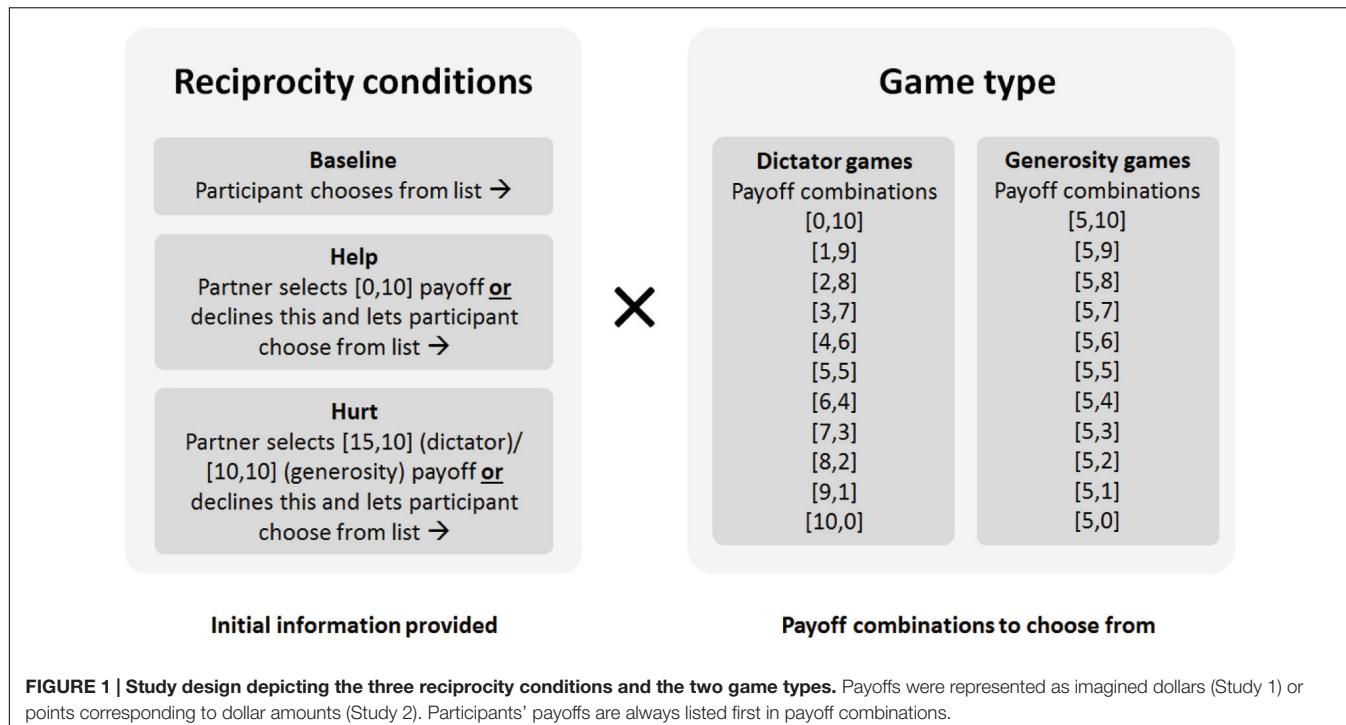
Participants played six economic games that were loosely based on a larger set of dictator and response games developed by Charness and Rabin (2002). All six games required the participant to select their preferred choice out of 11 combinations of payoffs for themselves and their partner, represented by imagined dollar amounts. All games were presented in a randomized order.

The six games were set up using a 2 (game type: dictator vs. generosity)  $\times$  3 (reciprocity: baseline, help, and hurt) repeated measures design, depicted in **Figure 1**. There were two types of games: dictator and generosity games. In the three dictator games (Kahneman et al., 1986; Forsythe et al., 1994), participants were asked to indicate their preferred choice out of 11 different payoff combinations, each of which summed to \$10. These ranged from \$0 for oneself and \$10 for one’s partner to \$10 for oneself and \$0 for one’s partner, varying in \$1 increments.

In the three generosity games (based on Güth et al., 2009, 2012; Güth, 2010)<sup>2</sup>, participants were again asked to indicate their preferred selection out of 11 different payoff combinations. This time, their own payoff was always fixed at \$5 and the choices ranged from \$0 to \$10 for their partner, varying in \$1 increments.

In addition, there were three types of reciprocity conditions: baseline, help, and hurt. In the two baseline games, participants were asked to indicate their preferred selection with no information provided about their partner. In the four remaining games, participants were provided information about their partner’s previous move, which involved passing on a decision that either helped or hurt the participant. In the two help games, participants read that their partner had passed on a decision with a payoff of \$0 to the participant, opting instead to defer to the participant to choose from the list of current options. In the two hurt games, participants read that their partner had passed on a decision with a payoff of \$15 (dictator version) or \$10 (generosity version) to the participant, opting instead to defer to the participant. In other words, the partner’s move in the help condition prevented the participant from going away

<sup>2</sup>While our generosity game was inspired by that designed by Güth et al., (2009, 2012) and Güth (2010), the two are not the same as players in the latter decide on the size of the entire pie so that the partner is the residual claimant. In contrast, our participants directly selected the payoffs for their partner, which aided ease of understanding for participants and allowed comparability with dictator games in our analysis.



**FIGURE 1 | Study design depicting the three reciprocity conditions and the two game types.** Payoffs were represented as imagined dollars (Study 1) or points corresponding to dollar amounts (Study 2). Participants' payoffs are always listed first in payoff combinations.

empty-handed, while their move in the hurt condition resulted in the participant missing out on \$15 (dictator version) or \$10 (generosity version). These different forgone payoffs between the dictator and generosity games correspond to the maximum amounts that could be earned in each of these games (\$10 in the dictator game, \$5 in the generosity game).

To summarize, this experimental setup would thus reveal an effect for generosity if there were greater allocations in the generosity games relative to the dictator game (i.e., a main effect for game type). In addition, reciprocity would be evident from varying allocations of wealth between the baseline, help, and hurt games (i.e., a main effect of reciprocity), in which higher allocations in the help games would be indicative of positive reciprocity and lower allocations in the hurt games indicative of negative reciprocity.

## Results and Discussion

### Preliminary Statistics

#### Game decisions

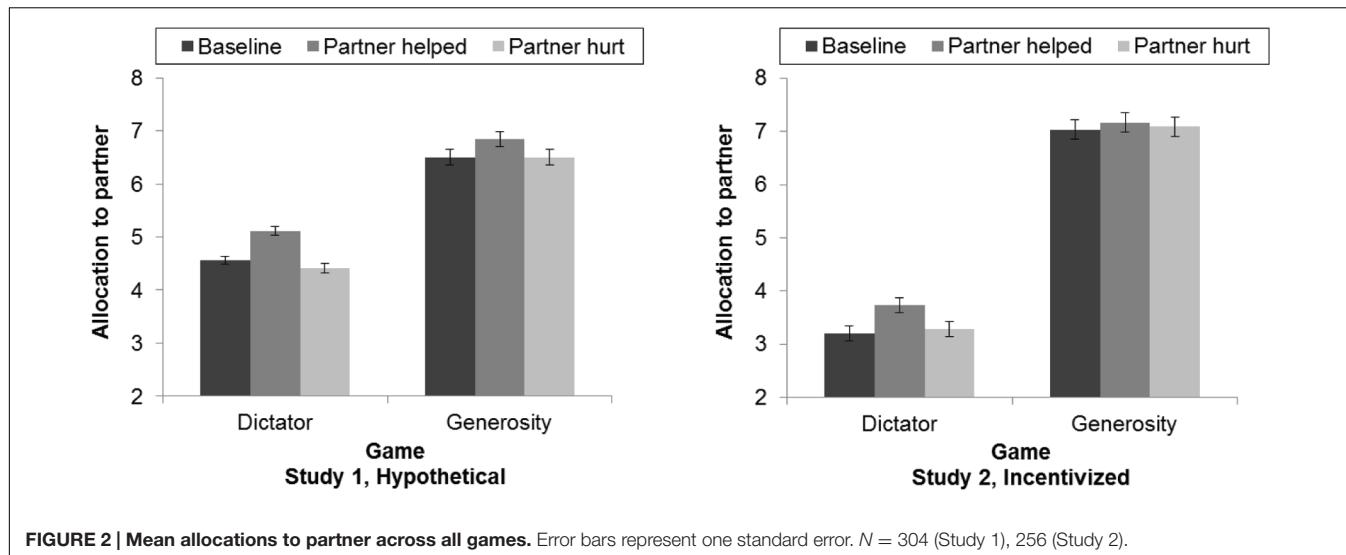
Mean allocations to a partner in each of the six economic bargaining games are presented in the left panel of **Figure 2**. A 2 (game type: dictator vs. generosity)  $\times$  3 (reciprocity: baseline, help, and hurt) repeated measures ANOVA was performed. Greenhouse-Geisser corrections were applied for sphericity violations of reciprocity,  $\chi^2(2) = 29.92$ ,  $p < 0.001$  ( $\epsilon = 0.91$ ), and its interaction with game type,  $\chi^2(2) = 14.62$ ,  $p = 0.001$  ( $\epsilon = 0.96$ ). There was a main effect for game type, with allocations in generosity games ( $M = 6.62$ ) higher than those in dictator games ( $M = 4.70$ ),  $F(1,303) = 212.12$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.41$ . There was also a main effect for reciprocity,  $F(1.83,553.77) = 15.68$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.05$ , for which allocations in the baseline

games ( $M = 5.54$ ) were significantly lower than in help games ( $M = 5.98$ ),  $F(1,303) = 27.16$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.08$ , but not in hurt games ( $M = 5.46$ ),  $F(1,303) = 0.57$ ,  $p = 0.45$ ,  $\eta_p^2 = 0.002$ . These findings thus indicate generosity and positive reciprocity, but not negative reciprocity.

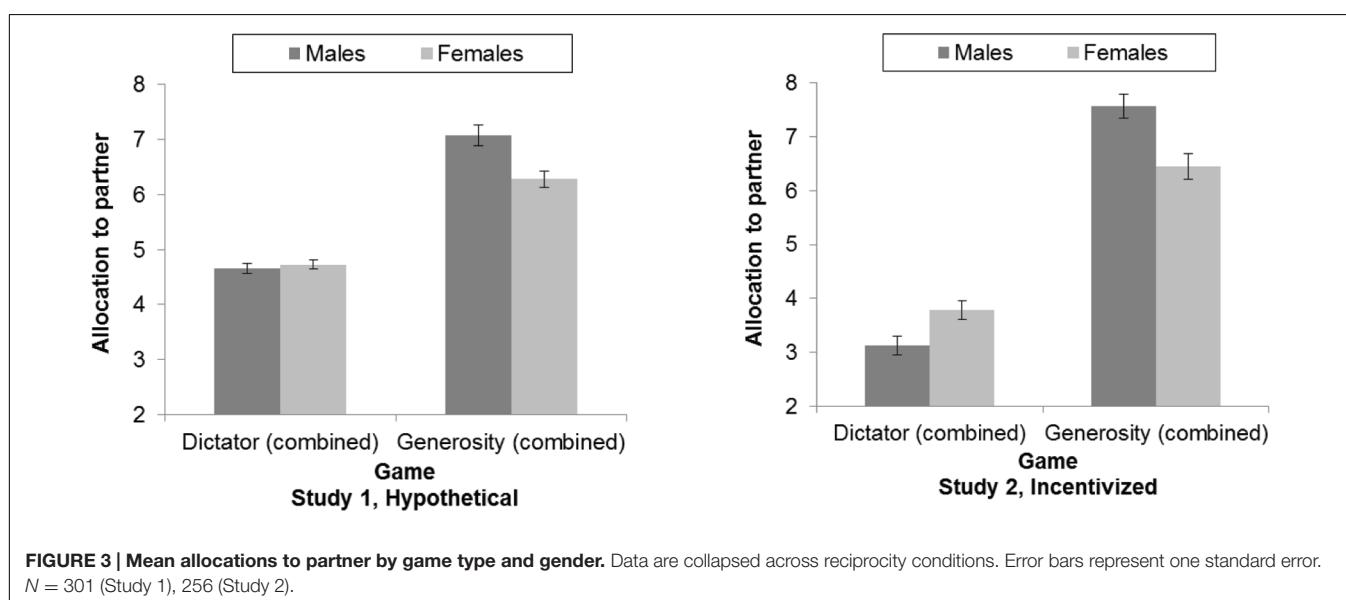
#### Demographic variables

Age and gender are important demographic variables frequently associated with social preferences (Andreoni and Vesterlund, 2001; for a discussion of age-related effects and possible confounds, see Kettner and Waichman, 2016). In the current study, age was not significantly correlated with any game decisions. In contrast, there was a significant interaction between gender and game type. After removing three participants who identified as neither male nor female, gender was included in the 2 (game type)  $\times$  3 (reciprocity) repeated measures ANOVA. This model produced a main effect for gender,  $F(1,299) = 7.45$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.02$ , with men allocating on average more than women,  $M_s = 5.86$  vs.  $5.50$ ,  $t(299) = 2.73$ ,  $p = 0.01$ . However, these findings were moderated by a significant interaction between gender and game type,  $F(1,299) = 10.88$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.04$ , and between gender and reciprocity,  $F(1.84,550.31) = 5.59$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.02$ .

Allocations by gender and game type are presented in the left panel of **Figure 3**, collapsed across reciprocity conditions. The main effect of gender appeared to be driven by men allocating more than women in generosity games,  $M_s = 7.07$  vs.  $6.28$ ,  $t(299) = 3.41$ ,  $p = 0.001$ , but no differently in dictator games,  $t(299) = -0.58$ ,  $p = 0.56$ . Meanwhile, men allocated more than women in the baseline,  $M_s = 5.76$  vs.  $5.37$ ,  $t(299) = 2.30$ ,  $p = 0.02$ , and help conditions,  $M_s = 6.36$  vs.  $5.68$ ,  $t(299) = 4.00$ ,  $p < 0.001$ ,



**FIGURE 2 | Mean allocations to partner across all games.** Error bars represent one standard error.  $N = 304$  (Study 1), 256 (Study 2).



**FIGURE 3 | Mean allocations to partner by game type and gender.** Data are collapsed across reciprocity conditions. Error bars represent one standard error.  $N = 301$  (Study 1), 256 (Study 2).

but not in the hurt conditions,  $M_s = 5.47$  vs.  $5.46$ ,  $t(299) = 0.05$ ,  $p = 0.96$ . All main effects for game type and reciprocity were replicated when including gender.

### Prosocial personality traits

Bivariate correlations between prosocial personality traits are shown in Table 2 and were generally consistent with previous research (Barford et al., 2015; Zhao et al., 2016). Both HEXACO honesty-humility and, to a lesser extent, HEXACO agreeableness were more strongly correlated with politeness ( $r_{ss} = 0.51$ , 0.26) than compassion ( $r_{ss} = 0.24$ , 0.14).

### Personality Predictors of Game Allocations

#### Bivariate correlations

Bivariate correlations between game allocations and prosocial personality traits are shown in Table 3 (see Supplementary Tables S1 and S2 in the Supplementary Material for correlations

with all personality traits). The politeness and compassion aspects of Big Five agreeableness were both correlated with all three dictator games ( $r_{ss} = 0.13$ –0.19). Similarly, HEXACO honesty-humility was associated with dictator ( $r_{ss} = 0.12$ –0.26), but not generosity ( $r_{ss} = -0.06$  to  $-0.01$ ) allocations. In contrast, HEXACO agreeableness was correlated with generosity ( $r_{ss} = 0.15$ –0.18) but not dictator allocations ( $r_{ss} = -0.01$ –0.10).

#### Repeated measures ANCOVAs

A series of 2 (game type)  $\times$  3 (reciprocity) repeated measures ANCOVAs was performed for each personality model with the relevant prosocial traits standardized and entered simultaneously as covariates. Interactions between prosocial personality traits and game type or reciprocity are presented in Tables 4 and 5.

Within the Big Five model, there was a main effect for agreeableness,  $F(1,302) = 6.23$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.02$ . Replacing this

**TABLE 2 | Correlations between prosocial personality traits.**

	Study 1: Hypothetical					Study 2: Incentivized				
	1	2	3	4	5	1	2	3	4	5
(1)	B5 Agreeableness	0.86				0.91				
(2)	B5 Politeness	0.84**	0.74			0.85**	0.82			
(3)	B5 Compassion	0.84**	0.44**	0.88		0.90**	0.56**	0.92		
(4)	HEX Honesty-Humility	0.44**	0.51**	0.24**	0.82	0.41**	0.48**	0.27**	0.87	
(5)	HEX Agreeableness	0.22**	0.26**	0.14*	0.19**	0.84	0.53**	0.52**	0.42**	0.31**
										0.88

Correlations calculated using Spearman's rho. Cronbach's  $\alpha$ s are shown in the diagonal. B5, Big Five model, measured using the Big Five Aspect Scales (BFAS; DeYoung et al., 2007). HEX, HEXACO model, measured using the HEXACO Personality Inventory—Revised (HEXACO-PI-R; Lee and Ashton, 2004). N = 304 (Study 1), 256 (Study 2). \* $p < 0.05$ , \*\* $p < 0.01$ .

**TABLE 3 | Correlations between prosocial personality traits and game allocations.**

	Study 1: Hypothetical						Study 2: Incentivized					
	DG	DG <sub>0</sub>	DG <sub>15</sub>	GG	GG <sub>0</sub>	GG <sub>10</sub>	DG	DG <sub>0</sub>	DG <sub>15</sub>	GG	GG <sub>0</sub>	GG <sub>10</sub>
<b>Big Five model</b>												
Agreeableness	0.20**	0.17**	0.16**	0.02	0.05	0.01	0.19**	0.11	0.18**	-0.01	0.0003	-0.07
Politeness	0.17**	0.16**	0.15**	0.02	0.07	-0.01	0.17**	0.12	0.15*	0.07	0.07	0.04
Compassion	0.19**	0.16**	0.13*	0.05	0.05	0.05	0.17**	0.07	0.17**	-0.09	-0.07	-0.14*
<b>HEXACO model</b>												
Honesty-Humility	0.20**	0.12*	0.26**	-0.06	-0.04	-0.01	0.30**	0.21**	0.31**	-0.04	-0.02	-0.08
Agreeableness	0.10	0.03	-0.01	0.17**	0.18**	0.15**	0.07	-0.02	0.02	0.04	0.004	-0.02

Correlations calculated using Spearman's rho. Game allocations indicate amount allocated to partner out of 10 units (i.e., dollars or points). Big Five traits are measured using the Big Five Aspect Scales (BFAS; DeYoung et al., 2007). HEXACO traits are measured using the HEXACO Personality Inventory—Revised (HEXACO-PI-R; Lee and Ashton, 2004). DG, baseline dictator game. DG<sub>0</sub>, dictator game after partner's decision cost the participant the 0 unit payoff. DG<sub>15</sub>, dictator game after partner's decision cost the participant the 15 unit payoff. GG, baseline generosity game. GG<sub>0</sub>, generosity game after partner's decision cost the participant the 0 unit payoff. GG<sub>10</sub>, generosity game after partner's decision cost the participant the 10 unit payoff. N = 304 (Study 1), 256 (Study 2). \* $p < 0.05$ , \*\* $p < 0.01$ .

**TABLE 4 | ANCOVA results for interactions between prosocial traits and game type.**

Interaction term	Study 1: Hypothetical					Study 2: Incentivized				
	df	F	p	$\eta^2_p$	df	F	p	$\eta^2_p$		
<b>Big Five model (B5A only)</b>										
Game × B5A	1, 302	0.84	0.36	0.003	1, 254	2.24	0.14	0.01		
<b>Big Five model</b>										
Game × B5Pol	1, 301	0.30	0.58	0.001	1, 253	1.97	0.16	0.01		
Game × B5Comp	1, 301	0.18	0.67	0.001	1, 253	7.70	0.01	0.03		
<b>HEXACO model</b>										
Game × HEXH	1, 301	12.84	<0.001	0.04	1, 253	11.48	0.001	0.04		
Game × HEXA	1, 301	9.62	0.002	0.03	1, 253	2.62	0.11	0.01		

B5, Big Five model, measured using the Big Five Aspect Scales (BFAS; DeYoung et al., 2007). B5A, B5 Agreeableness. B5Comp, B5 Compassion. B5Pol, B5 Politeness. HEX, HEXACO model, measured using the HEXACO Personality Inventory—Revised (HEXACO-PI-R; Lee and Ashton, 2004). HEXA, HEX Agreeableness. HEXH, HEXACO Honesty-Humility. N = 304 (Study 1), 256 (Study 2).

with covariates for politeness and compassion initially revealed no significant main effects for either. However, as men allocated more than women overall (primarily driven by generosity game allocations) and men were significantly lower on politeness and compassion than women, we also included gender in the same model. Here, a main effect for politeness,  $F(1,297) = 5.65$ ,  $p = 0.02$ ,  $\eta^2_p = 0.02$  [and a marginally significant effect for compassion,  $F(1,297) = 4.03$ ,  $p = 0.05$ ,  $\eta^2_p = 0.01$ ] emerged, suggesting that politeness was related to greater allocations

across all conditions (see Supplementary Tables S3–S5 in the Supplementary Material). None of the prosocial personality traits in the Big Five model interacted with game type or reciprocity.

Within the HEXACO model, there was a main effect for agreeableness,  $F(1,301) = 9.28$ ,  $p = 0.003$ ,  $\eta^2_p = 0.03$ , but not honesty-humility  $F(1,301) = 0.71$ ,  $p = 0.40$ ,  $\eta^2_p = 0.002$ . This was accompanied by significant interactions between honesty-humility and game type,  $F(1,301) = 12.84$ ,  $p < 0.001$ ,  $\eta^2_p = 0.04$ , and agreeableness and game type,

**TABLE 5 | ANCOVA results for interactions between prosocial traits and reciprocity.**

Interaction term	Study 1: Hypothetical				Study 2: Incentivized			
	df	F	p	$\eta_p^2$	df	F	p	$\eta_p^2$
<b>Big Five model (B5A only)</b>								
Reciprocity × B5A	1.83, 551.89	0.04	0.95	<0.001	1.89, 478.90	0.19	0.82	0.001
<b>Big Five model</b>								
Reciprocity × B5Pol	1.83, 550.24	0.22	0.78	0.001	1.89, 477.05	0.18	0.83	0.001
Reciprocity × B5Comp	1.83, 550.24	0.20	0.80	0.001	1.89, 477.05	0.002	0.99	<0.001
<b>HEXACO model</b>								
Reciprocity × HEXH	1.84, 553.30	4.42	0.02	0.01	1.89, 478.16	0.17	0.84	0.001
Reciprocity × HEXA	1.84, 553.30	0.78	0.45	0.003	1.89, 478.16	1.87	0.16	0.01

B5, Big Five model, measured using the Big Five Aspect Scales (BFAS; DeYoung et al., 2007). B5A, B5 Agreeableness. B5Comp, B5 Compassion. B5Pol, B5 Politeness. HEX, HEXACO model, measured using the HEXACO Personality Inventory—Revised (HEXACO-PI-R; Lee and Ashton, 2004). HEXA, HEX Agreeableness. HEXH, HEXACO Honesty-Humility. N = 304 (Study 1), 256 (Study 2).

$F(1,301) = 9.62$ ,  $p = 0.002$ ,  $\eta_p^2 = 0.03$ . This pattern of findings was replicated when gender was included in the model (see Supplementary Tables S3–S5 in the Supplementary Material).

To follow up on these interactions, we examined the effect of these two traits for dictator and generosity games separately, which revealed a “double dissociation” between the two, depicted in the left panel of **Figure 4**. Honesty-humility was uniquely associated with greater allocations in dictator games,  $F(1,301) = 23.77$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.07$ , but did not have a main effect in generosity games,  $F(1,301) = 2.32$ ,  $p = 0.13$ ,  $\eta_p^2 = 0.01$ . In contrast, HEXACO agreeableness was uniquely associated with greater allocations in generosity games,  $F(1,301) = 11.88$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.04$ , but did not have a main effect in dictator games,  $F(1,301) = 0.0004$ ,  $p = 0.98$ ,  $\eta_p^2 < 0.001$ .

In addition, there was a significant interaction between honesty-humility and reciprocity,  $F(1.84,553.30) = 4.42$ ,  $p = 0.02$ ,  $\eta_p^2 = 0.01$ . This revealed a significant main effect of honesty-humility in the hurt conditions,  $F(1,301) = 5.94$ ,  $p = 0.02$ ,  $\eta_p^2 = 0.02$ , but not in the baseline or help conditions ( $p = 0.99, 0.49$ , respectively).

## Summary

The results of Study 1 showed clear evidence of social preferences beyond inequality aversion and egalitarianism. Individuals allocated significantly more wealth to their partners when decisions were costless than when they were costly, demonstrating tendencies toward generosity. In addition, there was evidence of positive reciprocity, with individuals allocating more wealth to their partner after their partner had assisted them. However, we found no evidence of negative reciprocity, and individuals did not allocate any differently when they had been denied a higher payoff by a hurtful partner. These findings were further moderated by gender, with men allocating more than women in the generosity games and when their partner had not previously hurt them.

The results presented a mixed picture of predicted and unexpected findings regarding the role of personality, revealing

a main effect for politeness (but not so much compassion) in the Big Five model. In the HEXACO model, honesty-humility predicted greater allocations in the dictator game, in keeping with a large body of previous research (Hilbig and Zettler, 2009; Hilbig et al., 2015a; Zhao and Smillie, 2015). However, contrary to its putative mechanism of benevolence, honesty-humility did not play any role in the generosity game, where decisions were costless. Here, it was HEXACO agreeableness—or the tendency to be tolerant, lenient, and forgiving—which instead predicted greater generosity.

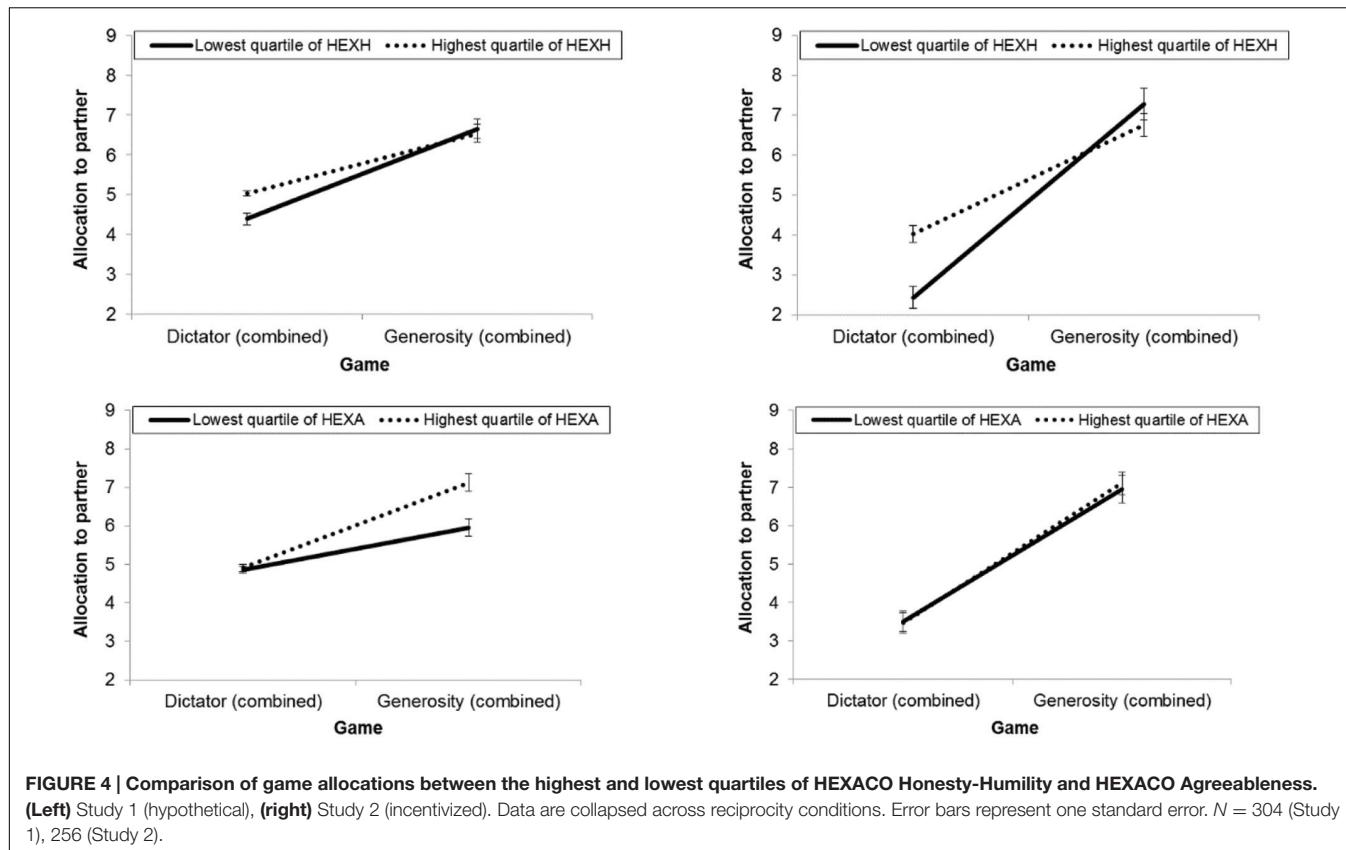
An important consideration in Study 1 is that the decisions were hypothetical, featuring imagined partners and stakes. Previous studies have been conflicted as to whether hypothetical paradigms produce comparable results to incentivized games, especially when trait effects are involved (Ben-Ner et al., 2008; Engel, 2011; Lönnqvist et al., 2011; Ferguson and Starmer, 2013; Hilbig et al., 2015a; Zhao et al., 2016). Another potential limitation stems from correlating self-reported personality traits with self-reported hypothetical responses, where there is a risk of inflated associations arising from common method variance (Podsakoff et al., 2003). Finally, the dearth of actual assessment of behavior has been a prominent issue in personality research, leading to calls for a broader range of data beyond self-reports and hypothetical scenarios (Funder, 2001; Baumeister et al., 2007). In light of these concerns, we ran a second study using an identical—but incentivized—paradigm with the aim of replicating our previous findings and identifying robust effects.

## STUDY 2

### Materials and Methods

#### Ethics Statement

This study was approved by the Human Ethics Advisory Group of the Melbourne School of Psychological Sciences, The University of Melbourne. All participants provided informed consent via an electronic survey according to the established guidelines of the Group.



**FIGURE 4 | Comparison of game allocations between the highest and lowest quartiles of HEXACO Honesty-Humility and HEXACO Agreeableness. (Left)** Study 1 (hypothetical), **(right)** Study 2 (incentivized). Data are collapsed across reciprocity conditions. Error bars represent one standard error.  $N = 304$  (Study 1), 256 (Study 2).

## Participants

The final sample consisted of 256 North American participants (aged 19–67 years,  $M_{age} = 34.76$ ,  $SD = 11.00$ ; 43% female) recruited from Amazon MTurk.

## Personality Measures

Participants completed the 100-item BFAS (DeYoung et al., 2007), along with the honesty-humility, agreeableness, and altruism scales (see Footnote 1) from the HEXACO-PI-R (Lee and Ashton, 2004), described in Study 1.

## Procedure

Participants completed the same demographic questions, personality measures, and economic games as Study 1, which were again programmed using Qualtrics Survey Software and administered through the MTurk requester interface. This time, however, the BFAS was presented before the HEXACO-PI-R and the two were separated by several other questionnaires (e.g., Major Life Goals, Roberts and Robins, 2000). In addition, the games of interest were preceded by a social mindfulness task involving the hypothetical selection of specific objects (Van Doesum et al., 2013) and subjective ratings of the payoff structures of social dilemmas (Halevy et al., 2012), both of which were beyond the scope of the aims of the current research. Neither involved any explicit themes of prosociality and were not expected to produce any carryover effects.

Unlike Study 1, participants' responses to all games were financially incentivized. This was done by informing participants

that their decisions for one of the games (which was pre-selected) would be matched to another participant and used to determine their payment at the end of the session. This approach is similar to the Conditional Information Lottery, which is a standard procedure in the literature (Bardsley, 2000). In the help and hurt reciprocity conditions, participants were asked to indicate their responses using the strategy method and assume that they would be matched to a partner who had picked a given move. Game payoffs were represented by points that corresponded with real dollar amounts at a rate of 1 point to US\$0.10. Bonus payments were then provided to participants at the end of the study using their anonymous response identification codes.

Participants completed the same two attention checks as in Study 1. Ten participants (3.8%) were excluded for failing at least one of these checks. The show-up fee was US\$8.00, in addition to bonus payments earned from study tasks (US\$0.50). The median time spent on the study was 42 min.

## Results and Discussion

### Preliminary Statistics

#### Game decisions

Mean allocations to a partner are presented in the right panel of Figure 2. Comparing across studies, all three dictator allocations were significantly lower in the incentivized Study 2 than the hypothetical Study 1 ( $p < 0.001$ ). Conversely, all but one generosity allocation (where a partner had previously helped the

participant,  $p = 0.16$ ) were significantly higher in Study 2 than Study 1 ( $ps < 0.05$ ).

A 2 (game type: dictator vs. generosity)  $\times$  3 (reciprocity: baseline, help, and hurt) repeated measures ANOVA was performed with Greenhouse-Geisser corrections for sphericity violations of reciprocity,  $\chi^2(2) = 16.00$ ,  $p < 0.001$  ( $\epsilon = 0.94$ ), and its interaction with game type,  $\chi^2(2) = 10.34$ ,  $p = 0.01$  ( $\epsilon = 0.96$ ). The results in Study 1 were replicated here, including main effects for game type,  $F(1,255) = 253.20$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.50$ , and reciprocity,  $F(1.89,480.67) = 9.40$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.04$ . There was also an interaction between game type and reciprocity,  $F(1.92,490.44) = 3.87$ ,  $p = 0.02$ ,  $\eta_p^2 = 0.02$ , which had been marginally significant ( $p = 0.09$ ) in Study 1. Post hoc comparisons with Bonferroni corrections revealed that the effect for reciprocity applied only to dictator games. Dictator allocations were significantly higher in the help games ( $M = 3.74$ ) compared with the baseline ( $M = 3.20$ ,  $p < 0.001$ ), and hurt games ( $M = 3.28$ ,  $p < 0.001$ ), but there were no significant differences across reciprocity conditions for the generosity games (all  $ps > 0.30$ ).

### Demographic variables

Again, age was not significantly correlated with any game decisions. There was an interaction between gender and game type when gender was included in the 2 (game type)  $\times$  3 (reciprocity) repeated measures ANOVA,  $F(1,254) = 15.32$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.06$ , shown in the right panel of **Figure 3**. Women allocated significantly more than men in dictator games,  $M_s = 3.79$  vs.  $3.13$ ,  $t(254) = 2.59$ ,  $p = 0.01$ , but this was reversed in the generosity game, where, as in Study 1, men allocated significantly more than women,  $M_s = 7.57$  vs.  $6.45$ ,  $t(254) = 3.38$ ,  $p = 0.001$ . All main effects and interactions for game type and reciprocity were replicated when including gender.

### Prosocial personality traits

Bivariate correlations between prosocial personality traits are shown in **Table 2** and were generally consistent with those in Study 1. However, HEXACO agreeableness was more strongly correlated with all other prosocial traits in Study 2 than in Study 1.

## Personality Predictors of Game Allocations

### Bivariate correlations

Bivariate correlations between game allocations and prosocial personality traits are shown in **Table 3** (see Supplementary Tables S1 and S2 in the Supplementary Material for correlations with all personality traits). Compared with Study 1, a stronger pattern of correlations was seen for honesty-humility, where it was again associated with dictator ( $r_{ss} = 0.21$ – $0.31$ )—but not generosity ( $r_{ss} = -0.08$  to  $-0.02$ )—allocations. In contrast to Study 1, however, HEXACO agreeableness was not associated with allocations in any game ( $r_{ss} = -0.02$ – $0.07$ ).

### Repeated measures ANCOVAs

A series of 2 (game type)  $\times$  3 (reciprocity) repeated measures ANCOVAs was again performed for each personality model with the relevant traits standardized and entered simultaneously as covariates (see **Tables 4** and **5**; **Figure 4**).

Within the Big Five model, there was again a main effect for agreeableness,  $F(1,254) = 7.12$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.03$ . Replacing this with covariates for politeness and compassion revealed a unique main effect for politeness only,  $F(1,253) = 17.89$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.07$ , and not compassion,  $F(1,253) = 1.87$ ,  $p = 0.17$ ,  $\eta_p^2 = 0.01$ . Unlike Study 1, there was a significant interaction between compassion and game type,  $F(1,253) = 7.70$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.03$ . Follow-up analysis revealed that compassion was associated with lower allocations in generosity games,  $F(1,253) = 7.20$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.03$ , but did not have a main effect in dictator games when politeness was controlled for,  $F(1,253) = 2.39$ ,  $p = 0.12$ ,  $\eta_p^2 = 0.01$ .

Within the HEXACO model, there was a main effect for honesty-humility,  $F(1,253) = 7.02$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.03$ , but not agreeableness  $F(1,253) = 0.003$ ,  $p = 0.95$ ,  $\eta_p^2 < 0.001$ . Whereas the interaction for HEXACO agreeableness observed in Study 1 fell short of significance here,  $F(1,253) = 2.62$ ,  $p = 0.11$ ,  $\eta_p^2 = 0.01$ , there was again a significant interaction between honesty-humility and game type,  $F(1,253) = 11.48$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.04$ . As in Study 1, this revealed a significant positive effect of honesty-humility in dictator games,  $F(1,253) = 26.98$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.10$ , but not in generosity games,  $F(1,253) = 0.68$ ,  $p = 0.41$ ,  $\eta_p^2 = 0.003$ .

The above analyses were repeated and the findings were largely the same when gender was included as an additional term (see Supplementary Tables S3–S5 in the Supplementary Material).

## Summary

The incentivized results of Study 2 replicated many of the main findings from the hypothetical paradigm of Study 1. Again, there was clear evidence of inequality aversion, generosity, and positive reciprocity, which were moderated by gender. When we examined the role of prosocial personality traits, honesty-humility once more interacted with game type, predicting greater allocations in dictator—but not generosity—games. In the Big Five model, we again observed a main effect of politeness—but not compassion—which was globally and uniquely associated with greater allocations across all games.

However, the results of Study 2 also introduced two non-trivial differences compared with Study 1. First, the previous interaction between agreeableness and game type in the HEXACO model disappeared in the incentivized paradigm. In fact, HEXACO agreeableness was not associated with allocations of any kind. Second, a novel and unpredicted interaction with game type emerged for compassion in the Big Five model, in which it was not related to dictator allocations, but predicted lower allocations in the generosity game, once politeness was controlled for. This combination of consistent and less consistent findings across the two studies demonstrates the importance of replication and comparisons across incentivized and hypothetical paradigms.

## GENERAL DISCUSSION

Prosociality is a complex, multidimensional construct, yet previous research on personality and social preferences has largely focused on simple games and broad trait domains.

Expanding on this literature, we developed a novel behavioral paradigm (inspired by Charness and Rabin, 2002), which integrated multiple social preferences using slight variations of the dictator game. We ran two studies—one with hypothetical decisions and one with incentivized games—across two large, and relatively diverse community samples to identify consistent effects. The findings provide clear evidence of inequality aversion, generosity, and positive reciprocity, which we mapped to a framework of distinct prosocial personality traits. This highlighted the unique roles of politeness from the Big Five model, honesty-humility from the HEXACO model, and more tentatively, traits reflecting irritability, anger, and (a lack of) tolerance and forgiveness.

The sizes for these effects are consistent with those previously observed for the role of personality in economic games, where the sample-size weighted average correlation with dictator allocations was  $r_s = 0.20$  for Big Five agreeableness (Zhao et al., 2016) and  $r_s = 0.25$  and  $r = 0.29$  for HEXACO honesty-humility (Hilbig et al., 2015a; Zhao et al., 2016). Though they may initially appear modest, these correlations—particularly for HEXACO honesty-humility—are at least as large as the average effect size in social and personality psychology ( $r = 0.21$ ; Richard et al., 2003; Fraley and Marks, 2007), and fall within the middle third of effect sizes in psychology as a whole (Hemphill, 2003). These findings will be discussed in detail in the following sections, with a focus on the robust and replicable effects across both two studies.

## Beyond Egalitarianism: Evidence for Generosity and Positive Reciprocity

In line with a large body of literature, our two studies showed that humans are responsive to additional social preferences that stray from both narrow self-interest and inequality aversion. The findings from the generosity game correspond to previous research showing that many individuals are willing to assist others even when it means being *relatively* less well off, as long as absolute costs are minimal (Charness and Rabin, 2002; Engelmann and Strobel, 2004; Güth et al., 2012). Generosity may be crowded out by the trade-off between self- and other-interests in the dictator game, but when it is costless, “most of us try to make the world a better place” (Güth et al., 2009, p. 13). Likewise, many real-world gestures of prosociality, such as giving directions to a stranger and offering a seat on public transport, are ubiquitous precisely because they are relatively inexpensive forms of benevolence.

In contrast, we found mixed results for reciprocity, with consistent evidence of positive—but not negative—reciprocity across both studies. This supports the idea that negative and positive reciprocity are indeed independent processes and are not driven by the same motivations (Yamagishi et al., 2012; Ackermann et al., 2014). Our results are also reminiscent of the original findings by Charness and Rabin (2002), where there was evidence of positive reciprocity but fewer acts of negative reciprocity, even when it was free to punish a misbehaving partner. This may reflect similar sentiments as those in the generosity game, in that individuals are generally benevolent—or at least non-spiteful—when the stakes are relatively inexpensive.

In addition, we found no consistent effects for personality with respect to positive or negative reciprocity, suggesting that individual differences in the propensity to reciprocate are subsumed more generally within broader prosocial tendencies.

Other factors may also contribute to the lack of negative reciprocity in our data. First, all decisions in the games were gain-framed. Even when a partner “hurt” a participant, it simply prevented them from receiving a higher amount rather than incurring a personal loss, which may have been too weak to provoke negative reciprocity. Second, the initial payoff combination (15 for the participant, 10 for the partner) declined by the partner in the hurt conditions of the dictator game was already unequal, which may have convinced participants that their partner’s decision to pass on this offer was justified and not deserving of retaliation. Third, the assessment of different social preferences within a single paradigm may trigger a desire among participants to behave consistently, thus artificially increasing consistency in behavior and nullifying any effects for negative reciprocity. However, the differential patterns of responding across generosity and positive reciprocity conditions provide evidence against any such response set. Future investigations using loss-framed manipulations, different configurations of payoffs, and measurements separated by time may be more appropriate for investigating negative reciprocity.

## Women More Egalitarian, Men More Generous?

One interesting finding to emerge across both studies was the interaction between gender and game type, with men consistently allocating more than women in the (costless) generosity games. In the (costly) dictator game, however, women allocated more than men in incentivized games while there were no gender differences in hypothetical responses. But given that decisions in the latter are already a costless form of prosociality—relying on words rather than actions—the absence of a gender gap here may reflect overestimates of allocations among men relative to women. Hence, while women were more inequality averse, they were not necessarily more altruistic when this involved promoting the welfare of others over and above their own.

Although these results were unpredicted and unrelated to the aims of this research, they provide a clear replication of previous research on gender and social preferences. Several studies have shown that women are more prosocial in simple dictator games, while men are more prosocial when the price of giving drops and when giving or cooperating maximizes efficiency (Eckel and Grossman, 1998; Andreoni and Vesterlund, 2001; Croson and Gneezy, 2009; Kuhn and Villeval, 2015). Even in middle childhood and early adolescence, girls more often than boys select egalitarian allocations of wealth over both selfish and generous allocations (Fehr et al., 2013).

These findings correspond to a wider literature on gender differences in preferences toward social and political inequality (i.e., social dominance orientation), which are largely stable across nations and cultures (Pratto et al., 1994, 1997; Sidanius et al., 2000). Such differences in egalitarianism are believed to arise from evolutionary differences in reproductive strategies, in

particular, the accumulation of economic resources and status for male, rather than female, reproductive success (Sidanius et al., 2000). Similarly, the literature on desirable mate qualities and costly signaling indicates that men may engage in greater acts of conspicuous consumption as a display of generosity and resources to increase prestige and status (Griskevicius et al., 2007). In the current paradigm, the safest and easiest way of doing this without putting one's actual stakes at risk is through costless allocations in the generosity game.

## Politeness as a General Prosocial Tendency in Economic Games

A prominent finding was that the politeness aspect of Big Five agreeableness consistently predicted greater overall allocations in both studies. Although we observed a trend for a main effect of compassion when decisions were hypothetical, this disappeared altogether in the incentivized paradigm. These results are in keeping with previous research demonstrating that politeness—rather than compassion—drives egalitarian allocations in the dictator game, with the divergence between the two clearest in incentivized rather than hypothetical paradigms (Zhao et al., 2016).

This unique effect of politeness suggests that prosociality in these decontextualized and neutrally framed paradigms is a function of the tendency to respect others and to adhere to social norms rather than emotional concern for others' wellbeing. While compassion plays a fundamental role in real-world forms of prosociality (Eisenberg and Miller, 1987; Bekkers, 2006) and the related construct of empathy is theorized to be the primary conduit through which humans engage in altruistic behavior (Batson, 1991), compassionate motives may not be elicited given the impersonal nature of economic games. This has important implications for the ecological validity of economic games, suggesting that social preferences and behaviors measured in these games only capture a limited form of norm-based prosociality. Indeed, in their commentary more than 20 years ago, Camerer and Thaler (1995) argued that the outcomes of such games reveal more about the economics of manners and etiquette than they do about altruism, which is empirically supported by the current findings.

## HEXACO Honesty-Humility, Agreeableness, and the Limits of Prosociality

A second major finding to appear consistently across studies was the interaction between honesty-humility and game type, where it predicted greater allocations in the dictator game but played no role in the generosity game. Honesty-Humility has been consistently linked to fair and prosocial (or at least the absence of antisocial) behaviors when there are personal profits to be made, such as delinquency (e.g., stealing money; Dunlop et al., 2012), workplace ethics and integrity (Lee et al., 2005; Cohen et al., 2014), and dishonesty (Hilbig and Zettler, 2015). Notably, across both Big Five and HEXACO models, honesty-humility is the trait most strongly and frequently associated with dictator allocations (Zhao and Smillie, 2015). Surprisingly, we found that

this link to prosociality disappears when such decisions do not involve a personal cost. In the generosity game where individuals could maximize their partner's payoffs for free, those high on honesty-humility allocated no differently from their low-scoring counterparts.

On the one hand, this implies that there are limits on the prosociality encompassed by honesty-humility, and, contrary to previous evidence (Hilbig et al., 2015b), suggests that honesty-humility is more closely tied to egalitarianism and fairness than benevolence. On the other hand, further inspection of Figure 4 shows that this interaction is more strongly driven by those at the *low* pole of honesty-humility. Although they are selfish when they can personally profit, they are neither competitive nor vindictive—and indeed appear concerned about efficiency—once their own stakes are secured.<sup>3</sup> These results highlight the importance of situational context in the expression of personality traits: Given that HEXACO honesty-humility represents the tendency to cooperate with others despite the opportunity for exploitation (i.e., active cooperation), it is no longer elicited when there is no invitation to exploit in the non-constant-sum structure of the generosity game<sup>4</sup>.

It is noteworthy that these findings for honesty-humility were accompanied by a complementary pattern of results for HEXACO agreeableness in the hypothetical games. While HEXACO agreeableness did not predict dictator allocations, consistent with previous research (Hilbig et al., 2013), it was associated with greater allocations in the generosity game. These findings are in keeping with the core features of HEXACO agreeableness, which capture individual differences in tolerance, lenience, flexibility, and a lack of irritability or anger (Ashton and Lee, 2007; Ashton et al., 2014). All of these tendencies are antithetical to spite and envy, two major motivations for curbing a partner's allocation in the generosity game. Nevertheless, we interpret this double dissociation with caution as it was not replicated in the incentivized paradigm, where HEXACO agreeableness was unrelated to any form of social preference.

Interestingly, however, we observed a near-identical interaction for the volatility aspect of Big Five neuroticism in the incentivized paradigm, which captures related constructs (i.e., anger and irritability; DeYoung et al., 2007, DeYoung, 2015) and is strongly negatively associated with HEXACO agreeableness ( $r_{ss} = -0.60, -0.62$  in Studies 1 and 2, respectively). Bivariate correlations and exploratory analyses for volatility (see Supplementary Tables S1 and S6 in the Supplementary Material) showed a similar but inverted pattern to that previously seen for HEXACO agreeableness. Volatility has been linked to psychopathic traits (Jonason et al., 2013), and may provoke envy and resentfulness when individuals face the prospect of disadvantageous inequality in the generosity game.

<sup>3</sup>A future extension is to examine these distinct prosocial traits at an even finer level of analysis, as honesty-humility too can be broken down into four facets: sincerity, greed avoidance, fairness, and modesty (Lee and Ashton, 2004). Our own data indicate a main effect for the modesty facet in both incentivized and hypothetical studies, while the interaction between honesty-humility and game type was driven by sincerity in Study 1 and fairness in Study 2.

<sup>4</sup>We thank one of our reviewers for this observation.

## Does Incentivization Pay Off?

The dual studies and their near-identical designs provide a useful comparison of trait effects across incentivized and hypothetical designs, which has been a topic of debate among psychologists and economists (Camerer and Hogarth, 1999; Ariely and Norton, 2007). Recent investigations of prosocial traits in the dictator game suggest that while average allocations drop between hypothetical and incentivized designs, the effects of certain traits—politeness and agreeableness from the Big Five model, and HEXACO honesty-humility—tend to be larger in incentivized paradigms (Zhao et al., 2016). Likewise, we also observed considerable discrepancies in average allocations between the two paradigms, with individuals overestimating dictator allocations when providing hypothetical responses. Such “hypothetical biases” are frequently seen in value elicitation methods, in which individuals overstate their willingness to pay for a given good (in this case, equality; Hertwig and Ortmann, 2001; List and Gallet, 2001). Yet, individuals also underestimated how benevolent or efficiency-maximizing they would be in the generosity game. It appears that in the absence of incentivization, all individuals gravitate toward the equality norm, leading to attenuated individual variation and muted trait effects. With incentivization, new trait effects emerged, including interactions for compassion and volatility. These can be understood in relation to a recent meta-analysis on the role of personality traits in cooperative game behaviors, which found moderating effects of incentivization on Big Five agreeableness and neuroticism (Ferguson et al., 2015). With incentivization, the effect for agreeableness became stronger while the effect for neuroticism went from weakly positive to negative.

## CONCLUSION

There have been recent calls for an integrated research agenda between personality psychology and economics (Ferguson et al.,

## REFERENCES

- Abibik, K., and Sadrieh, A. (2009). The pleasure of being nasty. *Econ. Lett.* 105, 306–308. doi: 10.1016/j.econlet.2009.08.024
- Ackermann, K. A., Fleiss, J., and Murphy, R. O. (2014). Reciprocity as an individual difference. *J. Conflict Resolut.* 23, 1–28. doi: 10.1177/0022002714541854
- Aghababaei, N., Mohammatabar, S., and Saffarinia, M. (2014). Dirty dozen vs. the H factor: comparison of the dark triad and honesty-humility in prosociality, religiosity, and happiness. *Per. Individ. Dif.* 67, 6–10. doi: 10.1016/j.paid.2014.03.026
- Andreoni, J., and Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *Q. J. Econ.* 116, 293–312. doi: 10.1162/003355301556419
- Ariely, D., and Norton, M. I. (2007). Psychology and experimental economics: a gap in abstraction. *Curr. Dir. Psychol. Sci.* 16, 336–339. doi: 10.1111/j.1467-8721.2007.00531.x
- Ashton, M. C., and Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Per. Soc. Psychol. Rev.* 11, 150–166. doi: 10.1177/108868306294907
- Ashton, M. C., and Lee, K. (2008). The prediction of honesty-humility-related criteria by the HEXACO and five-factor models of personality. *J. Res. Pers.* 42, 1216–1228. doi: 10.1016/j.jrp.2008.03.006
- Ashton, M. C., Lee, K., and de Vries, R. E. (2014). The HEXACO honesty-humility, agreeableness, and emotionality factors: a review of research and theory. *Pers. Soc. Psychol. Rev.* 18, 139–152. doi: 10.1177/108868314523838
- Axelrod, R., and Hamilton, W. D. (1981). The evolution of cooperation. *Science* 211, 1390–1396. doi: 10.1126/science.7466396
- Bardsley, N. (2000). Control without deception: individual behaviour in free-riding experiments revisited. *Exp. Econ.* 240, 215–240. doi: 10.1023/A:1011420500828
- Barford, K. A., Zhao, K., and Smillie, L. D. (2015). Mapping the interpersonal domain: translating between the Big Five, HEXACO, and interpersonal circumplex. *Per. Individ. Dif.* 86, 232–237. doi: 10.1016/j.paid.2015.05.038
- Bartling, B., Fehr, E., Maréchal, M. A., and Schunk, D. (2009). Egalitarianism and competitiveness. *Am. Econ. Rev.* 99, 93–98. doi: 10.1055/s-0033-1343446
- Batson, C. D. (1991). *The Altruism Question: Toward a Social Psychological Answer*. Hillsdale, NJ: Erlbaum.
- Batson, C. D., and Powell, A. A. (2003). “Altruism and prosocial behavior” in *Handbook of psychology: Personality and Social Psychology*, Vol. 5, eds T. Millon, M. J. Lerner, and I. B. Weiner (Hoboken, NJ: John Wiley & Sons), 463–484.
- Baumeister, R. F., Vohs, K. D., and Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? *Psychol. Sci.* 2, 396–403. doi: 10.1111/j.1745-6916.2007.00051.x
- 2011). In the current research, we mapped two models of personality onto individual differences in social preferences using a parsimonious behavioral paradigm. In the HEXACO model, honesty-humility (but not agreeableness) uniquely predicted egalitarian, but not generous, allocations of wealth. In the Big Five model, the politeness (but not compassion) aspect of agreeableness was uniquely associated with prosocial allocations of wealth more globally. The findings revealed important insights concerning the sources of heterogeneity in social preferences and the mechanisms driving prosocial behavior in economic games. Together, they demonstrate the value of a joint approach that combines theoretical predictions from personality psychology with behavioral paradigms from experimental economics.

## AUTHOR CONTRIBUTIONS

Conception and design: LS and KZ. Collection, analysis, and interpretation of data: EF, LS, and KZ. Drafting the article: KZ. Revising the article: EF, LS, and KZ.

## FUNDING

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Preparation of this manuscript was supported by funding from the Melbourne School of Psychological Sciences, The University of Melbourne. KZ was supported by an Australian Postgraduate Award and an Endeavour Research Fellowship.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01137>

- Baumert, A., Schlosser, T., and Schmitt, M. (2014). Economic games: a performance-based assessment of fairness and altruism. *Eur. J. Psychol. Assess.* 30, 178–192. doi: 10.1027/1015-5759/a000183
- Becker, A., Deckers, T., Dohmen, T., Falk, A., and Kosse, F. (2012). The relationship between economic preferences and psychological personality measures. *Annu. Rev. Econ.* 4, 453–478. doi: 10.1146/annurev-economics-080511-110922
- Bekkers, R. (2006). Traditional and health-related philanthropy: the role of resources and personality. *Soc. Psychol. Q.* 69, 349–366. doi: 10.1177/019027250606900404
- Ben-Ner, A., and Halldorsson, F. (2010). Trusting and trustworthiness: what are they, how to measure them, and what affects them. *J. Econ. Psychol.* 31, 64–79. doi: 10.1016/j.joep.2009.10.001
- Ben-Ner, A., Kong, F., and Putterman, L. (2004). Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving. *J. Econ. Psychol.* 25, 581–589. doi: 10.1016/S0167-4870(03)00065-5
- Ben-Ner, A., Kramer, A., and Levy, O. (2008). Economic and hypothetical dictator game experiments: incentive effects at the individual level. *J. Socio-Econ.* 37, 1775–1784. doi: 10.1016/j.socloc.2007.11.004
- Böckler, A., Tusche, A., and Singer, T. (2016). The structure of human prosociality: differentiating altruistically motivated, norm motivated, strategically motivated and self-reported prosocial behavior. *Soc. Psychol. Pers. Sci.* 7, 530–541. doi: 10.1177/1948550616639650
- Boehm, C. (1999). *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Cambridge, MA: Harvard University Press.
- Bolton, G. E., and Ockenfels, A. (2000). ERC: a theory of equity, reciprocity, and competition. *Am. Econ. Rev.* 90, 166–193. doi: 10.1257/aer.90.1.166
- Bowles, S., and Gintis, H. (2004). The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theor. Popul. Biol.* 65, 17–28. doi: 10.1016/j.tpb.2003.07.001
- Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NY: Princeton University Press.
- Camerer, C. F., and Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *J. Risk Uncertain.* 19, 7–42. doi: 10.1023/A:1007850605129
- Camerer, C. F., and Thaler, R. H. (1995). Anomalies: ultimatums, dictators and manners. *J. Econ. Perspect.* 9, 209–219. doi: 10.1257/jep.9.2.209
- Carlo, G., Okun, M. A., Knight, G. P., and de Guzman, M. R. T. (2005). The interplay of traits and motives on volunteering: agreeableness, extraversion and prosocial value motivation. *Pers. Individ. Dif.* 38, 1293–1305. doi: 10.1016/j.paid.2004.08.012
- Charness, G., and Rabin, M. (2002). Understanding social preferences with simple tests. *Q. J. Econ.* 117, 817–869. doi: 10.1017/S1368980012003515
- Cohen, T. R., Panter, A. T., Turan, N., Morse, L., and Kim, Y. (2014). Moral character in the workplace. *J. Pers. Soc. Psychol.* 107, 943–963. doi: 10.1037/a0037245
- Croson, R., and Gneezy, U. (2009). Gender differences in preferences. *J. Econ. Lit.* 47, 448–474. doi: 10.1257/jel.47.2.448
- DeYoung, C. G. (2015). Cybernetic big five theory. *J. Res. Pers.* 56, 33–58. doi: 10.1016/j.jrps.2014.07.004
- DeYoung, C. G., Quilty, L. C., and Peterson, J. B. (2007). Between facets and domains: 10 aspects of the big five. *J. Pers. Soc. Psychol.* 93, 880–896. doi: 10.1037/0022-3514.93.5.880
- Digman, J. M. (1990). Personality structure: emergence of the five-factor model. *Annu. Rev. Psychol.* 41, 417–440. doi: 10.1146/annurev.ps.41.020190.002221
- Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2008). Representative trust and reciprocity: prevalence and determinants. *Econ. Inquiry* 46, 84–90. doi: 10.1111/j.1465-7295.2007.00082.x
- Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2009). Homo reciprocans: survey evidence on behavioural outcomes. *Econ. J.* 119, 592–612. doi: 10.1111/j.1468-0297.2008.02242.x
- Dufwenberg, M., and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games Econ. Behav.* 47, 268–298. doi: 10.1016/j.geb.2003.06.003
- Dunlop, P. D., Morrison, D. L., Koenig, J., and Silcox, B. (2012). Comparing the Eysenck and HEXACO models of personality in the prediction of adult delinquency. *Eur. J. Pers.* 26, 194–202. doi: 10.1002/per.824
- Eckel, C. C., and Grossman, P. J. (1998). Are women less selfish than men? Evidence from dictator experiments. *Econ. J.* 108, 726–735.
- Eisenberg, N., and Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychol. Bull.* 101, 91–119. doi: 10.1037/0033-2950.101.1.91
- Elshaug, C., and Metzer, J. (2001). Personality attributes of volunteers and paid workers engaged in similar occupational tasks. *J. Soc. Psychol.* 141, 752–763. doi: 10.1080/00224540109600586
- Engel, C. (2011). Dictator games: a meta study. *Exp. Econ.* 14, 583–610. doi: 10.1007/s10683-011-9283-7
- Engelmann, B. D., and Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *Am. Econ. Rev.* 94, 857–869. doi: 10.1257/0002828042002741
- Evans, A. M., and Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *J. Res. Pers.* 42, 1585–1593. doi: 10.1016/j.jrp.2008.07.011
- Falk, A., and Fischbacher, U. (2006). A theory of reciprocity. *Games Econ. Behav.* 54, 293–315. doi: 10.1016/j.geb.2005.03.001
- Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149
- Fehr, E., Bernhard, H., and Rockenbach, B. (2008). Egalitarianism in young children. *Nature* 454, 1079–1083. doi: 10.1038/nature07155
- Fehr, E., Fischbacher, U., and Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Hum. Nat.* 13, 1–25. doi: 10.1007/s12110-002-1012-7
- Fehr, E., and Gächter, S. (2000). Fairness and retaliation: the economics of reciprocity. *J. Econ. Perspect.* 14, 159–181. doi: 10.1257/jep.14.3.159
- Fehr, E., Glätzle-Rützler, D., and Sutter, M. (2013). The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence. *Eur. Econ. Rev.* 64, 369–383. doi: 10.1016/j.eurocorev.2013.09.006
- Fehr, E., and Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *Q. J. Econ.* 114, 817–868. doi: 10.1162/00355399556151
- Fehr, E., and Schmidt, K. M. (2006). “The economics of fairness, reciprocity and altruism: experimental evidence and new theories,” in *Handbook of the Economics of Giving, Altruism and Reciprocity*, Vol. 1, eds S.-C. Kolm and J. M. Ythier (Amsterdam: Elsevier), 615–691. doi: 10.1016/S1574-0714(06)00108-6
- Ferguson, E., Gancarczyk, S., Wood, A. M., Delaney, L., and Corr, P. (2015). Personality and pro-social preferences: A meta-analysis. *Paper presented at the Meeting of the International Society for the Study of Individual Differences*, Ontario.
- Ferguson, E., Heckman, J. J., and Corr, P. (2011). Personality and economics: overview and proposed framework. *Pers. Individ. Dif.* 51, 201–209. doi: 10.1016/j.paid.2011.03.030
- Ferguson, E., and Starmer, C. (2013). Incentives, expertise, and medical decisions: testing the robustness of natural frequency framing. *Health Psychol.* 32, 967–977. doi: 10.1037/a0033720
- Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* 71, 397–404. doi: 10.1016/S0165-1765(01)00394-9
- Forsythe, R., Horowitz, J. L., Savin, N. E., and Sefton, M. (1994). Fairness in simple bargaining experiments. *Games Econ. Behav.* 6, 347–369. doi: 10.1006/game.1994.1021
- Fraley, R. C., and Marks, M. J. (2007). “The null hypothesis significance testing debate and its implications for personality research,” in *Handbook of Research Methods in Personality Psychology*, eds R. W. Robins, R. C. Fraley, and R. F. Krueger (New York, NY: Guilford Press), 149–169.
- Funder, D. C. (2001). Personality. *Annu. Rev. Psychol.* 52, 197–221. doi: 10.1146/annurev.psych.52.1.197
- Gallucci, M., and Perugini, M. (2000). An experimental test of a game-theoretical model of reciprocity. *J. Behav. Decis. Mak.* 13, 367. doi: 10.1002/1099-0771(200010/12)13:4<367::AID-BDM357>3.0.CO;2-9
- Goldberg, L. R. (1981). Language and individual differences: the search for universals in personality lexicons. *Rev. Pers. Soc. Psychol.* 2, 141–165.
- Gouldner, A. W. (1960). The norm of reciprocity: a preliminary statement. *Am. Sociol. Rev.* 25, 161–178. doi: 10.2307/2092623
- Graziano, W. G., and Eisenberg, N. (1997). “Agreeableness: a dimension of personality,” in *Handbook of Personality Psychology*, eds R. Hogan, J. Johnson, and S. Briggs (San Diego, CA: Academic Press), 795–824.
- Graziano, W. G., Habashi, M. M., Sheese, B. E., and Tobin, R. M. (2007). Agreeableness, empathy, and helping: a person x situation perspective. *J. Pers. Soc. Psychol.* 93, 583–599. doi: 10.1037/0022-3514.93.4.583

- Griskevicius, V., Tybur, J. M., Sundie, J. M., Cialdini, R. B., Miller, G. F., and Kenrick, D. T. (2007). Blatant benevolence and conspicuous consumption: when romantic motives elicit strategic costly signals. *J. Pers. Soc. Psychol.* 93, 85–102. doi: 10.1037/0022-3514.93.1.85
- Güth, W. (2010). The generosity game and calibration of inequity aversion. *J. Socio-Econ.* 39, 155–157. doi: 10.1016/j.soc.2009.10.012
- Güth, W., Levati, M., and Ploner, M. (2009). *Making the world a better place: Experimental evidence from the generosity game*. Jena Economic Research Paper, No. 2009-071. Jena: Friedrich Schiller University and Max Planck Institute of Economics.
- Güth, W., Levati, M. V., and Ploner, M. (2012). An experimental study of the generosity game. *Theory Decis.* 72, 51–63. doi: 10.1007/s11238-011-9253-8
- Halevy, N., Chou, E. Y., and Murnighan, J. K. (2012). Mind games: the mental representation of conflict. *J. Pers. Soc. Psychol.* 102, 132–148. doi: 10.1037/a0025389
- Haley, K. J., and Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evol. Hum. Behav.* 26, 245–256. doi: 10.1016/j.evolhumbehav.2005.01.002
- Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *Am. Psychol.* 58, 78–79. doi: 10.1037/0003-066X.58.1.78
- Hertwig, R., and Ortmann, A. (2001). Experimental practices in economics: a methodological challenge for psychologists? *Behav. Brain Sci.* 24, 383–451.
- Hilbig, B. E., Glöckner, A., and Zettler, I. (2014). Personality and prosocial behavior: linking basic traits and social value orientations. *J. Pers. Soc. Psychol.* 107, 529–539. doi: 10.1037/a0036074
- Hilbig, B. E., Thielmann, I., Hepp, J., Klein, S., and Zettler, I. (2015a). From personality to altruistic behavior (and back): evidence from a double-blind dictator game. *J. Res. Pers.* 55, 46–50. doi: 10.1016/j.jrp.2014.12.004
- Hilbig, B. E., Thielmann, I., Wührl, J., and Zettler, I. (2015b). From honesty-humility to fair behavior – benevolence or a (blind) fairness norm? *Pers. Individ. Dif.* 80, 91–95. doi: 10.1016/j.paid.2015.02.017
- Hilbig, B. E., and Zettler, I. (2009). Pillars of cooperation: honesty-humility, social value orientations, and economic behavior. *J. Res. Pers.* 43, 516–519. doi: 10.1016/j.jrp.2009.01.003
- Hilbig, B. E., and Zettler, I. (2015). When the cat's away, some mice will play: a basic trait account of dishonest behavior. *J. Res. Pers.* 57, 72–88. doi: 10.1016/j.jrp.2015.04.003
- Hilbig, B. E., Zettler, I., Leist, F., and Heydasch, T. (2013). It takes two: honesty-humility and agreeableness differentially predict active versus reactive cooperation. *Pers. Individ. Dif.* 54, 598–603. doi: 10.1016/j.paid.2012.11.008
- Hirsh, J. B., DeYoung, C. G., Xu, X., and Peterson, J. B. (2010). Compassionate liberals and polite conservatives: associations of agreeableness with political ideology and moral values. *Pers. Soc. Psychol. Bull.* 36, 655–664. doi: 10.1177/014616721036684
- Jensen, K., Vaish, A., and Schmidt, M. F. H. (2014). The emergence of human prosociality: aligning with others through feelings, concerns, and norms. *Front. Psychol.* 5:822. doi: 10.3389/fpsyg.2014.00822
- John, O. P., Naumann, L. P., and Soto, C. J. (2008). "Paradigm shift to the integrative big five trait taxonomy: history, measurement, and conceptual issues," in *Handbook of Personality: Theory and Research*, Vol. 3, eds O. P. John, R. W. Robins, and L. A. Pervin (New York, NY: Guilford Press), 114–158.
- Jonason, P. K., Kaufman, S. B., Webster, G. D., and Geher, G. (2013). What lies beneath the dark triad dirty dozen: varied relations with the big five. *Individ. Dif. Res.* 11, 81–90.
- Kagel, J., and McGee, P. (2014). Personality and cooperation in finitely repeated prisoner's dilemma games. *Econ. Lett.* 124, 274–277. doi: 10.1016/j.econlet.2014.05.034
- Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1986). Fairness and the assumptions of economics. *J. Bus.* 59, S285–S300. doi: 10.1086/296367
- Kettner, S. E., and Waichman, I. (2016). Old age and prosocial behavior: social preferences or experimental confounds? *J. Econ. Psychol.* 53, 118–130. doi: 10.1016/j.jeop.2016.01.003
- Knafo-Noam, A., Uzefovsky, F., Israel, S., Davidov, M., and Zahn-Waxler, C. (2015). The prosocial personality and its facets: genetic and environmental architecture of mother-reported behavior of 7-year-old twins. *Front. Psychol.* 6:112. doi: 10.3389/fpsyg.2015.00112
- Komorita, S. S., and Parks, C. D. (1999). "Reciprocity and cooperation in social dilemmas: Review and future directions," in *Games and Human Behavior: Essays in Honor of Amnon Rapoport*, ed. R. Zwick (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 315–330.
- Kuhn, P., and Villeval, M. C. (2015). Are women more attracted to cooperation than men? *Econ. J.* 125, 115–140. doi: 10.1111/eco.12122
- Lee, K., and Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behav. Res.* 39, 329–358. doi: 10.1207/s15327906mbr3902\_8
- Lee, K., and Ashton, M. C. (2012). Getting mad and getting even: agreeableness and honesty-humility as predictors of revenge intentions. *Pers. Individ. Dif.* 52, 596–600. doi: 10.1016/j.paid.2011.12.004
- Lee, K., Ashton, M. C., and de Vries, R. E. (2005). Predicting workplace delinquency and integrity with the HEXACO and five-factor models of personality structure. *Hum. Perform.* 18, 179–197. doi: 10.1207/s15327043hup1802\_4
- Li, J. A., and Chen, Y. (2012). *Personality, Affects, and Forgiving Behavior in Games (RAND Labor and Population Working Paper, No. WR-952)*. Santa Monica, CA: RAND Corporation.
- List, J. A., and Gallet, C. A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values? *Environ. Res. Econ.* 20, 241–254. doi: 10.1023/A:1012791822804
- Loewenstein, G. F., Thompson, L., and Bazerman, M. H. (1989). Social utility and decision making in interpersonal contexts. *J. Pers. Soc. Psychol.* 57, 426–441. doi: 10.1037/0022-3514.57.3.426
- Lönnqvist, J.-E., Verkasalo, M., and Walkowitz, G. (2011). It pays to pay – big five personality influences on co-operative behaviour in an incentivized and hypothetical prisoner's dilemma game. *Pers. Individ. Dif.* 50, 300–304. doi: 10.1016/j.paid.2010.10.009
- Mar, R. A., Spreng, R. N., and DeYoung, C. G. (2013). How to produce personality neuroscience research with high statistical power and low additional cost. *Cogn. Affect. Behav. Neurosci.* 13, 674–685. doi: 10.3758/s13415-013-0202-6
- Mehta, P. (2007). *The Endocrinology of Personality, Leadership, and Economic Decision Making* Doctoral dissertation, The University of Texas at Austin, Austin, TX.
- Moorlock, G., Ives, J., and Draper, H. (2014). Altruism in organ donation: an unnecessary requirement? *J. Med. Ethics* 40, 134–138. doi: 10.1136/medethics-2012-100528
- Müller, J., and Schwieren, C. (2012). *What Can the Big Five Personality Factors Contribute to Explain Small-Scale Economic Behavior? Tinbergen Institute Discussion Paper, No. TI 2012-028/1*. Amsterdam: Tinbergen Institute.
- Murphy, R. O., and Ackermann, K. A. (2014). Social value orientation: Theoretical and measurement issues in the study of social preferences. *Pers. Soc. Psychol. Rev.* 18, 123–139. doi: 10.1177/1088868313501745
- Osborne, D., Wootton, L. W., and Sibley, C. G. (2013). Are liberals agreeable or not? Politeness and compassion differentially predict political conservatism via distinct ideologies. *Soc. Psychol.* 44, 354–360. doi: 10.1027/1864-9335/a000132
- Ozer, D. J., and Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annu. Rev. Psychol.* 57, 401–421. doi: 10.1146/annurev.psych.57.102904.190127
- Penner, L. A., Dovidio, J. F., Piliavin, J. A., and Schroeder, D. A. (2005). Prosocial behavior: multilevel perspectives. *Annu. Rev. Psychol.* 56, 365–392. doi: 10.1146/annurev.psych.56.091103.070141
- Perugini, M., Gallucci, M., Presaghi, F., and Ercolani, A. P. (2003). The personal norm of reciprocity. *Eur. J. Pers.* 17, 251–283. doi: 10.1002/per.474
- Peysakhovich, A., Nowak, M. A., and Rand, D. G. (2014). Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nat. Commun.* 5, 1–8. doi: 10.1038/ncomms5939
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., and Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J. Appl. Psychol.* 88, 879–903. doi: 10.1037/0021-9010.88.5.879
- Pratto, F., Sidanius, J., Stallworth, L. M., and Malle, B. F. (1994). Social dominance orientation: a personality variable predicting social and political attitudes. *J. Pers. Soc. Psychol.* 67, 741–763. doi: 10.1037/0021-9010.67.4.741
- Pratto, F., Stallworth, L. M., and Sidanius, J. (1997). The gender gap: differences in political attitudes and social dominance orientation. *Br. J. Soc. Psychol.* 36, 49–68. doi: 10.1111/j.2044-8309.1997.tb01118.x

- Richard, F. D., Bond, C. F., and Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Rev. Gen. Psychol.* 7, 331–363. doi: 10.1037/1089-2680.7.4.331
- Roberts, B. W., and Robins, R. W. (2000). Broad dispositions, broad aspirations: the intersection of personality traits and major life goals. *Pers. Soc. Psychol. Bull.* 26, 1284–1296. doi: 10.1177/0146167200262009
- Rolison, J. J., Hanoch, Y., and Gummerum, M. (2013). Characteristics of offenders: the HEXACO model of personality as a framework for studying offenders' personality. *J. Forens. Psychiatry Psychol.* 24, 71–82. doi: 10.1080/14789949.2012.752024
- Saunders, B. (2012). Altruism or solidarity? The motives for organ donation and two proposals. *Bioethics* 26, 376–381. doi: 10.1111/j.1467-8519.2012.01989.x
- Shepherd, L., O'Carroll, R. E., and Ferguson, E. (2014). An international comparison of deceased and living organ donation/transplant rates in opt-in and opt-out systems: a panel study. *BMC Med.* 12:131. doi: 10.1186/s12916-014-0131-4
- Sidanius, J., Levin, S., Liu, J., and Pratto, F. (2000). Social dominance orientation, anti-egalitarianism and the political psychology of gender: an extension and cross-cultural replication. *Eur. J. Soc. Psychol.* 30, 41–67. doi: 10.1002/(SICI)1099-0992(200001/02)30:1<41::AID-EJSP976>3.0.CO;2-O
- Singer, T., and Steinbeis, N. (2009). Differential roles of fairness- and compassion-based motivations for cooperation, defection, and punishment. *Ann. N. Y. Acad. Sci.* 1167, 41–50. doi: 10.1111/j.1749-6632.2009.04733.x
- Thielmann, I., and Hilbig, B. E. (2014). Trust in me, trust in you: a social projection account of the link between personality, cooperativeness, and trustworthiness expectations. *J. Res. Pers.* 50, 61–65. doi: 10.1016/j.jrps.2014.03.006
- Thielmann, I., and Hilbig, B. E. (2015). The traits one can trust: dissecting reciprocity and kindness as determinants of trustworthy behavior. *Pers. Soc. Psychol. Bull.* 41, 1523–1536. doi: 10.1177/0146167215600530
- Thielmann, I., Hilbig, B. E., and Niedtfeld, I. (2014). Willing to give but not to forgive: borderline personality features and cooperative behavior. *J. Pers. Dis.* 28, 778–795. doi: 10.1521/pedi\_2014\_28\_135
- Tisserand, J.-C., Cochard, F., and Le Gallo, J. (2015). *Altruistic or Strategic Considerations: A Meta-Analysis on the Ultimatum and Dictator Games*. Besançon: CRESE, Université de Franche-Comté.
- Van Doesum, N. J., Van Lange, D. A. W., and Van Lange, P. A. M. (2013). Social mindfulness: skill and will to navigate the social world. *J. Pers. Soc. Psychol.* 105, 86–103. doi: 10.1037/a0032540
- Vazire, S. (2016). Getting papers accepted in social/personality journals post replicability crisis. *Paper Presented at the Annual Convention of the Society for Personality and Social Psychology*. San Diego, CA.
- Volk, S., Thöni, C., and Ruigrok, W. (2011). Personality, personal values and cooperation preferences in public goods games: a longitudinal study. *Pers. Individ. Dif.* 50, 810–815. doi: 10.1016/j.paid.2011.01.001
- Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., et al. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20364–20368. doi: 10.1073/pnas.1212126109
- Yamagishi, T., Mifune, N., Li, Y., Shinada, M., Hashimoto, H., Horita, Y., et al. (2013). Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality. *Organ. Behav. Hum. Decis. Process.* 120, 260–271. doi: 10.1016/j.obhdp.2012.06.002
- Zhang, L., and Ortmann, A. (2016). Pro-social or anti-social, or both? A within- and between-subjects study of social preferences. *J. Behav. Exp. Econ.* 62, 23–32. doi: 10.1016/j.soec.2016.03.001
- Zhao, K., Ferguson, E., and Smillie, L. D. (2016). Individual differences in good manners rather than compassion predict fair allocations of wealth in the dictator game. *J. Pers.* doi: 10.1111/jopy.12237 [Epub ahead of print].
- Zhao, K., and Smillie, L. D. (2015). The role of interpersonal traits in social decision making: exploring sources of behavioral heterogeneity in economic games. *Pers. Soc. Psychol. Rev.* 19, 277–302. doi: 10.1177/108868314553709

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Zhao, Ferguson and Smillie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Cognitive Reflection, Decision Biases, and Response Times

**Carlos Alós-Ferrer<sup>\*</sup>, Michele Garagnani and Sabine Hügelschäfer**

*Department of Economics, University of Cologne, Cologne, Germany*

We present novel evidence on response times and personality traits in standard questions from the decision-making literature where responses are relatively slow (medians around half a minute or above). To this end, we measured response times in a number of incentivized, framed items (decisions from description) including the Cognitive Reflection Test, two additional questions following the same logic, and a number of classic questions used to study decision biases in probability judgments (base-rate neglect, the conjunction fallacy, and the ratio bias). All questions create a conflict between an intuitive process and more deliberative thinking. For each item, we then created a non-conflict version by either making the intuitive impulse correct (resulting in an alignment question), shutting it down (creating a neutral question), or making it dominant (creating a heuristic question). For CRT questions, the differences in response times are as predicted by dual-process theories, with alignment and heuristic variants leading to faster responses and neutral questions to slower responses than the original, conflict questions. For decision biases (where responses are slower), evidence is mixed. To explore the possible influence of personality factors on both choices and response times, we used standard personality scales including the Rational-Experiential Inventory and the Big Five, and used them as controls in regression analysis.

## OPEN ACCESS

**Edited by:**

Nikolaos Georgantzis,  
*University of Reading, UK*

**Reviewed by:**

Michael Roy,  
*Elizabethtown College, USA*  
 Conny Ernst-Peter Wollbrant,  
*University of Gothenburg, Sweden*

**\*Correspondence:**

Carlos Alós-Ferrer  
 carlos.alos-ferrer@uni-koeln.de

**Specialty section:**

This article was submitted to  
*Personality and Social Psychology*,  
 a section of the journal  
*Frontiers in Psychology*

**Received:** 15 July 2016

**Accepted:** 01 September 2016

**Published:** 22 September 2016

**Citation:**

Alós-Ferrer C, Garagnani M and  
 Hügelschäfer S (2016) Cognitive  
 Reflection, Decision Biases, and  
 Response Times.  
*Front. Psychol.* 7:1402.  
 doi: 10.3389/fpsyg.2016.01402

## 1. INTRODUCTION

Human beings attempt to behave rationally, but they often struggle as intuitive impulses get in the way. Sometimes the latter are useful, sometimes they invite disaster. Modern economic thinking is shaping the view that decisions are often the result of the interaction between fast intuitive thinking and the attempt (often unsuccessful) to behave in a rational way. While neoclassic economics concentrated on rationalistic behavior, other branches as, e.g., the literature on learning in games (following Kandori et al., 1993; Young, 1993) focused on the study of behavioral rules of thumb. More recently, dual-process models from psychology (Epstein, 1994; Sloman, 1996; Strack and Deutsch, 2004; Evans, 2008; Alós-Ferrer and Strack, 2014) have received increasing attention in economics. These models postulate decision-process heterogeneity at the intra-individual level, that is, the interaction of more intuitive and more deliberative processes within a decision maker's mind.

Individual heterogeneity, however, remains an important topic. Across individuals, heterogeneity concerns whether each particular decision maker relies more or less on one or the other kind of process. To measure this dimension, a number of scales and questionnaires have been developed. Among them are the Rational-Experiential Inventory of Epstein et al. (1996), including its two subscales *Faith in Intuition* (FI) and *Need for Cognition* (NFC), and the three-item

Cognitive Reflection Test (CRT) of Frederick (2005), recently expanded by Toplak et al. (2014) and Primi et al. (2015). A recent branch of the literature has investigated interindividual differences regarding faulty probability judgments (heuristics and biases) using these scales. Oechssler et al. (2009) and Hoppe and Kusterer (2011) find that higher test scores in the CRT are correlated with lower incidences of certain biases, e.g., the conjunction fallacy (Tversky and Kahneman, 1983). As argued by Toplak et al. (2011), low CRT scores might indicate a tendency to act on impulse and give an intuitive response. Alós-Ferrer and Hügelschäfer (2012, 2016) showed that higher scores in Faith in Intuition are associated with higher error rates aligned with certain heuristics, e.g., based on representativeness or reinforcement, but found no systematic relation between the CRT and FI.

This work continues the exploration of individual differences in faulty probability judgments and extends previous works by considering process data. The dual-process literature naturally relies on process data for the analysis of multi-process decisions, an approach which allows inferences which would be impossible with choice data only. The simplest kind of process data arises from response times. However, the heuristics-and-biases literature typically relies on decisions made on the basis of verbal descriptions, that is, on relatively complex, non-repeatable questions related to a more or less artificial situation (as for instance, the LINDA problem from Tversky and Kahneman, 1983). The use of response times in such a setting faces two main difficulties.

The first difficulty is that within-subject comparisons for a single question are not possible. However, precisely those are the standard for response-times studies. In many behavioral studies, decisions are made in paradigms which allow for repetition, sometimes even for a large number of trials for each individual participant. In these cases, one can compare the response times of different responses for the same individual, which allows predictions linked to the very nature of processes. For instance, if (in an extreme case) it is assumed that a certain response overwhelmingly follows from a certain intuitive process, while another response overwhelmingly follows from a more deliberative one, one would predict the first response to be on average faster, simply because intuitive processes are faster. In a typical description-based decision, however, a paragraph-long decision situation is presented, the participant makes a decision, and moves on to a different question. Hence there is a unique observation per participant, which is either correct or not. It is not possible to test hypotheses on the relative speed of different responses, because such comparisons would be confounded with personal characteristics. For instance, if a process-based model predicted errors to be faster than correct responses in a given situation, and even if this prediction were correct, one might obtain the opposite result if participants giving correct responses had higher cognitive abilities, and the latter were associated with faster response times for the given situation.

For instance, Achtziger and Alós-Ferrer (2014) study a paradigm where a reinforcement-based heuristic can conflict or be aligned with more rational decision making (optimization based on Bayesian updating of beliefs). The main predictions

of the study (following the Dual-Process Diffusion Model, Alós-Ferrer, 2016) concern the relative speed of errors and correct responses *for a given individual*, i.e., a within-participant comparison. Those are testable because the paradigm allows for repetition, with 60 different decisions per participant, and hence one typically has multiple errors and multiple correct responses for a participant. In a paradigm with one decision per participant (say, measuring the CRT), errors and correct decisions can simply not be compared within participants.

The second difficulty is that, when measuring biases in probability judgments through standard decisions from description, response times are relatively long. In contrast, the dual-process literature focuses on rather short response times (a few seconds at most). Long response times (say, around half a minute) will always include some deliberation, and hence any response-time differences accruing from intrinsic differences among the decision processes involved are likely to be washed away (see, e.g., Myrseth and Wollbrant, 2016). However, this does not mean that long response times are useless. It is a well-established fact that decisions where the decision maker faces stronger tradeoffs, or is “closer to indifference,” are harder and result in longer response times (Dashiell, 1937; Mosteller and Nogee, 1951). This fact can be extended to longer response times, capturing the intuition that if one alternative is clearly preferred, a fast decision ensues, but if two alternatives are similarly desirable, an inner struggle results in a slower decision. Following this logic, longer response times should be considered evidence of longer deliberation due to opposed tendencies.

In view of these difficulties, our study focused on within-subject comparisons across different questions. To this purpose, we created a number of alternative versions of well-established questions. The logic is as follows. Many of the questions used to study biases in probability judgment pit the correct response against an intuitive alternative favored by a heuristic. For instance, in the LINDA question, an incorrect response is intuitively attractive because it is stereotype-consistent. The same is true for the items in the CRT, where an intuitive response conflicts with the correct one. To examine process data associated with the conflict, we created non-conflict versions of those questions. Depending on the content of the question, however, one ends with qualitatively different non-conflict items. In some cases it is possible to turn around the question in such a way that the intuitive process will remain active and favor the correct response. We refer to the resulting items as *alignment questions*, because both processes remain active but are aligned in terms of prescribed choices. In other cases, however, it is not possible to force the intuitive process to favor the correct response. The conflict can still be removed by shutting down the intuitive process (removing the cue on which it acts), creating a *neutral* version of the original question. In one extreme case, however, this manipulation was not possible, but it was still possible to create a non-conflict version where the heuristic points to the correct answer, but where the exact process (type of computation) underlying the deliberative process in the conflict version does not apply. The resulting altered item is called a *heuristic question*.

To the best of our knowledge, there is no systematic study on response times for this type of questions. Hence, the analysis in

this article is novel but exploratory. We collected choice data in a laboratory environment where participants answered a series of standard questions regarding probability judgments, the original CRT of Frederick (2005), and additional items from the extended CRT of Toplak et al. (2014). Crucially, we measured response times for those decisions. Additionally, we included a number of questionnaires measuring personality differences, including the short version of the Rational-Experiential Inventory of Epstein et al. (1996) (comprising FI and NFC) and the Big Five (McCrae and Costa, 1985).

The paper proceeds as follows. Section 2 details the experimental design and describes the sample, the methods, and the natural hypotheses regarding response times. Section 3 presents some preliminary, descriptive results of correlational nature. Section 4 presents results for the (extended) CRT questions, including evidence on response times. Section 5 presents the results for behavioral biases, including the relation to the CRT and evidence on response times. Section 6 concludes.

## 2. METHODS

### 2.1. Experimental Design

We investigated decision processes by measuring both choices and response times for a series of incentivized context-embedded scenarios (“decisions from description”). We focused on two types of problems. First, we employed items from the Cognitive Reflection Test introduced by Frederick (2005) and further extended by Toplak et al. (2014). Second, we used a sample of questions tackling typical decision biases in the domain of belief updating and probabilistic judgment, capturing the conjunction fallacy (Tversky and Kahneman, 1983), base-rate neglect (Kahneman and Tversky, 1972; Fiedler, 2000; Erev et al., 2008), and the ratio bias (Kirkpatrick and Epstein, 1992; Denes-Raj and Epstein, 1994).

For both types of problems, questions are assumed to create a situation of conflict between an “intuitive” answer favored by a certain heuristic process and the (normatively) correct response. We complemented each question with a non-conflict version, hence creating several pairs of items. We developed three categories of non-conflict versions. For some of the questions, we created alignment versions where the intuitive answer and the normatively correct answer coincide. For others we created neutral versions where the heuristic does not apply, so that there is no intuitive first answer. Further, for one of the CRT questions we created a heuristic version where the heuristic points to the correct answer, but where the computation process leading to the correct answer in the conflict version does not apply.

Presenting two versions of the same question within one experiment might potentially direct the participants’ attention to the deceitful property of these questions. To reduce this problem while keeping the rationale of the questions intact, the surface similarity between two paired items was reduced by using different contextual and numerical contents (see, e.g., De Neys et al., 2013). For the comparison of response times to be meaningful, we matched the length of the items for each pair (all items were translated to German as we relied on a sample of German-speaking participants). That is, we adapted the wording

of the questions to guarantee that the number of sentences was always the same for each pair. Further, the number of words, characters, and syllables of the German translations did not differ by more than 10% across the questions of a given pair. To this aim, in some cases we made slight cosmetic changes to the wording of the questions taken from the literature.

Overall, our sample of questions comprised the following items: Two pairs from the classic CRT (Frederick, 2005), plus the third original CRT item (without a matched non-conflict version) to be able to compute a CRT score for each participant; two pairs from the extended CRT by Toplak et al. (2014) (for other questions it was not possible to create non-conflict versions); a quartet referring to the conjunction fallacy; three pairs referring to base-rate neglect; and one pair referring to the ratio bias.

In addition, we investigated several individual correlates of the reliance on intuitive vs. deliberative decision making: Faith in Intuition and Need for Cognition (Epstein et al., 1996), Actively Open-Minded Thinking (Baron, 1993) (respectively referred to as FI, NFC, and AOT hereafter), and the Big Five personality scales (McCrae and Costa, 1985). We also controlled for numerical literacy (Lipkus et al., 2001), gender, and individual swiftness.

### 2.2. Participants

Participants were recruited using ORSEE (Greiner, 2004), a standard online recruitment system for economic experiments which allows for random recruitment from a predefined subject pool. Participants were native German-speaking students from the University of Cologne (Germany), excluding students majoring in psychology or economics. We only considered native speakers due to our focus on response times, since those are critically related to participants’ language skills for the text-based problems we used. In addition, our recruiting rules excluded participants who had previously participated in any experiment employing the CRT. A total of 158 participants (101 female; age range 18 – 44, mean 23.44) participated in exchange for performance-based payment plus a show-up fee of 4 Euros. Three further participants had to be excluded from data analysis because they did not comply with the instructions.

### 2.3. Procedure

The experiment was conducted at the Cologne Laboratory for Economic Research (CLER) using z-Tree (Fischbacher, 2007). Experimental procedures were in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments, and also standard practices in experimental economics (e.g., no-deception rule). In agreement with the ethics and safety guidelines at the CLER, participants were all pre-registered in the laboratory through ORSEE and had given written informed consent regarding the laboratory’s guidelines (no further informed consent is necessary for particular experiments). Potential participants were informed of their right to abstain from participation in the study or to withdraw consent to participate at any time without reprisal.

In a first phase, participants were asked 21 incentivized questions. Specifically, at the end of the experiment they received 0.50 Euro cent for each correct answer. These questions

comprised the (extended) CRT (9 items), the conjunction fallacy (4 items), base-rate neglect (6 items), and the ratio bias (2 items). All but two of the CRT items had to be answered in open format. That is, participants were required to type their numerical response into a blank box. The remaining two CRT items and all other questions were multiple-choice items with two or more possible answers each.

To control for possible order effects, participants were randomly assigned to four different counterbalance conditions (pseudo-randomized question order).<sup>1</sup> For each pair, half of the participants worked on the conflict version before the non-conflict version, whereas the other half started with the non-conflict version. In addition, for each participant, half of the item pairs were first shown in the conflict version and later in the non-conflict version, and vice versa for the other half of pairs. Further, the two items of each pair were separated by at least three other items.

In a second phase, participants worked on the 11 items of the numeracy scale (Lipkus et al., 2001). They were informed that the computer would randomly draw one of the 11 items at the end of the experiment, and that they would receive 0.50 Euro cent if their answer to the selected item was correct.

In a third phase, which was not incentivized, participants completed the self-report questionnaires. Those included FI and NFC (measured by means of the 10-item Rational-Experiential Inventory; Epstein et al., 1996), the Big Five Inventory-SOEP (15-item version; Gerlitz and Schupp, 2005), and AOT (7-item version by Haran et al., 2013). Participants rated questionnaire items by placing marks on continuous left-right scales ranging from 0 ("completely false") to 10 ("completely true"). An exception was AOT, which was rated on a 7-point scale for each item. Since participants might have been exposed to the CRT items in their daily life (e.g., through the press or the internet), we also asked them to indicate whether they had previously seen each of the classic CRT items. Finally, the questionnaire comprised socio-demographic questions (gender, age, and native language).

No time limit was imposed; participants were free to use as much time as needed for the incentivized questions and the questionnaires. As a proxy for swiftness (see Cappelen et al., 2015), we measured the time it took participants to read the brief introductory instructions for phases one and two, and the time it took them to answer the questions about age, gender, and native language in phase three. The sum of these two measures (reading time and demographic answer time in seconds) was used to create an (inverse) index of swiftness.

Payment was computed at the end of the experiment. A session lasted about 50 min and average earnings were 12.24 Euros ( $SD = 1.28$ ).

## 2.4. Basic Hypotheses for Response Times

Our basic hypotheses concern the comparison of response times for paired conflict and non-conflict questions. Following a dual-process logic (e.g., the Dual-Process Diffusion Model of Alós-Ferrer, 2016), the response time for a question where there is

a conflict between an intuitive and a deliberative process can be decomposed in two parts. First, the time needed for conflict detection and resolution. Second, the actual process time, that is, the time needed by the process which actually generates the response to do so. Let  $D_C$  be the expected time necessary for conflict detection and resolution in the presence of an actual decision conflict. Further, let  $T_H$  be the expected response time of the intuitive (heuristic) process, and let  $T_U > T_H$  be the expected response time of the deliberative (utilitarian) process (please note that, to simplify notation, all quantities are expected times).

Actual response time will be the sum of conflict detection and resolution time and process time. However, depending on conflict resolution, the process actually delivering the response might be either the intuitive or the deliberative one. Since we only observe one decision for a given participant, the expected response time is hence  $D_C + T_H$  or  $D_C + T_U$ , depending on which process is selected. The problem, of course, is that the actually selected process is unobservable. If a large enough set of answers for a fixed question was observed, the total expected response time would be

$$D_C + \Delta T_H + (1 - \Delta)T_U$$

where  $\Delta$  is the probability that the intuitive process is the one actually delivering the response.

These considerations are useful to derive experimental hypotheses for the comparison of response times across questions. Consider an alignment question where the conflict has been removed because both processes prescribe the same answer. Two effects can be expected. First, the conflict detection and resolution time  $D_C$  will be reduced, since there is no actual conflict. Second, there will be an increase in the probability  $\Delta$  that the faster, intuitive process is used, since there is no need to inhibit it (or, in other words, we have more observations of the type  $D_C + T_H$  than of the type  $D_C + T_U$ ). Both effects point in the same direction and deliver the following experimental hypothesis.

**H1.** Response times for alignment questions are shorter than response times for the analogous conflict questions.

Consider now a neutral question, where the intuitive process has been shut down by removing the cue on which it acts. The conflict detection and resolution time  $D_C$  will also be reduced in this case (absence of conflict). However, the probability that the intuitive process is actually used becomes  $\Delta = 0$ . Hence response times will be shorter with respect to conflict detection but all decisions will arise from the slower, deliberative process. Evidence from neuroscience points out that conflict detection and resolution occurs extremely early in decision making (see, e.g., Achtziger et al., 2014) and hence should have a moderate effect in response times of large magnitude. In Achtziger and Alós-Ferrer (2014), decisions where a reinforcement heuristic had been shut down were observed to be significantly slower (and error rates significantly lower) than decisions where the heuristic was active. On the basis of this evidence, we formulate the following hypothesis.

<sup>1</sup>The counterbalance condition did not significantly affect participants' response times or responses to any of the questions.

**H2.** Response times for neutral questions are longer than response times for the analogous conflict questions.

However, alternative hypotheses might also be reasonable. Following the interpretation of long response times as evidence for deliberative struggle, one could speculate that the presence of conflict in decisions as the ones considered here has an effect beyond conflict detection and resolution. However, at this point there is no empirical basis for a comparison of the magnitude of this effect and the slowing-down of decisions in neutral questions due to the shutdown of the intuitive process.

In one case, the non-conflict question involves the intuitive process becoming prescriptively correct while the original deliberative process is shut down (heuristic question). In this case, again  $D_C$  should be reduced, and either the likelihood of the intuitive process being selected should become  $\Delta = 1$ , or the deliberative process should be replaced with another, simpler and presumably faster one. In both cases, we would expect to observe faster decisions.

**H3.** Response times for heuristic questions are shorter than response times for the analogous conflict questions.

### 3. DESCRIPTIVE RESULTS

#### 3.1. Summary Statistics and Gender Effects

**Table 1** displays summary statistics for the main dependent variables and reports the presence or absence of gender differences (via Wilcoxon Rank-Sum tests on the whole sample). On average, participants correctly answered two out of the three classic CRT items by Frederick (2005), and one out of the two extended CRT items by Toplak et al. (2014). In line with previous studies (Frederick, 2005; Oechssler et al., 2009; Brañas-Garza et al., 2012; Alós-Ferrer and Hügelshäfer, 2016; Cueva et al., 2016), males had significantly higher scores in the classic three-item CRT; there was no difference concerning CRT2. The results regarding pre-experimental knowledge of the classic CRT imply that the test is becoming common knowledge in the student population: 13.92% of participants reported knowing one question, 26.58% two questions, and 36.08% all three. Participants with more previous knowledge of the items obtain significantly higher classic-CRT scores (Spearman's correlation,  $\rho = 0.307, p < 0.0001$ ).

Descriptive statistics for the numeracy scale (Lipkus et al., 2001) suggest that this measure is not particularly well-suited to capture interindividual differences. It exhibits a very low variance, with most of our participants answering either 10 or 11 out of 11 items correctly. Still, there is a significant gender difference, pointing to higher numeracy for males. Regarding personality traits, we find higher values of NFC for male compared to female participants, in line with previous research (Pacini and Epstein, 1999). Female participants have higher scores for Extraversion and Neuroticism, which is consistent with the literature (e.g., Feingold, 1994; Weisberg et al., 2011).

#### 3.2. Personality Measures

**Table 2** displays Spearman rank correlations among personality traits. We include numerical literacy, but this measure shows

no correlation with any of the personality traits. In contrast to theoretical assumptions of the Rational-Experiential Inventory (REI) (Epstein, 1994; Epstein et al., 1996), there is a weak positive correlation between FI and NFC in our sample (Spearman's correlation,  $\rho = 0.14, p = 0.079$ ). Concerning the relation between the REI and the Big Five, we found that FI is positively associated with Openness to Experience, Conscientiousness, and Extraversion, while NFC is positively correlated with Openness to Experience and Conscientiousness, and negatively with Neuroticism. These results are perfectly consistent with the findings of Pacini and Epstein (1999). The significant positive correlations of AOT with NFC and Openness are in line with results by Haran et al. (2013).

### 4. EXTENDED CRT QUESTIONS

For the analysis of response times, in a first step we removed outliers in order to exclude abnormal observations that might bias the results. To this end, we removed, for each item, response times that deviated more than two standard deviations from the respective mean of the whole sample of participants (see Miller, 1991, on this). This led to the exclusion of several very slow responses, but not of very fast ones. Further, we excluded response times of zero, which resulted from a few participants accidentally skipping a question by double-clicking. Hence, for every paired-observations test across the two questions in a pair, participants whose response times were outliers in either of the two questions are removed. In order to test our hypotheses on response times, we use non-parametric, two-tailed Wilcoxon Signed-Rank tests (for paired observations). To compare error rates across the two questions in a pair, we rely on McNemar's chi-squared test, which is based on the number of discordant pairs. For ease of presentation, instead of repeating the exclusion criteria for every single item pair, we report for each test the corresponding  $N$ , that is, the number of participants with valid response times in both of the two questions. The number of exclusions for each test is simply the difference between the reported number of observations and the total sample size of  $N = 158$ .

#### 4.1. Question-Level Analysis

In the following subsections we present the CRT questions used in the present study, together with the corresponding analyses of error rates and response times of the matched pairs. For each pair we briefly outline the rationale behind the conflict version and the construction of the non-conflict version. Given the frame modification that some of the original CRT questions underwent to minimize recognizability, we report the text (English translation of the German items) also for those original CRT questions.

##### 4.1.1. The Bat and the Ball: Conflict vs. Heuristic

The first pair of questions presented corresponded to the famous "bat and the ball" problem (Frederick, 2005). A non-conflict version of this question has been previously studied by De Neys et al. (2013) and Johnson et al. (2016).

**TABLE 1 | Summary statistics.**

Variable	Mean	Min.	Max.	Mean (females)	Mean (males)	WRT for gender effects
<b>NR. OF CORRECT ANSWERS</b>						
Total	16.475 (2.568)	10	22	16.149 (2.475)	17.053 (2.649)	$z = -2.35, p = 0.019^{**}$
Classic CRT	2.025 (1.009)	0	3	1.851 (1.043)	2.333 (0.873)	$z = -2.88, p = 0.004^{***}$
CRT2	1.38 (0.683)	0	2	1.366 (0.717)	1.404 (0.623)	$z = -0.08, p = 0.936$
Conjunction fallacy	3.076 (0.878)	1	4	3.05 (0.865)	3.123 (0.908)	$z = -0.63, p = 0.529$
Base-rate neglect	3.816 (0.943)	1	6	3.772 (0.937)	3.895 (0.958)	$z = -0.83, p = 0.409$
Ratio bias	1.892 (0.311)	1	2	1.871 (0.337)	1.93 (0.258)	$z = -1.14, p = 0.256$
Numeracy	9.652 (1.391)	5	11	9.386 (1.150)	10.122 (1.449)	$z = -3.38, p = 0.001^{***}$
<b>PERSONALITY TRAITS</b>						
Faith in intuition	6.354 (1.568)	1.94	10	6.313 (1.596)	6.427 (1.528)	$z = -0.44, p = 0.660$
Need for cognition	6.096 (1.372)	2.08	9.84	5.905 (1.462)	6.435 (1.131)	$z = -2.22, p = 0.027^{**}$
Openness to experience	5.764 (1.861)	1.333	9.467	5.808 (1.849)	5.685 (1.895)	$z = 0.45, p = 0.651$
Conscientiousness	6.391 (1.889)	0.333	10	6.359 (1.857)	6.449 (1.96)	$z = -0.13, p = 0.901$
Extraversion	6.289 (2.298)	0.433	10	6.549 (2.258)	5.829 (2.318)	$z = 2.00, p = 0.046^{**}$
Agreeableness	6.850 (1.729)	2.1	10	6.774 (1.718)	6.985 (1.755)	$z = -0.68, p = 0.500$
Neuroticism	5.547 (2.251)	0	10	5.942 (2.024)	4.847 (2.473)	$z = 2.87, p = 0.004^{***}$
Actively open-minded thinking	4.927 (4.783)	2.142	6	4.915 (4.318)	4.950 (5.55)	$z = -1.25, p = 0.211$
Known CRT items	1.753 (1.177)	0	3	1.653 (1.178)	1.93 (1.163)	$z = -1.49, p = 0.137$
Swiftness	94.546 (29.205)	36.941	169.449	95.178 (28.815)	93.426 (30.110)	$z = 0.48, p = 0.631$

Nr. of Correct Answers only refers to the conflict versions of the questions. The variable Classic CRT refers to the three original items of Frederick (2005). The variable CRT2 refers to the two extended-CRT conflict questions taken from Toplak et al. (2014). Numbers in parentheses are standard deviations.

**(Q1C)** A postcard and a pen cost 110 cents in total. The postcard costs 100 cents more than the pen. How much does the pen cost? (In cents)

[correct answer = 5]

**(Q1H)** A magazine and a banana together cost 290 cents. The magazine costs 200 cents. What is the price of the banana? (In cents)

[correct answer = 90]

For the classic (Q1C) question, there is an intuitive but wrong answer ("10"). This presumably involves participants focusing on the numbers, quickly segmenting the 110 cents into 100 and 10 cents, thereby neglecting the "more than" statement. Question

(Q1H) provides a control version of the problem, developed by De Neys et al. (2013). By eliminating the words "more than" from the question, it allows the intuitive segmentation mechanism to produce the correct answer. At the same time, however, the computation process that provides the correct solution in (Q1C) cannot be applied in this problem anymore. It becomes entirely inappropriate, since the solution is transparent. Hence this non-conflict version of the question (which, to the best of our knowledge, follows the obvious way to remove the conflict), neither generates process alignment nor shuts down the intuitive process. Rather, it corresponds to the *heuristic question* case we have described above.

**TABLE 2 | Spearman correlations among personality traits (and numeracy).**

	<b>1. Num</b>	<b>2. FI</b>	<b>3. NFC</b>	<b>4. Open</b>	<b>5. Consc</b>	<b>6. Extra</b>	<b>7. Agree</b>	<b>8. Neuro</b>
2. FI	0.067	–						
3. NFC	−0.107	0.140*	–					
4. Open	0.102	0.443***	0.190**	–				
5. Consc	0.047	0.163**	0.246**	0.034	–			
6. Extra	−0.038	0.261***	0.053	0.165**	0.154	–		
7. Agree	0.129	0.034	−0.030	0.130	0.090	0.043	–	
8. Neuro	−0.057	−0.072	−0.226**	−0.065	−0.065	−0.223***	0.033	–
9. AOT	0.049	0.062	0.266***	0.191**	0.074	−0.003	0.226***	−0.003

Num, Numeracy; FI, Faith in intuition; NFC, Need for cognition; Open, Openness to experience; Consc, Conscientiousness; Extra, Extraversion; Agree, Agreeableness; Neuro, Neuroticism; AOT, Actively open-minded thinking. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

**Figure 1** depicts the percentages of errors and correct responses (panel A) and the response times (panel B). Participants' answers to the conflict question were significantly slower than their answers to the heuristic question (median response time 29.14 s, mean 34.38 s,  $SD = 22.00$  in case of conflict; median 17.29 s, mean 18.71 s,  $SD = 6.42$  for the heuristic question; WSR test,  $N = 141$ ,  $z = 7.55$ ,  $p < 0.001$ ). This is consistent with hypothesis H3.

There were significantly more errors in the conflict question than in the heuristic version. For (Q1C), there were 39.72% (56) heuristic errors, 2.13% (3) non-heuristic errors (responses other than five or ten), and 58.16% (82) correct answers. For (Q1H), all answers were correct. Unsurprisingly, the proportion of errors in the conflict question was significantly larger than in the heuristic question [McNemar's test,  $N = 141$ ,  $\chi^2_{(1)} = 59.00$ ,  $p < 0.001$ ].

#### 4.1.2. Making Widgets: Conflict vs. Neutral

The second pair of questions again corresponds to one of the classic CRT items of Frederick (2005). The non-conflict version corresponds to our neutral category.

- (Q2C) If it takes 5 machines 5 minutes to make 5 car tires, how long would it take 100 machines to make 100 car tires? (In minutes)  
[correct answer = 5]
- (Q2N) If it takes 60 machines 100 minutes to make 60 bricks, how long would it take 100 machines to make 100 bricks? (In minutes)  
[correct answer = 100]

The number repetition in (Q2C) induces many participants to complete the pattern and give the intuitive but wrong answer “100.” (Q2N) provides a control version where the pattern is broken. By excluding the possibility of recognizing and reproducing a simple pattern, (Q2N) excludes the possibility of using a heuristic shortcut as in (Q2C). However, the same computation process that provides the correct solution in the conflict version can still be applied in this problem. Therefore, (Q2N) is a neutral counterpart of the conflict item (Q2C).

**Figure 1** depicts the percentages of errors and correct responses (panel A) and the response times (panel B). Answers to the conflict question were significantly faster than the answers to

the neutral question (median response time 21.61 s, mean 27.08 s,  $SD = 18.10$  in case of conflict; median 37.16 s, mean 48.69 s,  $SD = 33.07$  for the neutral question; WSR test,  $N = 139$ ,  $z = -7.68$ ,  $p < 0.001$ ). This is in agreement with our Hypothesis H2.

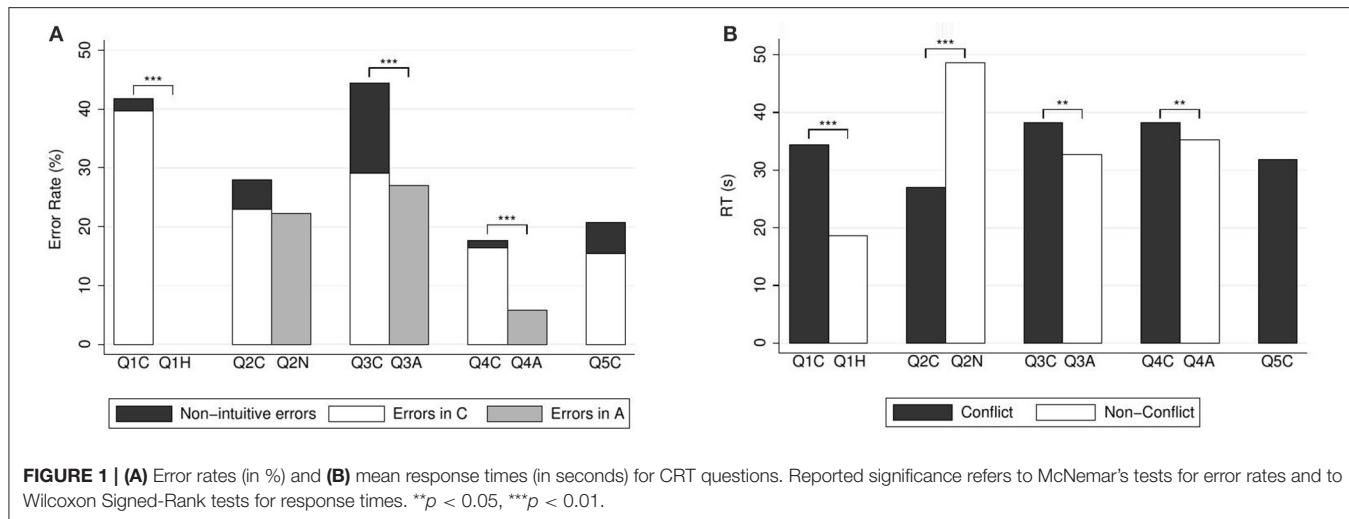
Regarding choice data, for (Q2C) there were 23.02% (32) heuristic errors, 5.04% (7) non-heuristic errors, and 71.94% (100) correct answers. For (Q2N), there were 22.30% (31) errors, and 77.70% (108) correct answers. According to McNemar's test, the proportion of errors in the conflict question was not significantly different than in the neutral question [ $N = 139$ ,  $\chi^2_{(1)} = 2.29$ ,  $p = 0.131$ ]. Please note, however, that throughout the paper we rely on two-sided tests. If we used a one-sided test here (based on our directional prediction), the result would of course be (marginally) significant.

#### 4.1.3. Buying and Selling: Conflict vs. Alignment

The third pair of questions we used was taken from the extended-CRT questions of Toplak et al. (2014), for which we developed an alignment version.

- (Q3C) A man buys a pig for 60 Euros, sells it for 70 Euros, buys it back for 80 Euros, and finally sells it for 90 Euros. How much has he made? (In Euros)  
[correct answer = 20]
- (Q3A) A man buys a cow for 60 Euros, sells it for 70 Euros, buys it back for 70 Euros, and finally sells it for 90 Euros. How much has he made? (In Euros)  
[correct answer = 30]

For the (Q3C) question, there is an intuitive but wrong answer (“10”). This is due to a miscalculation of the earnings where the difference between each two consecutive buying or selling actions is computed, instead of computing the profits or losses from every buy-and-sell operation. That is, participants compute  $(70 - 60) + (70 - 80) + (90 - 80) = 10$  instead of  $(70 - 60) + (90 - 80) = 20$ . In (Q3A), by having equal numbers in the middle of the question, this heuristic but incorrect way of thinking provides the correct answer. Importantly, the computation process that provides the correct solution is the same in (Q3C) and (Q3A). Therefore (Q3A) is an alignment counterpart of the conflict item (Q3C).



**Figure 1** depicts the percentages of errors and correct responses (panel A) and the response times (panel B). Answers to the conflict question were significantly slower than the answers to the alignment question (median response time 33.74 s, mean 38.23 s,  $SD = 18.40$  in case of conflict; 28.92 s, mean 32.66 s,  $SD = 15.03$  in case of alignment; WSR test,  $N = 144$ ,  $z = 2.42$ ,  $p = 0.015$ ). This is in agreement with our Hypothesis H1.

Alignment of course produces a simpler question, since the intuitive process becomes a cognitive shortcut. It was hence expected that there would be less errors under alignment. For (Q3C), there were 29.17% (42) heuristic errors, 15.27% (22) non-heuristic errors, and 55.56% (80) correct answers. For (Q3A), there were 27.08% (39) errors, and 72.92% (105) correct answers. According to McNemar's test, the proportion of errors in the conflict question was significantly larger than in the alignment question [ $N = 144$ ,  $\chi^2_{(1)} = 15.24$ ,  $p < 0.001$ ].

#### 4.1.4. Up and Down: Conflict vs. Alignment

The fourth pair of questions presented to participants is the seventh item in the list of extended CRT questions by Toplak et al. (2014), which follows a multiple-choice format. Our non-conflict version follows one developed by Bieleke and Gollwitzer for a different purpose (manuscript in preparation).

- (Q4C)** In January, Lucas invested 8000 Euros in the stock market. In July, the stocks he had purchased were down 50%. Fortunately, from July to October, they went up 75%. In October, Lucas

- has broken even in the stock market.
- is ahead of where he began.
- has lost money.

[correct answer = has lost money]

- (Q4A)** On Monday, the temperature was 22°C in the shade. Two days later, the temperature was down by 50%. Fortunately, the temperature went up 125% again by Saturday. Compared to Monday, the temperature on Saturday is
- warmer.

- the same.
  - colder.
- [correct answer = warmer]

In this problem, participants typically focus on the fact that the later percentage increase is larger than the earlier percentage decrease, neglecting that the amount to which the increase is applied is not the starting amount. Hence, many participants erroneously select the second option in (Q4C). In (Q4A), by making the percentage increase larger, the heuristic shortcut provides the correct answer even if the way of thinking is erroneous. Still, the correct answer can also be reached by means of the same computation mechanism that is required to correctly answer (Q4C). Therefore, (Q4A) is an alignment counterpart of the conflict item (Q4C).

**Figure 1** depicts the percentages of errors and correct responses (panel A) and the response times (panel B). Answers to the conflict question were significantly slower than the answers to the alignment question (median response time 35.95 s, mean 38.32 s,  $SD = 15.06$  in case of conflict; median 33.24 s, mean 35.25 s,  $SD = 13.50$  in case of alignment; WSR test,  $N = 152$ ,  $z = 2.25$ ,  $p = 0.024$ ). Again, this is in agreement with our Hypothesis H1.

As in the previous pair, there were significantly less errors under alignment. For (Q4C), there were 16.45% (25) heuristic errors, 1.32% (2) non-heuristic errors, and 82.24% (125) correct answers. For (Q4A), there were 5.92% (9) errors, and 94.08% (143) correct answers. According to McNemar's test, the proportion of errors in the conflict question was significantly larger than in the alignment question [ $N = 152$ ,  $\chi^2_{(1)} = 9.00$ ,  $p = 0.003$ ].

#### 4.1.5. Growing in the Lake

In order to be able to compute the standard CRT score, we also included the last of the classic three CRT items of Frederick (2005).

- (Q5C)** In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover

the entire lake, how long would it take for the patch to cover half of the lake? (In days)  
[correct answer = 47]

For this question, there is an intuitive but wrong answer ("24"), produced by halving the number of days, ignoring the exponential growth of the lily pads. The structure of the question makes it impossible to create a non-conflict version without making it exceedingly trivial. Hence, this item was not paired with a non-conflict version.

After removing response-time outliers, our sample for the (Q5C) question contains  $N = 149$  observations. The median response time was 27.33 s, mean 31.85 s,  $SD = 16.79$  (Figure 1B). There were 15.44% (23) heuristic errors, 5.37% (8) non-heuristic errors, and 79.19% (118) correct answers (Figure 1A).

## 4.2. Regression Analysis for Extended CRT Questions

Our data forms a perfectly balanced panel with 9 decisions per participant. Hence we rely on random-effects panel regressions. This allows us to control for a variety of variables that might affect choices or response times, such as the number of words and letters and participants' swiftness.

**Table 3** displays the results of panel regressions for response times. Contrary to the individual tests, we did not drop participants with outlier response times. Dropping those participants would have greatly reduced the sample since the regression covers all questions simultaneously. Instead, we relied on random effects and a log-transformation of response times (and controlling for swiftness). We only had to drop one of the participants from the whole sample because he left one of the answers blank. Model 1 contains dummies for the different versions of questions (heuristic, neutral, and alignment; conflict is the reference category). All dummies are significant, implying longer response times for neutral questions and shorter response times for heuristic and alignment questions, in agreement with Hypotheses H1, H2, and H3. Further, they remain significant when controlling for interindividual heterogeneity (Model 2). Not surprisingly, previous knowledge of the classic CRT items reduces response times.

In Model 3, we introduce dummies for intuitive and non-intuitive errors. The results imply that participants making non-intuitive errors are slower. A *post-hoc* test further shows that participants making an intuitive error under conflict are significantly slower than those giving the correct answer under conflict (coefficient 0.167,  $SD = 0.060$ ,  $z = 2.84$ ,  $p = 0.004$ ). However, there was no difference between those committing an error and those giving a correct answer under alignment (*post-hoc* test, coefficient 0.004,  $SD = 0.044$ ,  $z = 0.10$ ,  $p = 0.923$ ). In any case, this should not be confused with a statement on the relative speed of errors, which would be a within-subject comparison. Since this is a comparison across subjects, it merely points out that participants giving incorrect answers might be cognitively slower than participants giving correct answers.

We now turn to random-effects probit panel regressions on correct answers to the CRT questions (Table 4). As expected, the likelihood of a correct answer is higher in the absence of

**TABLE 3 | Random-effects regressions on log response times of CRT questions.**

Log(response time)	Model 1	Model 2	Model 3
Heuristic	-0.438*** (0.075)	-0.438*** (0.075)	-0.356*** (0.073)
Neutral	0.443*** (0.158)	0.443*** (0.159)	0.482** (0.157)
Alignment	-0.063* (0.037)	-0.063* (0.038)	-0.002 (0.035)
Number of words	0.006 (0.011)	0.006 (0.011)	0.006 (0.010)
Number of letters	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Log(swiftness)		0.101* (0.053)	0.103* (0.061)
Numeracy		-0.040 (0.033)	-0.026 (0.030)
Male		0.057 (0.058)	0.053 (0.053)
Known CRT items		-0.114*** (0.036)	-0.109*** (0.037)
Faith in intuition		-0.006 (0.016)	-0.007 (0.016)
Need for cognition		-0.020 (0.018)	-0.018 (0.020)
Openness to experience		-0.005 (0.016)	-0.006 (0.016)
Conscientiousness		-0.022*** (0.006)	-0.021*** (0.007)
Extraversion		0.022*** (0.003)	0.020*** (0.003)
Agreeableness		0.014 (0.010)	0.017 (0.012)
Neuroticism		0.009 (0.010)	0.006 (0.010)
Actively open-minded thinking		0.022 (0.028)	0.020 (0.033)
Non-intuitive error			0.458*** (0.027)
Intuitive error			0.167*** (0.059)
Intuitive error $\times$ alignment			-0.163** (0.073)
Observations	1413	1413	1413

The dummies Heuristic, Neutral, and Alignment take the value 1 for the respective versions of the questions; Conflict is the reference category. Standard errors (in parentheses) are (conservatively) clustered at the level of counterbalance condition (question order).

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

conflict, as reflected by a dummy pooling alignment, neutral, and heuristic questions. Participants scoring high in numeracy are more likely to answer the CRT questions correctly, in spite of the low variance in this scale. There is also a gender effect, with males providing correct responses more often. However,

**TABLE 4 | Random-effects probit regressions on correct answers to CRT questions.**

Correct answer	Model 1	Model 2	Model 3	Model 4
Male	0.259** (0.123)	0.119 (0.144)	0.118 (0.144)	0.026 (0.130)
Non-conflict	0.632*** (0.027)	0.629*** (0.027)	0.931** (0.373)	0.923** (0.377)
Faith in intuition (FI)		-0.007 (0.016)	-0.018 (0.017)	0.006 (0.019)
Need for cognition (NFC)		0.010 (0.033)	0.039 (0.041)	0.026 (0.036)
Numeracy		0.186*** (0.034)	0.185*** (0.033)	0.187*** (0.033)
Non-conflict × FI			0.029 (0.026)	0.031 (0.027)
Non-conflict × NFC			-0.080 (0.063)	-0.081 (0.063)
Openness to experience				-0.051* (0.030)
Conscientiousness				0.030 (0.035)
Extraversion				-0.022 (0.023)
Agreeableness				0.071*** (0.022)
Neuroticism				-0.053*** (0.012)
Actively open-minded thinking				-0.009 (0.048)
Observations	1413	1413	1413	1413

The dummy Non-Conflict subsumes alignment, neutral, and heuristic problems; Conflict is the reference category. Standard errors (in parentheses) are (conservatively) clustered at the level of counterbalance condition (question order).

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

once FI, NFC, and numerical literacy are included (Model 2), the gender difference disappears. Effects remain significant when controlling for further heterogeneity, including the Big Five and AOT (Model 4).

### 4.3. Discussion: Extended CRT Questions

Response times are typically in the 15 – 35 s range, which is considerably longer than response times studied in the dual-process literature. Hence, it is clear that practically all decisions involve deliberation and no relevant part of the observations can be viewed exclusively as the result of a fast, automatic (intuitive or heuristic) process in the sense of the dual-process literature (Epstein, 1994; Strack and Deutsch, 2004). However, a mechanistic interpretation of process conflict and alignment, as given in Section 2.4, might still help organize and understand the data.

The predictions derived from this interpretation (Section 2.4) were clearly supported by the data. Overall, the study of response times related to CRT questions suggests that even at this long time

scale, this kind of questions fall well within the domain of dual-process theories. It is conceptually useful to identify behavioral tendencies with decision processes and consider intuitive ones as more automatic (hence faster) processes.

When analyzing response times, a large individual heterogeneity has to be expected, and differences will become more important at longer time scales. The regression analysis confirmed our basic findings while controlling for individual differences, including a number of personality factors, and an individual measure of swiftness (which was, as expected, significant). To study actual choices, we also conducted probit regressions on individual answers, which showed the expected effects, e.g., non-conflict question versions were easier.

The regressions also allowed us to examine the influence of personality factors on behavior. Interestingly, scales as FI, NFC, and AOT had no impact on CRT questions (neither on answers nor on response times), in agreement with previous evidence that these measures appear to diverge greatly at the individual level (Alós-Ferrer and Hügelschäfer, 2016). However, the inclusion of these measures eliminates apparent gender effects, pointing out that gender differences in performance in CRT-style questions might be explained by personality differences correlated with gender. Regarding the Big-Five Inventory, Conscientiousness led to faster responses and Extraversion to slower ones, but neither had a significant effect on responses. Agreeableness led to significantly more correct answers and Neuroticism and Openness to Experience to more errors, but none of them had a significant effect on response times.

Finally, it is important to note that the items we have considered (as the ones related to decision biases analyzed in the next section) belong to the category of *decisions from inference*, in the sense that there is an objectively correct answer which needs to be identified. This is in contrast to *preferential choice*, where by definition there is no objectively correct response (for example, consider lottery-choice questions). For decisions from inference, it is in principle possible to derive natural hypotheses on the nature of the involved processes in advance, as our discussions above illustrate. For preferential choice, the picture is less clear, because the very nature of the involved processes is part of the research question. We will return to this point in the discussion.

## 5. DECISION BIASES

In the following subsections we present the questions capturing the decision biases investigated in the present study, together with the corresponding analyses of error rates and response times of the matched items. For each question we report the English translation of the German items we used, and briefly outline the rationale behind the conflict version and the construction of the non-conflict version. The same criteria for outliers and tests were used as in Section 4.

To explore the influence of personality factors, we also test whether participants' proneness to decision biases is related to their CRT, FI, and NFC scores, following Alós-Ferrer and Hügelschäfer (2012, 2016). In particular, Alós-Ferrer and Hügelschäfer (2016) observed that higher CRT scores were linked

**TABLE 5 | Random-effects probit regressions on correct answers to decision-bias questions.**

Correct answer	BRN1	BRN2	BRN3	Conj. fall.	Ratio bias
Classic CRT	0.535 (0.764)	-0.025 (0.109)	0.320** (0.144)	0.379*** (0.113)	0.513*** (0.161)
CRT2	0.708 (1.112)	0.211 (0.165)	0.401* (0.216)	-0.113 (0.169)	-0.138 (0.237)
Faith in intuition	0.198 (0.508)	0.133* (0.069)	0.036 (0.085)	0.070 (0.070)	0.088 (0.102)
Need for cognition	0.246 (0.543)	-0.035 (0.083)	0.102 (0.102)	-0.054 (0.081)	-0.144 (0.117)
Observations	158	158	158	158	158

Correct Answer refers to the conflict version of the respective questions. BRN1-3 indicate the three base-rate-neglect problems. Standard errors (in parentheses) are (conservatively) clustered at the level of counterbalance condition (question order).

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

to a lower likelihood of committing the conjunction fallacy and base-rate neglect, in line with previous research (Oechssler et al., 2009; Hoppe and Kusterer, 2011). Similarly, lower FI scores were associated with a lower likelihood of these biases, albeit not as consistently as CRT scores. Whereas those results were based on median splits, we ran random-effects probit regressions in order to take advantage of the full range of scores, regressing correct answers to the conflict versions of the bias questions on participants' CRT, FI, and NFC score. We defined the additional variable CRT2 as the score in the two additional conflict items taken from Toplak et al. (2014) (hence, CRT2 can take the values 0, 1, or 2). Results are shown in **Table 5** and are discussed in the respective subsections below.

## 5.1. Base-Rate Neglect

The first group of questions on decision biases refers to base-rate neglect. This phenomenon occurs when decision makers overweight sample information at the expense of the base rate. To examine this bias, we used three pairs of questions.

### 5.1.1. Taxicabs and Base-Rate Neglect: Conflict vs. Alignment

The first question is the celebrated “Taxicab question” from Kahneman and Tversky (1972), studied by Tversky and Kahneman (1980) and Bar-Hillel (1980), which we implemented as a multiple-choice problem.

**(BR1C)** In a city there are two cab companies, the Green and the Blue. 85% of the cabs in the city are Green and 15% are Blue. A cab was involved in a hit-and-run accident last night. A witness identified the cab as a Blue cab. The court tested his ability to distinguish between Green and Blue cabs at night. The witness made correct identifications in 80% of the cases and erred in 20% of the cases. The probability that the cab involved in the accident was Blue rather than Green is

- larger than 50%.
- smaller than 50%.

**(BR1A)** In a city there are two limousine companies, the Yellow and the Pink. 60% of the limousines in the city are Yellow and 40% are Pink. A limousine was involved in a hit-and-run accident last night. A witness identified the limousine as a Yellow. The court tested his ability to distinguish between Yellow and Pink limousines at night. The witness made correct identifications in 70% of the cases and erred in 30% of the cases. The probability that the limousine involved in the accident was Yellow rather than Pink is

- larger than 50%.
- smaller than 50%.

Bayes' Rule yields a posterior probability of  $\sim 41\%$  in the conflict version (BR1C), and a probability of  $\sim 78\%$  in the alignment version (BR1A). However, in studies involving probability estimates, median answers in (BR1C) are typically around 80% (e.g., Bar-Hillel, 1980). This is because decision makers typically underweight the base rate, and their answers are dominated by the witness' credibility instead. Hence, for (BR1C), the intuitive but normatively wrong answer is to choose the first option (larger than 50%). In (BR1A), by increasing the base rate, the same heuristic that misled participants in (BR1C) now provides the correct answer. Therefore (BR1A) is an alignment counterpart of the conflict item (BR1C). We remark that Bar-Hillel (1980) developed a different and more extreme non-conflict version, but we developed our own because in that version, the statement of a witness is replaced by more specific information which actually dominates the base rate, so that neglecting the base rate is appropriate.

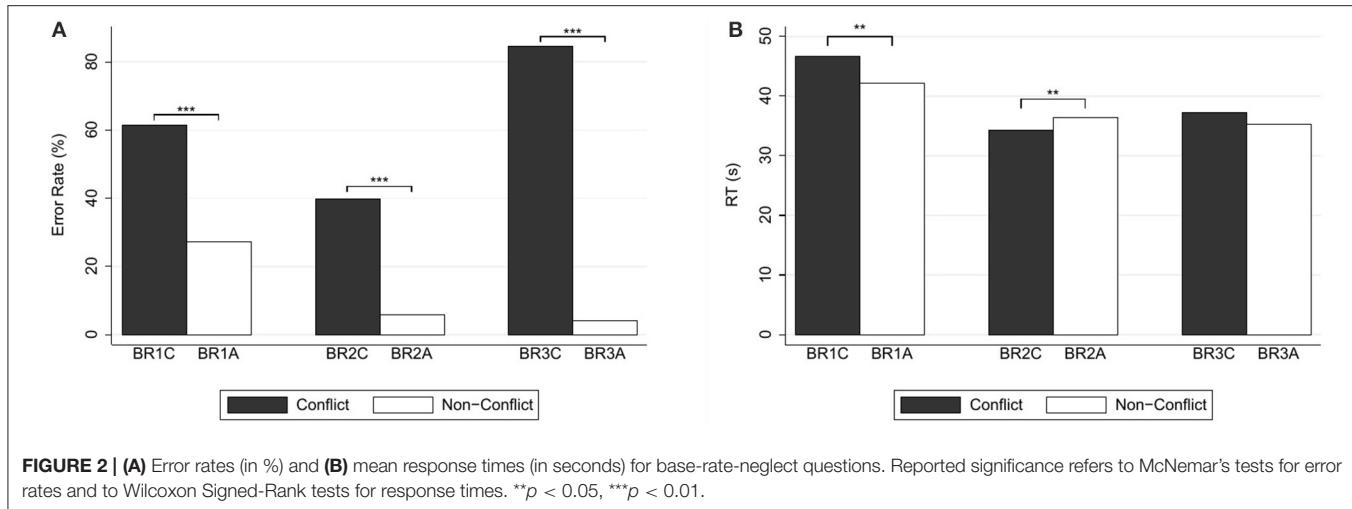
**Figure 2** depicts the percentages of errors and correct responses (panel A) and the response times (panel B). Answers to the conflict question were significantly slower than the answers to the alignment question (median response time 43.73 s, mean 46.78 s,  $SD = 19.13$  in case of conflict; 38.70 s, mean 42.23 s,  $SD = 14.43$  in case of alignment; WSR test,  $N = 146$ ,  $z = 2.01$ ,  $p = 0.044$ ), in agreement with our basic Hypothesis H1.

Regarding error rates, for the conflict question (BR1C), there were 61.64% (90) errors, and 38.36% (56) correct answers. For (BR1A), there were 27.40% (40) errors, and 72.60% (106) correct answers. As is to be expected for a comparison between a conflict and an alignment version, the proportion of errors in the conflict question was significantly larger than in the alignment question [McNemar's test,  $N = 146$ ,  $\chi^2_{(1)} = 27.17$ ,  $p < 0.001$ ].

Last, we report on the relation to the CRT and the FI and NFC scales. Alós-Ferrer and Hügelshäfer (2016) found that the CRT had no informative value for the base-rate fallacy as captured by this particular question. We obtain the same null result. There were no effects of CRT score, FI or NFC on the likelihood of correctly answering the conflict version of this question (BR1C) (see **Table 5**).

### 5.1.2. Detecting Criminals and Base-Rate Neglect: Conflict vs. Alignment

The next pair of questions used to measure base-rate neglect is analogous to a classic problem from Eddy (1982) and has



**FIGURE 2 | (A)** Error rates (in %) and **(B)** mean response times (in seconds) for base-rate-neglect questions. Reported significance refers to McNemar's tests for error rates and to Wilcoxon Signed-Rank tests for response times. \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

been used by Hoppe and Kusterer (2011) and Alós-Ferrer and Hügelschäfer (2016).

- (BR2C)** In a city with 10 criminals and 10,000 innocent citizens there is a surveillance camera with an automatic face recognition software. If the camera sees a known criminal, it will trigger the alarm with 70% probability; if the camera sees an innocent citizen, it will trigger the alarm with a probability of 30%. The probability that indeed a criminal was filmed when the alarm is triggered is
  - larger than 50%.
  - smaller than 50%.**(BR2A)** In a prison with 9000 criminals and 90 innocent people appeals to court take place regularly. When an appeal happens, it will sentence guilty a criminal with 75% probability; if the trial involves an innocent citizen, it will sentence him guilty with a probability of 25%. The probability that indeed a criminal was sentenced if the appeal judged him guilty is
  - larger than 50%.
  - smaller than 50%.

The posterior probability in the (BR2C) question is only  $\sim 23\%$ , but due to base-rate neglect participants typically overweight the reliability of the test. Hence, the intuitive but incorrect response is to select the first option. In (BR2A), by increasing the base rate to put it in agreement with the diagnostic information, the same heuristic that misled participants in (BR2C) now provides the correct answer. Therefore (BR2A) is an alignment counterpart of the conflict item (BR2C).

**Figure 2** depicts the percentages of errors and correct responses (panel A) and the response times (panel B). Answers to the conflict question were significantly *faster* than the answers to the alignment version (median response time 30.63 s, mean 34.38 s,  $SD = 15.95$  in case of conflict; 34.80 s, mean 36.45 s,  $SD = 11.77$  in case of alignment; WSR test,  $N = 138$ ,  $z = 2.08$ ,  $p = 0.037$ ). This is inconsistent with the results for the previous question pair and with our Hypothesis H1.

Error rates, however, do not suggest a qualitative difference with the previous pair. For (BR2C) there were 39.86% (55) errors, and 60.14% (83) correct answers. For (BR2A) there were 5.80% (8) errors, and 94.20% (130) correct answers. According to McNemar's test, the proportion of errors in the conflict question was significantly larger than in the alignment question [ $N = 138$ ,  $\chi^2_{(1)} = 37.44$ ,  $p < 0.001$ ].

As shown in **Table 5**, participants' CRT score did not affect the likelihood of a correct answer to the (BR2C) item, in contradiction with the results by Hoppe and Kusterer (2011) and Alós-Ferrer and Hügelschäfer (2016). In the same way, NFC was not predictive for this item. FI level was predictive, but not in the expected direction (marginally significantly higher likelihood of answering correctly with higher FI level).

### 5.1.3. Genetic Disorders and Base-Rate Neglect: Conflict vs. Alignment

The third pair of items is based on the original question of Eddy (1982).

- (BR3C)** Jonathan has been tested for a rare genetic disorder at his doctor. Only one in 10,000 people have this disorder. The test has very high detection rate: 99%. That means if Jonathan has the disorder, there is a 99% chance that the test is positive. The test also has a very low false-positive rate: 1%. That means that if Jonathan does not have the disorder, there is only a 1% chance that the test is positive. Unfortunately, Jonathan has tested positive for this disorder. The probability with which Jonathan has the genetic disorder is
  - larger than 50%.
  - smaller than 50%.**(BR3A)** A doctor has tested a population of obese patients for high cholesterol in the blood. 8000 in 10,000 obese people have high cholesterol. The test has a very high detection rate: 95%. That means if a patient has high cholesterol, there is a 95% chance that the test is positive. The test also has a very low false-positive rate:

5%. That means that if a patient does not have high cholesterol, there is only a 5% chance that the test is positive. A patient has been tested positive for this condition. The probability with which the patient has high cholesterol is

- larger than 50%.
- smaller than 50%.

The logic here is the same as for (BR2C) and (BR2A). The posterior probability in the (BR3C) question is only around ~1%, but participants are tempted to select the first option. Due to the altered base rate, (BR3A) is an alignment counterpart of the conflict item (BR3C).

**Figure 2** depicts the percentages of errors and correct responses (panel A) and the response times (panel B). Response times of answers to the conflict question were not significantly different from those to the alignment question (median response time 32.32 s, mean 37.22 s,  $SD = 18.47$  in case of conflict; median 31.57 s, mean 35.38 s,  $SD = 14.77$  in case of alignment; WSR test,  $N = 144$ ,  $z = 0.62$ ,  $p = 0.536$ ). Hence, with respect to our Hypothesis H1, we cannot reject the null hypothesis of no differences in this case.

As in the case of the previous pair of questions, however, choice data reflect the normal difference between a conflict and an alignment version. For (BR3C), there were 84.97% (122) errors, and 15.28% (22) correct answers. For (BR3A), there were 4.17% (6) errors, and 95.83% (138) correct answers. According to McNemar's test, the proportion of errors in the conflict question was significantly larger than in the alignment question [ $N = 144$ ,  $\chi^2_{(1)} = 114.03$ ,  $p < 0.001$ ].

A higher CRT score significantly increased the likelihood of giving a correct answer to (BR3C) (see **Table 5**). The effect was significant when considering the classic CRT, and marginally significant when considering the two items from the extended CRT contained in CRT2. There was no effect of FI or NFC scores.

#### 5.1.4. Regression Analysis (Base-Rate Neglect)

Our data forms a perfectly balanced panel with 6 decisions per participant. **Table 6** reports random-effects panel regressions on response times, transformed logarithmically. Answers to the alignment versions are significantly slower, also when controlling for personality traits (Model 2). Swiftness is again predictive of the time participants need to work on the base-rate-neglect questions. There is a significant gender effect, suggesting that females are faster in answering the problems.

Regarding choice data, we ran random-effects probit panel regressions on correct answers to the base-rate-neglect items (**Table 7**). The variable conflict is significant across all models, indicating an increased likelihood of giving a correct answer to alignment compared to conflict questions. The score obtained in the classic CRT does not affect correct answers across all base-rate-neglect questions (Model 2). Controlling for conflict (Model 3), we obtain the unexpected result that there is a significant negative effect of CRT score on the likelihood of correct answers to alignment questions. Further, the CRT score does not significantly predict a correct answer to the conflict versions (Model 3; *post-hoc* test of the linear combination of CRT2 plus Conflict  $\times$  CRT2: coefficient 0.135,  $SD = 0.059$ ,  $z = 2.30$ ,  $p = 0.022$ ). Results remain stable when including personality traits (Model 4).

**TABLE 6 | Random-effects regressions on log response times of base-rate-neglect questions.**

Log(response time)	Model 1	Model 2	Model 3
Alignment	0.175*** (0.049)	0.175*** (0.049)	0.126*** (0.027)
Number of words	0.026*** (0.002)	0.026*** (0.002)	0.025*** (0.002)
Number of letters	-0.005*** (0.001)	-0.005*** (0.001)	-0.005*** (0.001)
Swiftness		0.174*** (0.067)	0.170*** (0.064)
Numeracy		0.041** (0.020)	0.042** (0.020)
Male		0.156*** (0.050)	0.152*** (0.043)
Faith in intuition		0.004 (0.019)	0.004 (0.019)
Need for cognition		-0.015 (0.025)	-0.014 (0.026)
Openness to experience		0.012* (0.007)	0.011 (0.008)
Conscientiousness		-0.004 (0.005)	-0.004 (0.005)
Extraversion		0.009*** (0.001)	0.009*** (0.001)
Agreeableness		-0.008 (0.029)	-0.008 (0.029)
Neuroticism		0.014 (0.016)	0.015 (0.017)
Actively open-minded thinking		0.010 (0.050)	0.010 (0.049)
Error			-0.053 (0.090)
Error $\times$ alignment			0.114 (0.145)
Observations	942	942	942

The dummy Alignment takes the value 1 for the respective versions of the questions; Conflict is the reference category. Standard errors (in parentheses) are (conservatively) clustered at the level of counterbalance condition (question order).

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Classic CRT plus Conflict  $\times$  Classic CRT: coefficient 0.010,  $SD = 0.057$ ,  $z = 0.17$ ,  $p = 0.861$ ). In contrast, the score obtained in the two items of CRT2 is a significant positive predictor for correctly answering to the alignment versions of the questions, and also for the conflict versions (Model 3; *post-hoc* test of the linear combination of CRT2 plus Conflict  $\times$  CRT2: coefficient 0.135,  $SD = 0.059$ ,  $z = 2.30$ ,  $p = 0.022$ ). Results remain stable when including personality traits (Model 4).

#### 5.1.5. Discussion (Base-Rate Neglect)

Response times for questions focusing on base-rate neglect were clearly longer than for the typical CRT questions, with

**TABLE 7 | Random-effects probit regressions on correct answers to base-rate-neglect questions.**

Correct answer	Model 1	Model 2	Model 3	Model 4
Male	0.032 (0.134)	0.016 (0.129)	0.015 (0.130)	0.014 (0.153)
Conflict	-1.377*** (0.133)	-1.391*** (0.174)	-1.639*** (0.518)	-1.662** (0.648)
Classic CRT		-0.037 (0.043)	-0.116*** (0.041)	-0.130* (0.072)
CRT2		0.150*** (0.038)	0.174*** (0.054)	0.209*** (0.070)
Faith in intuition		0.041 (0.046)	0.050 (0.059)	0.032 (0.058)
Need for cognition		0.008 (0.026)	-0.005 (0.079)	-0.017 (0.061)
Numeracy		0.032 (0.036)	0.033 (0.037)	0.030 (0.034)
Conflict × classic CRT			0.126** (0.052)	0.132* (0.071)
Conflict × CRT2			-0.039 (0.087)	-0.045 (0.091)
Conflict × FI			-0.014 (0.030)	-0.017 (0.030)
Conflict × NFC			0.021 (0.087)	0.025 (0.087)
Openness to experience				0.071** (0.036)
Conscientiousness				-0.026 (0.018)
Extraversion				-0.034** (0.014)
Agreeableness				0.017 (0.028)
Neuroticism				0.002 (0.017)
Actively open-minded thinking				0.052 (0.066)
Observations	942	942	942	942

The dummy Conflict takes the value 1 for the respective versions of the questions; Alignment is the reference category. Standard errors (in parentheses) are (conservatively) clustered at the level of counterbalance condition (question order).

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

medians above 30 seconds. Such long response times suggest that significant deliberation was involved. The behavioral results (error rates) indicate that in all three pairs, the constructed alignment version was easier than the conflict version. Hence we are confident that the constructed pairs worked as intended. However, the evidence on response times is mixed. For one of the pairs, responses to the alignment question were significantly faster than responses to the conflict version, for another the relation was the opposite, and for the third no significant differences were found. Pooling all the data, a panel regression controlling for swiftness and numeracy indicated a significantly

positive effect of alignment on response times. This provides evidence of a tendency for longer response times in the *easier* decisions, which stands in direct contradiction with both Hypothesis H1 and the idea that easier decisions are made faster (Dashiell, 1937; Mosteller and Nogee, 1951).

In Alós-Ferrer and Hügelschäfer (2016), it was already found that responses to different conflict questions used to measure base-rate neglect were affected differently by personality factors. From a conservative point of view, the only conclusion that can be drawn at this point is that, in spite of their apparent similarities at the abstract level, the heavily-framed, context-rich questions might activate quite different processes and process combinations. To fully understand base-rate neglect, and in particular its roots in different decision processes, future research should concentrate on separating framing effects and process conflict or alignment, moving away from the standard questions used in the literature.

The regression analysis allows us to examine the effect of personality differences on both choice data and response times. Scales as FI, NFC, and AOT had no effect in our sample when aggregating across questions. Higher scores in the classic CRT had no effect on error likelihood for the conflict versions of the questions, and surprisingly even increased errors for the alignment versions. In contrast, higher scores in the two items of the extended CRT reduced errors both for conflict and alignment questions. Regarding the Big Five Inventory, Extraversion resulted in longer response times and more errors, and Openness to Experience significantly reduced errors and increased response times.

## 5.2. Conjunction Fallacy

### 5.2.1. Question Analysis (Conjunction Fallacy)

The following four questions refer to the conjunction fallacy. To examine this bias, we employed problems analogous to the classic LINDA question from Tversky and Kahneman (1983).

- (CFC) Tom is 34 years old. He is intelligent, punctual but unimaginative, and somewhat lifeless. In school, he was strong in mathematics but weak in social studies and humanities. Which of the following statements is more likely to be true?

- Tom plays in a rock band for a hobby.
- Tom plays in a rock band for a hobby and is an accountant.

- (CFA) Klaus is 41 years old, single, introverted, and very intelligent. He majored in physics. As a student, he played pen-and-paper role-playing games, and also participated in several chess tournaments. Which of the following statements is more likely to be true?

- Klaus DJs on the weekend.
- Klaus DJs on the weekend or is a university professor.

- (CFN1) Claire is 30 years old, single, open-minded, and very smart. As a student of literature, she was deeply concerned with issues of discrimination and social justice, and also participated in several demonstrations.

- Which of the following statements is more likely to be true?
- Claire is active in the animal-rights movement.
  - Claire is active in the animal-rights movement and works in an international company.
- (CFN2) Richard is 31 years old, married with no children. A man of high ability and high motivation, he promises to be successful in his field. He is well liked by his colleagues. Which of the following statements is more likely to be true?
- Richard is an engineer.
  - Richard is an engineer and is active in the civil rights movement.

The (CFC) item, which is adapted from Tversky and Kahneman (1983) (see also De Neys and Bonnefon, 2013), is analogous to the LINDA problem. Intuition prescribes to select the second option, because the frame seems in line with the stereotype of an accountant more than with that of a rock-band member. This is obviously incorrect, because the simultaneous realization of two disjoint events cannot be more probable than one of the events.

(CFA), (CFN1), and (CFN2) represent different non-conflict versions of the same problem. First, by substituting “and” with “or” in (CFA), the stereotypical answer suggested by the frame becomes logically valid. The change does not affect the mechanism used to correctly answer to the problem, which is still the same as in (CFC). Therefore (CFA) is an alignment counterpart of the conflict item (CFC). Second, the frame of (CFN1) is adapted from the original LINDA problem (Tversky and Kahneman, 1983). By presenting the cue linked to the stereotype (“animal-rights movement”) in both answers, the heuristic which misleads participants in (CFC) cannot directly be applied because it does not have a favored option. Therefore (CFN1) is a neutral counterpart of the conflict item (CFC). Third, (CFN2) is adapted from Kahneman and Tversky (1973). The description of Richard is neutral with respect to the two suggested answers; hence the heuristic process activated in (CFC) is no longer available. Therefore (CFN2) represents another possible neutral counterpart of (CFC).

**Figure 3** depicts the percentages of errors and correct responses (panel A) and the response times (panel B). We compared the response times in the conflict question to those of the non-conflict variants, but we found no significant differences whatsoever, neither for the alignment question (CFA) (median 24.02 s, mean 25.33 s,  $SD = 7.79$ , compared to median 23.34 s, mean 24.63 s,  $SD = 9.90$  for (CFC); WSR test,  $N = 143$ ,  $z = -1.58$ ,  $p = 0.113$ ) nor for the neutral questions (CFN1) (median 23.61 s, mean 25.32 s,  $SD = 9.32$ , compared to median 23.24 s, mean 24.67 s,  $SD = 9.99$  for (CFC); WSR test,  $N = 144$ ,  $z = -0.64$ ,  $p = 0.524$ ) and (CFN2) (median 22.64 s, mean 23.52 s,  $SD = 8.67$  for (CFN2), compared to median 23.14 s, mean 24.58 s,  $SD = 9.93$  for (CFC); WSR test,  $N = 145$ ,  $z = 0.78$ ,  $p = 0.437$ ).

We also compared the percentages of errors in the non-conflict questions to those of the conflict question. The proportion of errors in the conflict question was significantly larger than in the alignment question (CFA) [9.09% (13)

compared to 42.66% (61) for (CFC); McNemar’s test:  $N = 143$ ,  $\chi^2_{(1)} = 32.91$ ,  $p < 0.001$ ] and in the neutral questions (CFN1) [27.78% (40), compared to 41.67% (60) for (CFC); McNemar’s test:  $N = 144$ ,  $\chi^2_{(1)} = 7.41$ ,  $p = 0.007$ ] and (CFN2) [13.79% (20), compared to 42.07% (61) for (CFC); McNemar’s test:  $N = 145$ ,  $\chi^2_{(1)} = 32.96$ ,  $p < 0.001$ ].

As can be seen from **Table 5**, the likelihood of a correct answer to the standard conjunction-fallacy problem (CFC) was significantly increased with increasing CRT score when considering the classic 3-item version, but not when considering only the two additional items from Toplak et al. (2014). This result is in line with the findings of Oechssler et al. (2009) and Alós-Ferrer and Hügelschäfer (2016), and also with Liberali et al. (2012), who found a negative correlation between CRT score and number of conjunction fallacies. In contrast, there were no effects of FI and NFC.

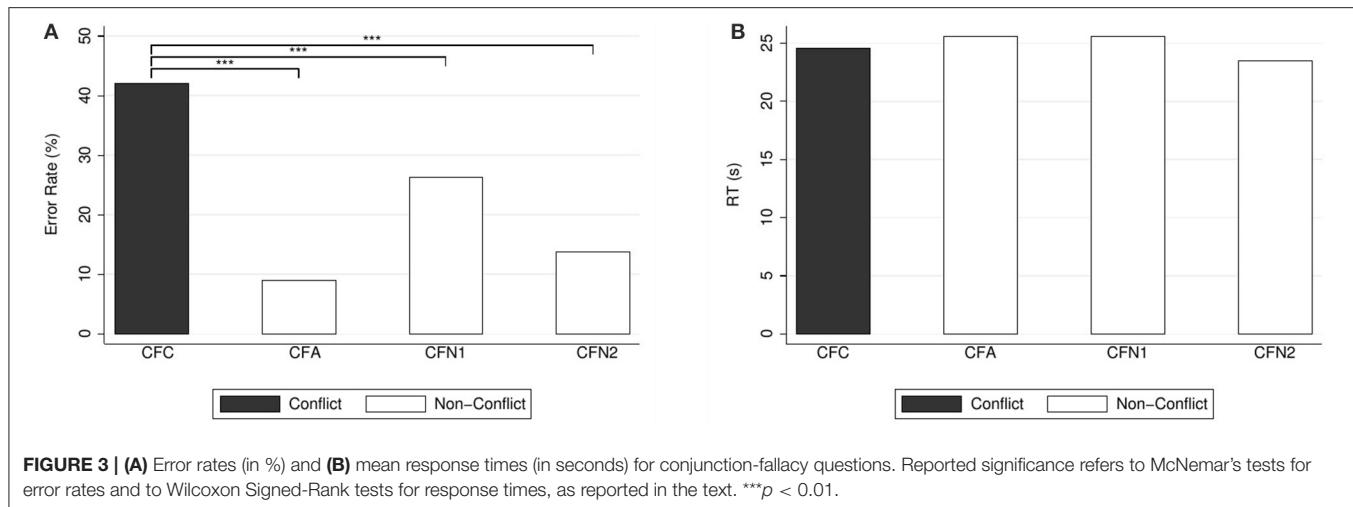
### 5.2.2. Regression Analysis (Conjunction Fallacy)

Our data forms a perfectly balanced panel with 4 decisions per participant. **Table 8** reports random-effects panel regressions on response times, transformed logarithmically. The alignment dummy is significantly positive in all three models. The effect of swiftness is as in previous sections. A high score in Need for Cognition is negatively related to response time. Further, the error dummy is significant and positive, meaning that participants making an error in the conflict and neutral versions of the question need more time than participants giving a correct answer. Again, this is a strictly between-subjects comparison which might simply reflect cognitive-capacity correlates.

To analyze actual choices, we ran random-effects probit panel regressions on correct answers (**Table 9**). In the basic model, the dummies conflict and neutral are significant and negative, indicating a lower probability of being answered correctly compared to the alignment counterpart. Scoring high in the numeracy scale is associated with an increased probability of giving a correct answer to the questions. Higher scores in the classic CRT are a significant positive predictor for correct answers, in particular for the conflict item (Model 3). This is in agreement with Liberali et al. (2012), who reported a significant negative association between CRT score and committing the conjunction fallacy. In contrast, the number of correct answers to the two items of CRT2 is not predictive (Model 3; *post-hoc* test of the linear combination of CRT2 plus Conflict  $\times$  CRT2: coefficient  $-0.136$ ,  $SD = 0.130$ ,  $z = -1.04$ ,  $p = 0.297$ ). Results remain stable when controlling for interindividual heterogeneity by including personality traits (Model 4).

### 5.2.3. Discussion (Conjunction Fallacy)

Median response times for conjunction-fallacy questions were in the 22–26 s range. Error rates show that the non-conflict versions of the basic (conflict) conjunction-fallacy question were easier. Hence we are confident that the question manipulation worked as intended. However, there were no significant differences in response times. Taking advantage of the panel structure of the data, and controlling for individual differences, the regression revealed a significant positive effect of the alignment question



on response times (contrary to Hypothesis H1), but no effect of neutral questions.

One possible explanation for these disappointing results is related to the structure of the questions in detail. By their very nature, these questions seek to consider stereotypes. One alternative presents an event  $E$ , the other alternative the conjunction of events  $E$  and  $F$  (or, in the case of (CFA), their disjunction). In the conflict question (CFC), the frame is stereotypically consistent with  $F$ , hence " $E$  and  $F$ " becomes an incorrect, intuitive response. In (CFN1) and (CFN2), the intention was to have an event  $F$  unrelated to the frame, hence shutting down stereotypical thinking. In (CFA), the frame is stereotypically consistent with  $F$ , but the introduction of a disjunction makes the answer " $E$  or  $F$ " correct.

The process logic operates under the assumption that a stereotype-based, intuitive process will select one answer or the other on the basis of the match between frame and events, and a more deliberative process will operate on the basis of the logic of probability. That the latter is indeed active is evidenced by the sharp drop in the error rate from (CFC) to (CFA), where the disjunction is introduced. However, the characteristics of the stereotypical process might not be fully understood. For instance, in all four questions, there is a basic stereotypical inconsistency between events  $E$  and  $F$ . This might activate stereotypical thinking even in the neutral questions (CFN1) and (CFN2), and create a conflict in the alignment question (CFA). In other words, the basic structure of conjunction-fallacy questions might make it difficult to disentangle stereotypical thinking and deliberative processes. Further research should hence try to isolate the actual process involved in stereotypical thinking for this kind of questions.

Personality factors had no effect on response times for conjunction-fallacy questions, with the exception of Need for Cognition, for which higher scores resulted in faster responses. Regarding actual responses, higher scores in the CRT reduced errors (particularly under conflict) as did higher numeracy scores, but FI, NFC, and AOT had no effect. From the Big Five Inventory, only Conscientiousness and Neuroticism had

a significant (positive) effect on correct answers, reducing errors.

### 5.3. Ratio Bias

The last problem refers to the ratio bias (Kirkpatrick and Epstein, 1992; Denes-Raj and Epstein, 1994), which is the tendency to judge a low-probability event as more likely when it is presented as a ratio of large numbers (e.g., 10 in 100) than as a smaller-numbered ratio (e.g., 1 in 10). For instance, in a study by Denes-Raj and Epstein (1994), a majority of participants preferred an 8-in-100 chance of winning to a 1-in-10 chance of winning.

#### 5.3.1. Question Analysis (Ratio Bias, Conflict vs. Neutral)

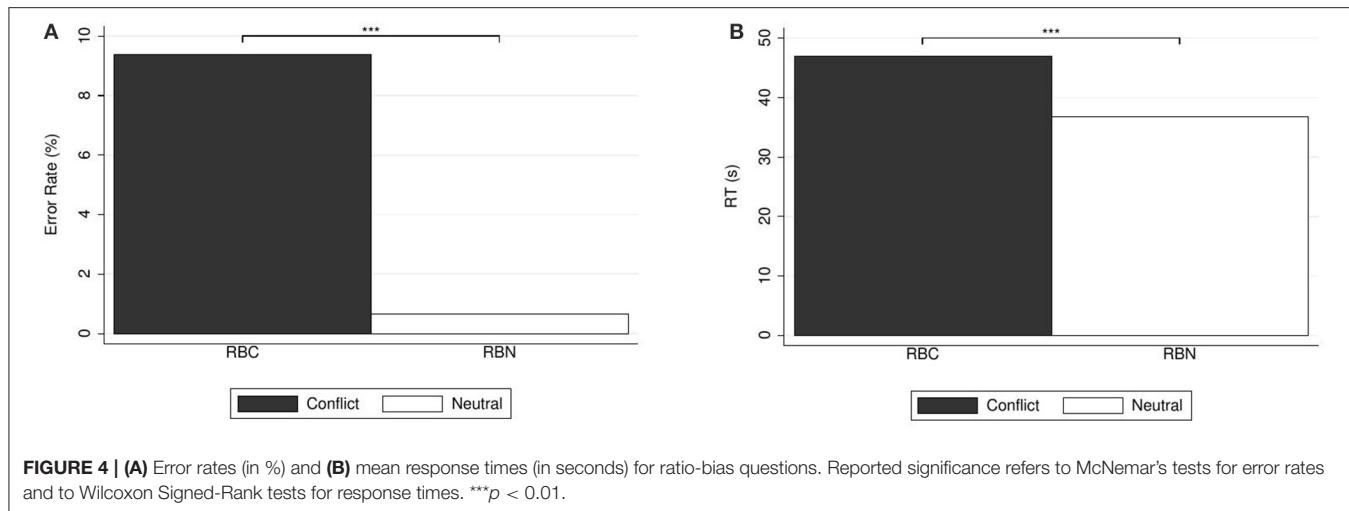
We selected one of the scenarios used by Denes-Raj and Epstein (1994) and complemented it with a non-conflict version as follows.

**(RBC)** There are two urns, a large one containing 100 balls, and a small one containing 10 balls. You must choose one of the urns. A single ball will be extracted at random from the urn you choose, and if the ball is black, you will win. If it is white, you will lose. The small urn contains 1 black ball and 9 white balls. The large urn contains 8 black balls and 92 white balls. Which urn should you choose to maximize the probability of winning?

- The large urn.
- The small urn.

**(RBN)** There are two urns, a large one containing 100 balls, and a small one containing 90 balls. You must choose one of the urns. A single ball will be extracted at random from the urn you choose, and if the ball is black, you will win. If it is white, you will lose. The small urn contains 8 black balls and 82 white balls. The large urn contains 8 black balls and 92 white balls. Which urn should you choose to maximize the probability of winning?

- The large urn.
- The small urn.



For the (RBC) item, the intuitive but incorrect answer is to select the first option. In (RBN), the number of winning balls in the small urn is changed to make both urns contain the same number of winning balls. The heuristic which led participants in (RBC) to choose the urn with the biggest number of winning balls cannot be applied anymore, but comparing proportions is still possible. Therefore (RBN) is a neutral counterpart of the conflict item (RBC).

Figure 4 depicts the percentages of errors and correct responses (panel A) and the response times (panel B). Answers to the conflict question were significantly slower than the answers to the neutral question (median response time 42.10 s, mean 46.93 s,  $SD = 17.73$  in case of conflict; median 35.40 s, mean 36.89 s,  $SD = 12.89$  in case of neutral; WSR test,  $N = 149$ ,  $z = 5.36$ ,  $p < 0.001$ ). This is inconsistent with Hypothesis H2 (and also opposite to the results for CRT questions (Q2C) and (Q2N), which did conform to H2).

Regarding error rates, for (RBC) there were 9.40% (14) errors, and 90.60% (135) correct answers. For (RBN) there was only 1 error (0.67%), and all other 148 answers (99.33%) were correct. Of course, the proportion of errors in the conflict question was significantly larger than in the alignment question [McNemar's test,  $N = 149$ ,  $\chi^2_{(1)} = 11.27$ ,  $p = 0.001$ ].

As can be seen from Table 5, a higher score in the classic 3-item CRT led to a significantly higher likelihood of answering the (RBC) item correctly. In contrast, none of the scores in CRT2, FI, or NFC affected the likelihood of correct answers. In particular, we fail to reproduce the result of Pacini and Epstein (1999), who reported a more pronounced ratio bias for participants low in NFC.

### 5.3.2. Discussion (Ratio Bias)

Error rates were quite low even for the conflict question, but response times were relatively long (medians in the 35–40 s range). The difference in error rates shows that the neutral question worked as intended, with the conflict being removed by shutting down the intuitive process. However, response times were longer for the conflict question, in contradiction

with Hypothesis H2. In contrast, the result is compatible with the view that the conflict question induces a struggle between different tendencies which makes the decision more difficult and results in longer deliberation times, analogously to the “closeness to indifference” argument inspired by Dashiell (1937) and Mosteller and Nogee (1951). The dual-process logic under which Hypothesis H2 was derived (which views conflict resolution as a relatively short part of the decision process) might be more appropriate for shorter decisions as those studied for CRT questions, but the “closeness to indifference” view of tradeoffs and conflicts might be more appropriate for long decisions as those related to our ratio-bias questions. This points out to a need for more detailed models of decision processes, especially if they are to encompass relatively long decisions.

## 6. GENERAL DISCUSSION

Our work provides novel evidence on response times and the multiplicity of decision processes for a category of questions which are extensively used in the decision-making literature. Since responses in experiments in this domain are relatively slow (medians around half a minute or above), our research had an exploratory character.

We selected two kinds of items, those arising from the Cognitive Reflection Test and extensions thereof, and those used to measure decision biases for probability judgments. All such questions create a conflict between an intuitive process and more deliberative thinking, in the terms of dual-process theories. Our strategy of research was to create a non-conflict version for each item, by either making the intuitive impulse correct (resulting in an alignment question), shutting it down (creating a neutral question), or making it dominant (creating a heuristic question).

For CRT items, results were encouraging. The differences in response times are as predicted by dual-process theories, with alignment and heuristic variants leading to faster responses and neutral questions to slower responses than the original, conflict questions. That is, even though response times are relatively long (well above those found in typical experimental

**TABLE 8 | Random-effects regressions on log response times of conjunction-fallacy questions.**

Log(response time)	Model 1	Model 2	Model 3
Neutral	−0.056 (0.125)	−0.056 (0.126)	−0.007 (0.128)
Alignment	0.084*** (0.028)	0.084*** (0.028)	0.143*** (0.045)
Number of letters	0.005** (0.002)	0.005** (0.002)	0.003* (0.002)
Swiftness		0.190*** (0.069)	0.183*** (0.070)
Numeracy		−0.002 (0.022)	0.004 (0.023)
Male		−0.012 (0.041)	−0.010 (0.041)
Faith in intuition		0.005 (0.028)	0.008 (0.026)
Need for cognition		−0.029** (0.012)	−0.028** (0.011)
Openness to experience		−0.009 (0.031)	−0.010 (0.030)
Conscientiousness		0.006 (0.014)	0.007 (0.014)
Extraversion		0.003 (0.011)	0.003 (0.011)
Agreeableness		0.013 (0.026)	0.013 (0.027)
Neuroticism		−0.006 (0.012)	−0.003 (0.012)
Actively open-minded thinking		−0.005 (0.024)	0.000 (0.024)
Error			0.156*** (0.048)
Error × alignment			−0.104 (0.132)
Observations	632	632	632

The dummies Neutral and Alignment take the value 1 for the respective versions of the questions; Conflict is the reference category. Standard errors (in parentheses) are (conservatively) clustered at the level of counterbalance condition (question order).

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

workhorses for dual-process theories), evidence is consistent with the involvement of different decision processes and the diagnosticity of their interaction (conflict or not).

For decision-bias items, results were sobering. Results on conflict vs. alignment for base-rate neglect questions were inconclusive on the aggregate in spite of significant effects for some individual items. In our opinion, this points out that the heavily-framed questions employed in this area are not stylized enough to properly identify the involved processes, and further efforts are needed in order to disentangle framing and the effects of conflict or alignment among decision processes. For the conjunction fallacy, response-time differences were generally not significant, even though the manipulations worked

as intended in terms of error rates. In view of the structure of the items, we tentatively conclude that stereotypical thinking cannot be properly isolated with the standard frames used to study the conjunction fallacy, and recommend further research to move away from this basic structure. For the ratio-bias item, where response times are particularly long, we obtained a clear result showing that a neutral version of the original, conflict question results in lower error rates and shorter response times. This is compatible with the view that, in this case, process conflict reflects a stronger behavioral struggle resulting in longer deliberation (following a classic “closeness to indifference” argument).

## 6.1. Response Times and Underlying Assumptions

It is worth discussing possible explanations for the differences in results between the CRT items and the items on decision biases. Two avenues are apparent, one procedural and one conceptual.

The procedural avenue concerns the fact that our implementation of the decision-bias items involved binary choices, while the CRT items were open-ended [with the exception of (Q4C) and (Q4A)]. The reason is that, for the CRT items, the exactly correct answer is still reasonably easy to arrive at, and the alternative, intuitive process provides a specific answer. Hence the open-answer format is natural. In contrast, for the base-rate questions the postulated processes do not deliver precise answers. Correct answers are the result of complex, precise calculations while “intuitive” tendencies have a directional nature (high or low probability estimate). Hence, we presented those items with binary-choice answers (larger or smaller than 50%). However, it is unlikely that this procedural difference is determinant for the difference in results. First, we did not compare response times of different answers for a fixed question, but rather the response times for different *questions*. Whatever answers the different processes led to, differences among types of questions should subsist. Second, for the items related to the conjunction fallacy and the ratio bias, the binary-choice format is indeed natural, because the correct response is easy to arrive at, and the alternative intuitive processes do provide a clear response. However, it remains at least conceivable that for the base-rate-bias item pairs, the presentation of binary-choice answers lowered participant involvement (since there was no need to arrive at an exact numerical estimate), hence activating decision processes different from those postulated in the analysis. This, however, would have no bearing on the conjunction-fallacy and ratio-bias items.

The conceptual avenue arises from our discussion in Section 2.4. Our hypotheses are always derived from the confluence of two effects. First, the time required for conflict detection and resolution should be smaller if there is no actual conflict. Second, the kind of question should affect the percentage of intuitive (hence faster) decisions. Evidence on conflict detection and resolution, however, indicates that the required time might be relatively short. Specifically, EEG research shows that conflict detection and resolution are probably associated with activity in the anterior cingulate cortex occurring as

**TABLE 9 | Random-effects probit regressions on correct answers to conjunction-fallacy questions.**

Correct answer	Model 1	Model 2	Model 3	Model 4
Male	0.062 (0.151)	-0.092 (0.116)	-0.088 (0.117)	0.003 (0.110)
Conflict	-1.173*** (0.226)	-1.175*** (0.226)	-1.072 (0.788)	-1.128 (0.775)
Neutral	-0.491*** (0.119)	-0.492*** (0.123)	-0.489*** (0.122)	-0.499*** (0.117)
Classic CRT		0.169*** (0.047)	0.089** (0.037)	0.111* (0.059)
CRT2		-0.063 (0.058)	-0.030 (0.030)	-0.013 (0.055)
Faith in intuition		0.070 (0.066)	0.076 (0.064)	0.085 (0.074)
Need for cognition		-0.005 (0.041)	0.012 (0.028)	0.017 (0.048)
Numeracy		0.081* (0.048)	0.084* (0.048)	0.074** (0.036)
Conflict × classic CRT		0.250*** (0.095)	0.248** (0.091)	
Conflict × CRT2		-0.106 (0.110)	-0.098 (0.112)	
Conflict × FI		-0.014 (0.116)	-0.011 (0.108)	
Conflict × NFC		-0.059 (0.104)	-0.057 (0.105)	
Openness to experience			-0.012 (0.049)	
Conscientiousness			0.020** (0.006)	
Extraversion			-0.001 (0.026)	
Agreeableness			-0.017 (0.055)	
Neuroticism			0.092*** (0.032)	
Actively open-minded thinking			0.094 (0.068)	
Observations	632	632	632	632

The dummies Conflict and Neutral take the value 1 for the respective versions of the questions; Alignment is the reference category. Standard errors (in parentheses) are (conservatively) clustered at the level of counterbalance condition (question order).

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

early as 200 milliseconds (see, e.g., Nieuwenhuis et al., 2003; Coderre et al., 2011; Achziger et al., 2014). Although this evidence has been gathered for paradigms with simple stimuli (suitable for EEG analysis), it can be speculated that even for paradigms with longer response times as the ones considered here, the time necessary for conflict detection and resolution in the sense of dual-process theories is relatively short (if at all relevant), implying that response-time effects should be driven by the second phenomenon described above, namely

the shift in likelihood from one type of process to the other.

If, for the sake of the argument, we accept this preliminary hypothesis, we can reexamine our results. Evidence from our CRT item pairs is compatible with the postulate that, relative to conflict items, the balance is shifted toward automatic processes in case of alignment, and obviously toward deliberative processes for neutral items. Since the former are faster than the latter, this observation suffices to explain our data. It is also compatible with the fact that smaller error rates are observed in both cases, since in case of alignment intuitive processes also deliver the correct answer.

For our decision-bias items, as shown in the analysis sections above, it remains true that error rates for alignment and neutral items are lower than for the corresponding conflict items (the comparison was significant in every single case). Hence, based on the error-rate evidence, we have no reason to doubt that in every item pair, the process shift occurred as postulated. However, response-time evidence appears inconsistent. At this point of the argument, the original response-time predictions rest on a single assumption, namely that the expected response times of the deliberative processes for these items are indeed always larger than the expected response times of the corresponding intuitive processes ( $T_U > T_H$ ). Given the simple nature of the processes involved in the CRT questions, there is little reason to question this assumption in that setting. For decision-bias items, however, it has been argued that many of the involved heuristics might not be fast shortcuts, but rather “cognitive heuristics” including multi-step operations (even if they are sometimes called “fast and frugal,” Gigerenzer and Goldstein, 1996). If this is the case for the intuitive processes involved in decision biases of the type examined here, then the assumption of a significant difference in expected response times among processes for this particular case might not be justified. Our data is consistent with this interpretation, but further evidence is needed.

## 6.2. Long Response Times and Types of Decisions

At this point, we can conclude that the scope of response times has a definite influence on their interpretation. For short response times it is comparatively easier to identify the involved decision processes and simple dual-process models deliver instructive predictions. For longer decisions, the exact length thereof might reflect moderators of deliberation, and predictions should be more modest at this point. Clearly, there is a need for improved models of deliberation and the associated process data.

It should be kept in mind that we have concentrated on decisions from inference where an objectively correct decision can be identified in advance and natural hypotheses on the nature of the involved processes are available. This is in stark contrast to preferential-choice settings, where the nature of the involved processes is open to discussion. For instance, Cappelen et al. (2015) examined response times in the dictator game and argued that choosing a fair allocation of resources among two people (as opposed to keeping most of a given

resource for oneself) might be more intuitive, because the average response times were shorter (but still quite long). As observed by Myrseth and Wollbrant (2016), this might amount to a reverse-inference fallacy, especially since the conclusion is not based on a theoretical model, but rather operates as if there was a one-to-one correspondence between processes and choices (see Alós-Ferrer, 2016 for a discussion of this point). Further, preferential choice presents an added difficulty. We have concentrated our analysis on response-time differences *across questions*, which enable paired comparisons of data. This is important because there exists a large response-time heterogeneity across individuals, which becomes exacerbated for long response times as the ones we study. In studies of preferential choice as Cappelen et al. (2015), there is exactly one observation per individual, and the population of subjects is partitioned according to the response. Hence, individual heterogeneity is harder to control for. This is why our analysis focused on paired-observations tests, moving to regressions only to clarify the possible effect of additional individual correlates.

### 6.3. Personality Measures

Our analysis also points out the necessity of further research on the influence of personality traits on decision-making biases. In spite of some clear general trends, evidence is still mixed. We found that higher scores in the CRT resulted in significantly more correct responses for both the conjunction fallacy and the ratio bias. However, we did not find a clear predictive effect of higher scores in the CRT on correct responses for base-rate-neglect questions. Faith in Intuition, Need for Cognition, and Actively Open-Minded Thinking were generally non-predictive for correct responses in our sample. However, we used the short REI-10 version with 5 items per subscale, while Alós-Ferrer and Hügelschäfer (2016) used a 15-item version.

Regarding the Big Five Inventory, we confirmed the typical correlations with other personality traits found in the literature. We included them as controls in regressions on both choices

and response times for the base-rate-neglect and conjunction-fallacy items. We found significant effects, but none of the five personality traits showed a consistent effect for base-rate neglect and the conjunction fallacy. For instance, Extraversion resulted in more base-rate-neglect errors but had no effect on the conjunction-fallacy items. This is especially interesting, because this personality trait has been related to a more sensitive midbrain dopaminergic reward system, leading to difficulties in regulating impulsiveness (Depue and Collins, 1999; Cohen et al., 2005).

We conclude that the effects of personality measures often appear to be bias-specific, and apparently related constructs, which are supposed to measure related traits, often have different effects. The CRT is predictive for different decision biases, but the scale is becoming generally known and, contrary to self-report questionnaires, cannot be reliably measured repeatedly. Subscales from the Rational-Experiential Inventory have a predictive value (recall Alós-Ferrer and Hügelschäfer, 2012, 2016), but the effects appear to be small in general. Personality traits from the Big Five Inventory often have significant effects, but those are generally inconsistent across biases. Larger datasets, allowing for the study of multiple interactions, might contribute to obtain a more clear picture.

## AUTHOR CONTRIBUTIONS

All authors contributed equally to this work. The listing of authors is alphabetical.

## ACKNOWLEDGMENTS

This research was financed by the Research Unit “Psychoeconomics,” funded by the German Research Foundation (DFG, FOR 1882). The studies were conducted at the CLER (Cologne Laboratory for Economic Research). The CLER gratefully acknowledges financial support from the German Research Foundation (DFG).

## REFERENCES

- Achtziger, A., and Alós-Ferrer, C. (2014). Fast or rational? A response-times study of Bayesian updating. *Manage. Sci.* 60, 923–938. doi: 10.1287/mnsc.2013.1793
- Achtziger, A., Alós-Ferrer, C., Hügelschäfer, S., and Steinhäuser, M. (2014). The neural basis of belief updating and rational decision making. *Soc. Cogn. Affect. Neurosci.* 9, 55–62. doi: 10.1093/scan/nss099
- Alós-Ferrer, C. (2016). A dual-process diffusion model. *J. Behav. Decis. Making.* doi: 10.1002/bdm.1960
- Alós-Ferrer, C., and Hügelschäfer, S. (2012). Faith in intuition and behavioral biases. *J. Econ. Behav. Organ.* 84, 182–192. doi: 10.1016/j.jebo.2012.08.004
- Alós-Ferrer, C., and Hügelschäfer, S. (2016). Faith in intuition and cognitive reflection. *J. Behav. Exp. Econ.* 64, 61–70. doi: 10.1016/j.soec.2015.10.006
- Alós-Ferrer, C., and Strack, F. (2014). From dual processes to multiple selves: implications for economic behavior. *J. Econ. Psychol.* 41, 1–11. doi: 10.1016/j.jeop.2013.12.005
- Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychol.* 44, 211–233. doi: 10.1016/0001-6918(80)90046-3
- Baron, J. (1993). Why teach thinking? An essay. *Appl. Psychol.* 42, 191–214. doi: 10.1111/j.1464-0597.1993.tb00731.x
- Brañas-Garza, P., García-Muñoz, T., and González, R. H. (2012). Cognitive effort in the beauty contest game. *J. Econ. Behav. Organ.* 83, 254–260. doi: 10.1016/j.jebo.2012.05.018
- Cappelen, A. W., Nielsen, U. H., Tungodden, B., Tyran, J.-R., and Wengström, E. (2015). Fairness is intuitive. *Exp. Econ.* doi: 10.1007/s10683-015-9463-y
- Coderre, E., Conklin, K., and van Heuven, W. J. B. (2011). Electrophysiological measures of conflict detection and resolution in the Stroop task. *Brain Res.* 1413, 51–59. doi: 10.1016/j.brainres.2011.07.017
- Cohen, M. X., Young, J., Baek, J.-M., Kessler, C., and Ranganath, C. (2005). Individual differences in extraversion and dopamine genetics predict neural reward responses. *Cogn. Brain Res.* 25, 851–861. doi: 10.1016/j.cogbrainres.2005.09.018
- Cueva, C., Iturbe-Ormaetxe, I., Mata-Pérez, E., Ponti, G., Sartarelli, M., Yu, H., et al. (2016). Cognitive (Ir)reflection: new experimental evidence. *J. Behav. Exp. Econ.* 64, 81–93. doi: 10.1016/j.soec.2015.09.002
- Dashiell, J. F. (1937). Affective value-distances as a determinant of aesthetic judgment-times. *Am. J. Psychol.* 50, 57–67. doi: 10.2307/1416620
- De Neys, W., and Bonnefon, J.-F. (2013). The ‘whys’ and ‘whens’ of individual differences in thinking biases. *Trends Cogn. Sci.* 17, 172–178. doi: 10.1016/j.tics.2013.02.001

- De Neys, W., Rossi, S., and Houdé, O. (2013). Bats, balls, and substitution sensitivity: cognitive misers are no happy fools. *Psychon. Bull. Rev.* 20, 269–273. doi: 10.3758/s13423-013-0384-5
- Denes-Raj, V., and Epstein, S. (1994). Conflict between intuitive and rational processing: when people behave against their better judgment. *J. Pers. Soc. Psychol.* 66, 819–829. doi: 10.1037/0022-3514.66.5.819
- Depue, R. A., and Collins, P. F. (1999). Neurobiology of the structure of personality: dopamine, facilitation of incentive motivation, and extraversion. *Behav. Brain Sci.* 22, 491–569. doi: 10.1017/S0140525X99002046
- Eddy, D. (1982). "Probabilistic reasoning in clinical medicine: problems and opportunities," in *Judgment Under Uncertainty: Heuristics and Biases*, eds D. Kahneman, P. Slovic, and A. Tversky (Cambridge: Cambridge University Press), 249–267. doi: 10.1017/CBO9780511809477.019
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *Am. Psychol.* 49, 709–724. doi: 10.1037/0003-066X.49.8.709
- Epstein, S., Pacini, R., Denes-Raj, V., and Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *J. Pers. Soc. Psychol.* 71, 390–405. doi: 10.1037/0022-3514.71.2.390
- Erev, I., Shimonowitch, D., Schurr, A., and Hertwig, R. (2008). "Base rates: how to make the intuitive mind appreciate or neglect them," in *Intuition in Judgment and Decision Making*, eds H. Plessner, C. Betsch, and T. Betsch (Mahwah, NJ: Lawrence Erlbaum), 135–148.
- Evans, J. S. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Ann. Rev. Psychol.* 59, 255–278. doi: 10.1146/annurev.psych.59.103006.093629
- Feingold, A. (1994). Gender differences in personality - a metaanalysis. *Psychol. Bull.* 116, 429–456. doi: 10.1037/0033-2909.116.3.429
- Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychol. Rev.* 107, 659–676. doi: 10.1037/0033-295X.107.4.659
- Fischbacher, U. (2007). z-Tree: zurich toolbox for ready-made economic experiments. *Exp. Econ.* 10, 171–178. doi: 10.1007/s10683-006-9159-4
- Frederick, S. (2005). Cognitive reflection and decision making. *J. Econ. Perspect.* 19, 25–42. doi: 10.1257/089533005775196732
- Gerlitz, J.-Y., and Schupp, J. (2005). Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP. *DIW Research Notes* 2005-04.
- Gigerenzer, G., and Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychol. Rev.* 103, 650–669. doi: 10.1037/0033-295X.103.4.650
- Greiner, B. (2004). An online recruiting system for economic experiments. *Forschung und wissenschaftliches Rechnen GWDG Bericht* 63, 79–93.
- Haran, U., Ritov, I., and Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgm. Decis. Making* 8, 188–201.
- Hoppe, E. I., and Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. *Econ. Lett.* 110, 97–100. doi: 10.1016/j.econlet.2010.11.015
- Johnson, E. D., Tubau, E., and De Neys, W. (2016). The doubting system 1: evidence for automatic substitution sensitivity. *Acta Psychol.* 164, 56–64. doi: 10.1016/j.actpsy.2015.12.008
- Kahneman, D., and Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cogn. Psychol.* 3, 430–454. doi: 10.1016/0010-0285(72)90016-3
- Kahneman, D., and Tversky, A. (1973). On the psychology of prediction. *Psychol. Rev.* 80, 237–271. doi: 10.1037/h0034747
- Kandori, M., Mailath, G. J., and Rob, R. (1993). Learning, mutation, and long run equilibria in games. *Econometrica* 61, 29–56. doi: 10.2307/2951777
- Kirkpatrick, L. A., and Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: further evidence for two conceptual systems. *J. Pers. Soc. Psychol.* 63, 534–544. doi: 10.1037/0022-3514.63.4.534
- Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., and Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *J. Behav. Decis. Making* 25, 361–381. doi: 10.1002/bdm.752
- Lipkus, I. M., Samsa, G., and Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Med. Decis. Making* 21, 37–44. doi: 10.1177/0272989X0102100105
- McCrae, R., and Costa, P. (1985). *The NEO Personality Inventory Manual*. Odessa, FL: Psychological Assessment Resources.
- Miller, J. (1991). Reaction time analysis with outlier exclusion: bias varies with sample size. *Q. J. Exp. Psychol.* 43, 907–912. doi: 10.1080/14640749108400962
- Mosteller, F., and Nogee, P. (1951). An experimental measurement of utility. *J. Polit. Econ.* 59, 371–404. doi: 10.1086/257106
- Myrseth, K. O., and Wollbrant, C. E. (2016). Commentary: fairness is intuitive. *Front. Psychol.* 7:654. doi: 10.3389/fpsyg.2016.00654
- Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., and Ridderinkhof, K. (2003). Electrophysiological correlates of anterior cingulate function in a Go/No-Go task: effects of response conflict and trial type frequency. *Cogn. Affect. Behav. Neurosci.* 3, 17–26. doi: 10.3758/CABN.3.1.17
- Oechssler, J., Roider, A., and Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *J. Econ. Behav. Organ.* 72, 147–152. doi: 10.1016/j.jebo.2009.04.018
- Pacini, R., and Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *J. Pers. Soc. Psychol.* 76, 972–987. doi: 10.1037/0022-3514.76.6.972
- Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., and Hamilton, J. (2015). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *J. Behav. Decis. Making*. doi: 10.1002/bdm.1883. [Epub ahead of print].
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychol. Bull.* 119, 3–22. doi: 10.1037/0033-2909.119.1.3
- Strack, F., and Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Pers. Soc. Psychol. Rev.* 8, 220–247. doi: 10.1207/s15327957pspr0803\_1
- Toplak, M. E., West, R. F., and Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Mem. Cogn.* 39, 1275–1289. doi: 10.3758/s13421-011-0104-1
- Toplak, M. E., West, R. F., and Stanovich, K. E. (2014). Assessing miserly information processing: an expansion of the cognitive reflection test. *Think. Reason.* 20, 147–168. doi: 10.1080/13546783.2013.844729
- Tversky, A., and Kahneman, D. (1980). "Causal schemas in judgments under uncertainty," in *Progress in Social Psychology*, Vol. I, ed M. Fishbein (Hillsdale, NJ: Lawrence Erlbaum Associates), 49–72.
- Tversky, A., and Kahneman, D. (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol. Rev.* 90, 293–315. doi: 10.1037/0033-295X.90.4.293
- Weisberg, Y. J., DeYoung, C. G., and Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the Big Five. *Front. Psychol.* 2:178. doi: 10.3389/fpsyg.2011.00178
- Young, H. P. (1993). The evolution of conventions. *Econometrica* 61, 57–84. doi: 10.2307/2951778
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2016 Alós-Ferrer, Garagnani and Hügelshäfer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Creativity and Cognitive Skills among Millennials: Thinking Too Much and Creating Too Little

Brice Corgnet<sup>1</sup>, Antonio M. Espín<sup>2,3\*</sup> and Roberto Hernán-González<sup>3,4</sup>

<sup>1</sup> EMLYON Business School, Univ Lyon, GATE L-SE UMR 5824, Ecully, France, <sup>2</sup> Economics Department, Middlesex University Business School, London, UK, <sup>3</sup> Granada Lab of Behavioral Economics, Universidad de Granada, Granada, Spain, <sup>4</sup> Business School, University of Nottingham, Nottingham, UK

## OPEN ACCESS

### Edited by:

Nikolaos Georgantzas,  
University of Reading, UK

### Reviewed by:

Noelia Sánchez-Pérez,  
University of Murcia, Spain  
Conny Ernst-Peter Wollbrant,  
University of Gothenburg, Sweden

### \*Correspondence:

Antonio M. Espín  
a.espin@mdx.ac.uk

### Specialty section:

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

Received: 20 July 2016

Accepted: 05 October 2016

Published: 25 October 2016

### Citation:

Corgnet B, Espín AM and  
Hernán-González R (2016) Creativity  
and Cognitive Skills among Millennials:  
Thinking Too Much and Creating Too  
Little. *Front. Psychol.* 7:1626.  
doi: 10.3389/fpsyg.2016.01626

Organizations crucially need the creative talent of millennials but are reluctant to hire them because of their supposed lack of diligence. Recent studies have shown that hiring diligent millennials requires selecting those who score high on the Cognitive Reflection Test (CRT) and thus rely on effortful thinking rather than intuition. A central question is to assess whether the push for recruiting diligent millennials using criteria such as cognitive reflection can ultimately hamper the recruitment of creative workers. To answer this question, we study the relationship between millennials' creativity and their performance on fluid intelligence (Raven) and cognitive reflection (CRT) tests. The good news for recruiters is that we report, in line with previous research, evidence of a positive relationship of fluid intelligence, and to a lesser extent cognitive reflection, with *convergent creative thinking*. In addition, we observe a positive effect of fluid intelligence on originality and elaboration measures of *divergent creative thinking*. The bad news for recruiters is the inverted U-shape relationship between cognitive reflection and fluency and flexibility measures of *divergent creative thinking*. This suggests that thinking too much may hinder important dimensions of creative thinking. Diligent and creative workers may thus be a rare find.

**Keywords:** creativity, cognitive reflection, intelligence, cognition, intuition

## INTRODUCTION

Evidence from a recent survey reports that managers are three times more likely to hire a mature worker than to hire a *millennial* (born between 1980 and 2000; Rainer and Rainer, 2011) despite desperately needing their creative talent<sup>1</sup>. Mature workers are appealing to recruiters because they are seen as more reliable and more committed than millennials. The dilemma for managers is thus to hire millennials that are both diligent and creative.

Recent studies have shown that firms can secure the hiring of diligent millennials by relying on measures of cognitive skills. For example, intelligence has been found to be the main predictor of overall work performance in a wide variety of occupations and across age and gender (e.g., Hunter and Hunter, 1984; Olea and Ree, 1994; see Schmidt, 2009 for a review). Standard measures of cognitive ability have been found to correlate positively with task performance (Schmidt et al., 1986; Murphy, 1989) and negatively with counterproductive work behaviors such as theft or absenteeism

<sup>1</sup>See the following press release: <http://www.forbes.com/sites/susanadams/2012/09/24/older-workers-theres-hope-study-finds-employers-like-you-better-than-millennials/#1f5799cb4aa6> (accessed September 21, 2016).

(Dilchert et al., 2007). Moreover, the results of a recent study suggest that these effects may be mediated by individuals' cognitive styles (Corgnet et al., 2015b). In particular, Corgnet et al. (2015b) find that millennials characterized by a more reflective style (as measured by the Cognitive Reflection Test; Frederick, 2005) are more diligent, displaying higher levels of task performance and lower levels of counterproductive work behaviors<sup>2</sup>. A crucial caveat is whether hiring millennials based on cognitive measures may ultimately select less creative workers. To address this point we need to assess the relationship between cognitive skills and creativity.

Traditionally, intelligence, and creativity have been considered to be unrelated (Getzels and Jackson, 1962; Wallach and Kogan, 1965; Batey and Furnham, 2006; Sawyer, 2006; Weisberg, 2006; Runco, 2007; Kaufman, 2009; Kim et al., 2010). In a meta-analysis, Kim (2005) finds that the correlation between creativity test scores and IQ varies widely and is, on average, small ( $r = 0.174$ ).

However, a growing consensus has emerged in recent research stressing a close relationship between intelligence and creative performance (see Silvia, 2015, for a review). This emerging consensus heavily relies on recent studies that have employed more sophisticated statistical techniques and more robust assessment methods than prior research on the topic. For example, the use of latent variable models has allowed researchers to uncover a positive and significant relationship between creativity and intelligence using data from previous studies that reported non-significant correlations (Silvia, 2008b). The recent wave of research on intelligence and creativity has also improved upon traditional assessment of creativity that exclusively relied on scoring methods based on the originality and uniqueness of responses in creative tasks (such as finding unusual uses for an object). These traditional scoring methods are imprecise because they confound several factors, such as fluency and sample size (Hocevar, 1979; Silvia et al., 2008), and can thus lead to inaccurate estimates of the relationship between intelligence and creativity (Silvia, 2008a; Nusbaum and Silvia, 2011). The results of this new wave of research on creativity and intelligence have been taken as evidence that executive cognition is undoubtedly beneficial to creative thinking (Silvia, 2015).

Yet, although there is an obvious link between intelligence and executive cognition, from the point of view of modern dual-process theory (Evans, 2008, 2009; Stanovich, 2009, 2010; Evans and Stanovich, 2013), one should distinguish between algorithmic and reflective cognitive processes. Algorithmic processes are typically associated with computational efficiency and are measured by standard intelligence tests whereas reflective processing is associated with a disposition to employ the resources of the algorithmic mind, that is, to switch from autonomous "Type 1" thought to analytic "Type 2" (working memory-dependent) thought. The reflective mind thus has

a disposition-based definition ("cognitive styles", reflective vs. intuitive) and is not adequately measured by standard intelligence tests (which assess "cognitive ability") but by tasks of cognitive reflection like the Cognitive Reflection Test (CRT; Frederick, 2005). Individuals characterized by a more reflective mind tend to show higher levels of self-control and lower levels of "cognitive impulsivity" (Frederick, 2005; Kahneman and Frederick, 2007; Cokely and Kelley, 2009; Oechsler et al., 2009; Toplak et al., 2011; Brañas-Garza et al., 2012).

From this perspective, one can conjecture that cognitive reflection may relate negatively to creativity. This is the case because a number of studies suggest that the capacity to control one's attention and behavior may even be detrimental for creative thinking (for a review, see Wiley and Jarosz, 2012a). For example, creative problem solving has been shown to relate positively to moderate alcohol intoxication (Jarosz et al., 2012), which is known to impair inhibition and attentional control (Peterson et al., 1990; Kovacevic et al., 2012; Marinkovic et al., 2012). Similarly, an "experiential" thinking style (which maps onto Type 1 processing) has been found to correlate positively with creative performance (Norris and Epstein, 2011).

As mentioned, past literature arrived at conflicting conclusions regarding whether executive cognition favors (e.g., Nusbaum and Silvia, 2011; Beaty and Silvia, 2012; Silvia, 2015) or hampers (e.g., Eysenck, 1993; Kim et al., 2007; Ricks et al., 2007; Norris and Epstein, 2011; Jarosz et al., 2012; Wiley and Jarosz, 2012b) creative thinking. Dual-process theory can reconcile these apparently conflicting findings by positing that creativity may be generated by a mix of Type 1 and Type 2 processes (Allen and Thomas, 2011; Ball et al., 2015; Barr et al., 2015; see Sowden et al., 2015, for a review). It follows that the dual-process approach lays out a promising research agenda based on assessing the exact mix of Type 1 and Type 2 processes that bolsters creativity as well as analyzing separately the effect of algorithmic and reflective Type 2 processes on creative thinking.

Following a dual-process approach, Barr et al. (2015) find experimental evidence of an important effect of controlled Type 2 analytic processes on both convergent and divergent (Guilford, 1967) creative thinking. In particular, they find that both cognitive ability (measured as the combination of numeracy and verbal skills) and reflective cognitive style (average of scores in the CRT and base-rate problem tasks) covary positively with one's capacity to make remote associations, that is, with convergent creative thinking. Regarding divergent creative thinking, Barr et al. (2015) show that cognitive ability but *not* cognitive reflection predicts higher originality scores in an alternate uses task. Fluency in the latter task, however, was not correlated with either cognitive measure.

In this paper, we use a similar approach to Barr et al. (2015) and investigate how both types of cognitive processes affect creativity. In particular, we analyze how cognitive abilities (measured using Raven as a test of fluid intelligence) and cognitive styles (intuitive vs. reflective; as measured by the CRT) relate to convergent and divergent creative thinking. We extend Barr et al. (2015) by analyzing other measures of divergent thinking such as flexibility and elaboration and

<sup>2</sup>Positive effects of cognitive reflection on people's willingness to choose socially-efficient resource allocations (Lohse, 2016; Capraro et al., 2016) as well as to trust strangers (Corgnet et al., 2016) suggest other possible channels through which organizations may benefit from hiring individuals with a more reflective cognitive style. Cognitive reflection has also been found to play a key role in moral judgment (e.g., Paxton et al., 2012; Pennycook et al., 2014).

by exploring possible non-linearities between creativity and cognitive measures.

Given the conflicting results regarding whether executive cognition is beneficial or detrimental for creative thinking, we conjecture that there might exist a non-linear relationship between different measures of creativity and cognition. Specifically, it might be that a minimum level of executive cognition is necessary for creative performance but, beyond some level, the relationship disappears or even turns negative. This might explain why previous findings seem to be inconsistent. A related line of reasoning has been proposed in the so-called “threshold hypothesis” of the relationship between IQ and creativity (Guilford, 1967; Jauk et al., 2013). The threshold hypothesis states that intelligence is positively related to creative thinking for low IQ levels but the relationship blurs for high IQ levels. Similar arguments arise in recent accounts of the “mad genius hypothesis”: moderate levels of inhibitory or top-down control dysfunction, characteristic of subclinical psychiatric populations (e.g., mild ADHD and schizophrenia disorders), can spur creativity under some conditions whereas clinical-severe levels typically lead to impoverished creative thinking (Schulberg, 2005; Abraham et al., 2007; Jaracz et al., 2012; Acar and Sen, 2013; Abraham, 2014).

## METHODS

### Participants and General Protocol

Participants were 150 students (46.67% female; age: mean  $\pm SD = 20.23 \pm 1.96$ ) from Chapman University in the U.S. These participants were recruited from a database of more than 2000 students. Invitations to participate in the current study were sent to a random subset of the whole database. This study is part of a larger research program on cognition and economic decision making. The local Institutional Review Board approved of this research. All participants provided written informed consent prior to participating. We conducted a total of 12 sessions, nine had 12 participants and three had 14 participants. On average, sessions lasted for 45 min. All subjects completed the same tasks in the following order: (1) CRT, (2) Raven test, (3) Remote associates task, (4) Alternate uses task. Subjects had 6 min to complete each task and a 2-min break after completing the Raven test.

## Measures

### Cognitive Ability Assessment

Participants completed a subset of Raven progressive matrices test (Raven, 1936). Specifically, we used the odd number of the last three series of matrices (Jaeggi et al., 2010; Corgnet et al., 2015a). The number of matrices correctly solved in the Raven test (in our sample, ranging from 9 to 18, mean  $\pm SD = 14.40 \pm 2.42$  for males and  $14.47 \pm 2.16$  for females) is a conventional measure of cognitive ability. This test captures an important aspect of cognitive processing which is referred to as fluid intelligence and is closely related to algorithmic thinking (Stanovich, 2009, 2010).

### Cognitive Style Assessment

We measured the participants’ tendency to rely on intuition vs. reflection using the CRT introduced by Frederick (2005). The test is characterized by the existence of an incorrect response which automatically comes to mind but has to be overridden in order to find the correct solution. To the original CRT questions, we added four questions recently developed by Toplak et al. (2014). This extended task (see Text S1) will allow us to uncover potentially non-linear relationships that would be hard to observe using the classical three-item task (Frederick, 2005). In Table S1, we display the proportion of subjects answering each question correctly, split by gender. As expected, males performed better in the test than females (Frederick, 2005; Bosch-Domènec et al., 2014). Our measure of cognitive reflection is given by the total number of correct answers (from 0 to 7). The full distribution of correct answers by males (mean  $\pm SD = 4.09 \pm 2.31$ ) and females (mean  $\pm SD = 2.89 \pm 2.03$ ) is provided in Figure S1.

### Convergent Creative Thinking

We used a subset of the Remote Associate Test (RAT; Mednick, 1962) to measure subjects’ ability to make remote associations. In particular, subjects were shown 13 sets of three words (e.g., widow-bite-monkey) and asked to find a word which relates to all the three words provided (in this example the solution is “spider”). Our measure of convergent thinking is the number of problems correctly solved (from 0 to 13).

### Divergent Creative Thinking

We measured divergent thinking using a variant of the Alternate Uses Task (AUT; Guilford, 1967). Participants were instructed to provide as many unusual uses of a pen as possible during 6 min. We construct four different measures of divergent thinking: *fluency*, *originality*, *flexibility*, and *elaboration*. We measured *fluency* as the total number of answers provided by a participant. Three raters were presented with a random list of answers and asked to score the degree of originality of each entry using a 1 (*not at all*) to 5 (*very much*) Likert scale. We computed *originality* as the sum of the average score of the three raters for all the entries provided by a participant, divided by the total number of answers. Following Troyer and Moscovitch (2006) and Gilhooly et al. (2007), all the answers were classified in broad differentiated categories (e.g., uses of the pen as cloth or hair accessories). Then, *flexibility* was measured as the number of different categories provided by each participant. Finally, *elaboration* refers to the average amount of detail (from 0 to 2) provided by each participant.

### Statistical Analysis

For the data analysis, we start by showing the descriptive statistics of all the measures used and their zero-order correlations. To further assess the relationships between creativity and cognitive measures, we first provide a graphical representation using LOWESS smoothing (Cleveland, 1979; Cleveland and McGill, 1985). We then run ordinary least squares regressions which allow us to test the statistical significance of the linear and non-linear relationships which were shown in the LOWESS graphs. All the analyses were performed using Stata 14.0.

## RESULTS

### Descriptive Statistics and Correlations

Means, standard deviations, and correlations are shown in **Table 1**. Unsurprisingly, we find moderate positive correlation between the number of correct answers in the CRT and Raven tests ( $r = 0.26$ ,  $p < 0.01$ ) which suggests that CRT and Raven are not entirely measuring the same cognitive skills (Frederick, 2005; Stanovich, 2009, 2010). Similarly, the different measures of divergent thinking (AUT) are significantly correlated (all  $p$ 's  $< 0.01$ ), except for *originality* and *flexibility* ( $p = 0.28$ ).

Regarding our cognitive measures, we find that both Raven ( $p < 0.01$ ) and CRT scores ( $p = 0.03$ ) are positively correlated with convergent thinking (RAT). However, the relationship between cognitive skills and divergent thinking is more complicated. High levels of cognitive ability (Raven) relate positively with *originality* ( $p = 0.01$ ) and *elaboration* ( $p < 0.01$ ), but negatively with the number of answers provided (*fluency*;  $p = 0.04$ ) and non-correlated with *flexibility* ( $p = 0.20$ ). Finally, we do not find a significant correlation between cognitive styles (CRT scores) and any measure of divergent thinking (all  $p$ 's  $> 0.26$ ).

### Non-linear Effects and Regression Analysis

We now turn to the study of possible non-linear relationships between our measures of cognition and creativity. **Figure 1** displays all the relationships under study using LOWESS (bandwidth = 0.8; Cleveland, 1979; Cleveland and McGill, 1985). LOWESS is a model-free smoothing technique based on locally-weighted regressions which can detect both linear and non-linear relationships. In order to compare the effect sizes, we standardize all measures (standard deviations from the mean). We also ran ordinary least squares regressions to assess the statistical significance of the observed relationships. In Tables S2–S6, we present the results of a series of regressions in which we estimated both linear and quadratic effects of each of the predictors (Raven and CRT) separately on each creativity measure (columns [1] to [4]). From these regressions, we selected the models with the best fit, either linear or quadratic in each case, using the Akaike Information Criterion (AIC) and report them in summary **Table 2**. In addition, we ran similar

regressions in which both predictors (linear and quadratic terms) are included simultaneously (columns [5] and [6] in Tables S2–S6) in order to test for possible mediation or confounding effects. The interaction between CRT and Raven scores is never significant in predicting creativity (all  $p$ 's  $> 0.3$ ) and is thus not reported in the tables for the sake of brevity. The results remain qualitatively similar if we also control for gender and age.

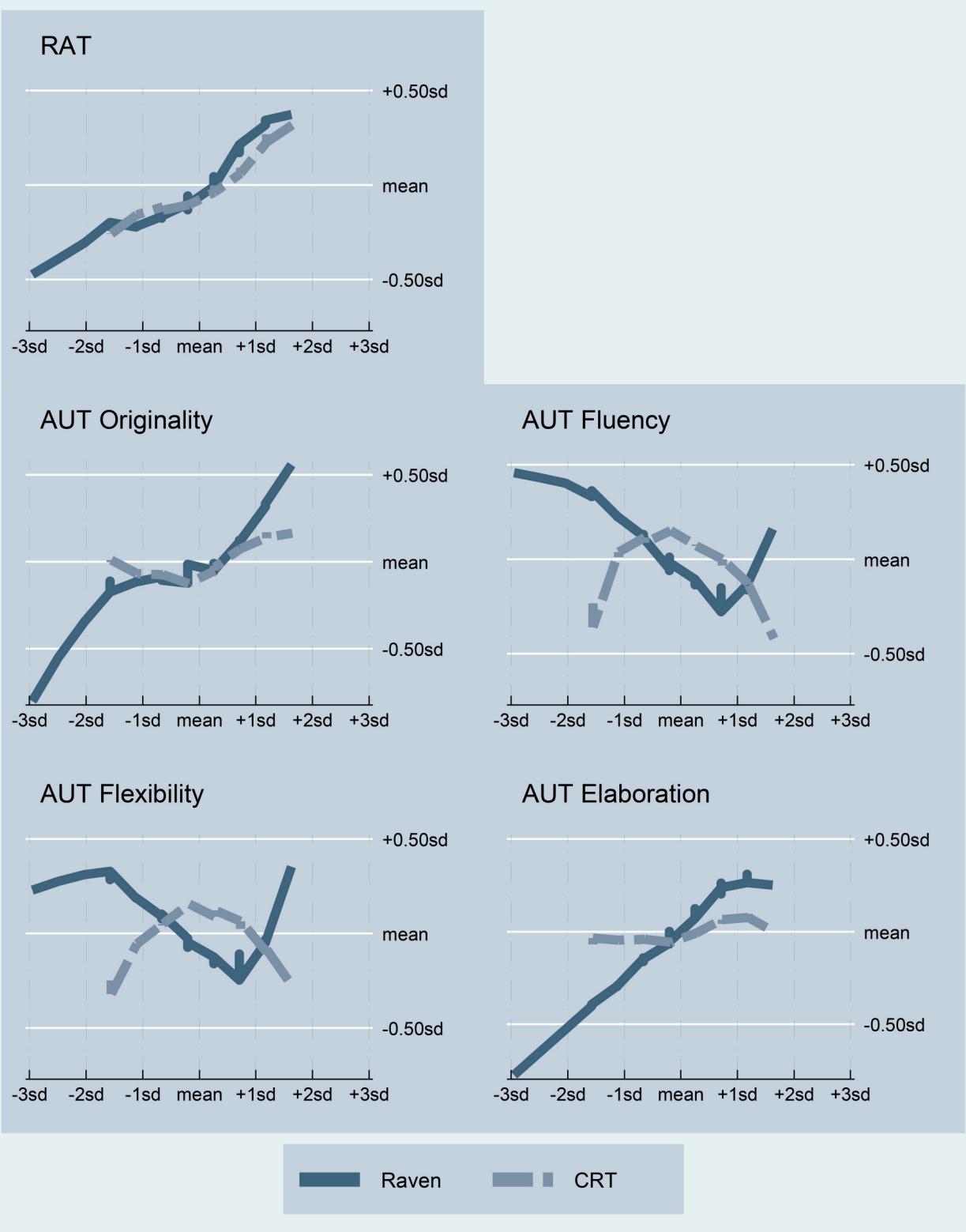
The models with the best fit (**Table 2**) report a positive linear relationship of convergent thinking (RAT) with both Raven ( $p < 0.01$ ) and CRT scores ( $p = 0.03$ ), which is consistent with the positive and significant correlations reported in the previous section. Effect sizes are substantial: in both cases, one SD increase in the predictor is associated with about 20% of one SD increase in RAT (0.22 and 0.17 for Raven and CRT, respectively; see coefficients in **Table 2**). Interestingly, the effect of Raven on RAT remains significant ( $p = 0.02$ ) if we include both Raven and CRT scores as predictors (see column [5] in Table S2) whereas the effect of CRT becomes non-significant ( $p = 0.15$ ). This result suggests that the significant effect of CRT scores on convergent thinking is driven more by cognitive ability (basic computational skills are also necessary for solving the CRT correctly) rather than by reflectiveness.

The relationship between our cognitive measures and divergent thinking is more complex. The models with the best fit report a linear and significant relationship between cognitive ability and all the measures of divergent thinking (all  $p$ 's  $< 0.03$ ), except for *flexibility* ( $p = 0.22$ ; see **Table 2**). Subjects with a higher Raven score tend to generate less uses (lower *fluency*), although these are more elaborated and original. Again, for these three creativity measures, one SD increase in Raven produces a variation in the dependent variable of about 20% of one SD. The effect of Raven on *flexibility* appears to be slightly U-shaped in **Figure 1** but the regressions do not report any significant linear or quadratic relationship (all  $p$ 's  $> 0.22$ ; see columns [1] and [2] in Table S5). As shown in columns [5] and [6] of Tables S3–S6, the effect of Raven on the divergent thinking measures remains virtually identical when controlling for CRT, which indicates that cognitive reflection does not mediate any of these relationships.

**TABLE 1 | Descriptive statistics and Pearson correlations.**

	Mean	Std Dev	[1]	[2]	[3]	[4a]	[4b]	[4c]	[4d]
<b>COGNITIVE MEASURES</b>									
1. Raven	14.43	2.30	–						
2. CRT	3.53	2.26	0.26**	–					
<b>CREATIVITY</b>									
3. RAT	3.69	2.97	0.23**	0.17*	–				
4. AUT									
4.a. Originality	1.33	0.54	0.20*	0.09	0.14	–			
4.b. Fluency	16.47	8.90	–0.17*	–0.06	–0.06	–0.25**	–		
4.c. Flexibility	11.17	4.22	–0.10	–0.01	–0.02	–0.09	0.85***	–	
4.d. Elaboration	0.23	0.29	0.26**	0.06	0.10	0.37***	–0.36***	–0.31***	–

$N = 150$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .



**FIGURE 1 | Relationship between cognitive measures and creative thinking.** The relationships are represented using locally weighted smoothing (LOWESS) techniques. All variables are standardized.

**TABLE 2 |** The effect of cognitive abilities and cognitive styles on creativity (best fitting models).

	RAT		AUT Originality		AUT Fluency		AUT Flexibility		AUT Elaboration	
	[Raven]	[CRT]	[Raven]	[CRT]	[Raven]	[CRT]	[Raven]	[CRT]	[Raven]	[CRT]
Raven <sub>std</sub>	0.219*** (0.076)		0.195*** (0.071)		-0.162** (0.074)		-0.099 (0.081)		0.246*** (0.067)	
Raven <sub>std</sub> <sup>2</sup>										
CRT <sub>std</sub>		0.169** (0.076)		0.089 (0.082)		-0.043 (0.066)		0.011 (0.073)		0.053 (0.082)
CRT <sub>std</sub> <sup>2</sup>						-0.207*** (0.076)		-0.194** (0.079)		
Constant	0.001 (0.080)	-0.010 (0.080)	0.001 (0.080)	-0.005 (0.081)	-0.001 (0.081)	0.224 (0.136)	-0.001 (0.082)	0.206 (0.127)	0.001 (0.079)	-0.003 (0.082)
F	8.229	4.891	7.562	1.179	4.732	4.255	1.490	2.990	13.302	0.420
prob > F	0.005	0.029	0.007	0.279	0.031	0.016	0.224	0.053	0.000	0.518
R <sup>2</sup>	0.053	0.030	0.042	0.008	0.029	0.039	0.011	0.031	0.067	0.003
LI	-208.277	-210.018	-209.129	-211.706	-210.139	-209.359	-211.522	-210.000	-207.174	-212.112
AIC	420.555	424.036	422.258	427.411	424.278	424.717	427.044	426.000	418.348	428.225

OLS estimates. N = 150. All variables are standardized. Robust standard errors are shown in parentheses. See Tables S2–S6 for alternative specifications. \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001.

Contrary to the results observed with Raven, we do not find any significant linear relationship between cognitive styles and divergent thinking (all p's > 0.28; see column [3] in Tables S3–S6). These results hold when we control for Raven (all p's > 0.63; see column [5] in Tables S3–S6). However, we find a significant inverted U-shape relationship of CRT with both *fluency* and *flexibility*, as reported in **Table 2** ( $p < 0.01$  and  $p = 0.02$ , respectively). Subjects with an average level of cognitive reflection tend to produce more answers and use more categories than those subjects characterized by either a more intuitive or a more reflective cognitive style. Moreover, the fact that the coefficient of the linear term in the quadratic regression specification is not significantly different from zero in either case ( $p = 0.52$  and  $p = 0.88$ , respectively) indicates that the maximum levels of fluency and flexibility are observed at the mean CRT score, as suggested by **Figure 1**. Effect sizes are comparable to those reported above insofar as, in both cases, moving one SD either above or below the mean CRT is associated with a decrease of about 20% of one SD in the dependent variable. Yet, the effects are larger for more extreme CRT values. Note that half of the observations fall outside the range mean  $\pm$  one SD (see also Figure S1). Controlling for Raven does not alter these relationships ( $p = 0.01$  and  $p = 0.02$ , respectively; see column [6] in Tables S4, S5), which again indicates an absence of mediation effects.

## DISCUSSION

The dual-process approach of cognition has been recently suggested to reconcile previous conflictive findings on the relationship between creativity and executive cognition (Allen and Thomas, 2011; Ball et al., 2015; Barr et al., 2015;

Sowden et al., 2015). We contribute to this literature by differentiating between the algorithmic and reflective minds (Evans and Stanovich, 2013), and by analyzing their separate effects on convergent thinking and four different dimensions of divergent thinking. We partially replicate the results of Barr et al. (2015) by finding that individuals' ability to make remote associations correlates positively with cognitive ability and cognitive reflection. However, we find that this effect on convergent thinking is mainly driven by cognitive ability. Similarly to Barr et al. (2015), we also find that higher levels of cognitive ability are related with higher originality scores and lower fluency scores in divergent thinking. Unlike Barr et al. (2015), we also analyze non-linear effects and find an inverted U-shape relationship between cognitive reflection and our measures of flexibility and fluency on the divergent thinking task. These new results suggest that individuals who are highly deliberative may have a disadvantage in producing a large number of new and creative ideas.

Dual-process models of creativity suggest that both generative and evaluative processes interact during the creative process (Finke et al., 1992; Basadur, 1995; Howard-Jones, 2002; Gabora, 2005; Nijstad et al., 2010; Gabora and Ranjan, 2013). Although these models do not have a straightforward mapping onto dual-process models of cognition, the interaction between Type 1 and Type 2 cognitive processes may play a different role in different phases of the creative process. In this line, Sowden et al. (2015) call for future research "... to investigate the extent to which creativity is determined by the ability to shift between Type 1 and Type 2 thinking processes as a function of the circumstances and the stage of the creative processes" (p. 55). Our results suggest that cognitive reflection, that is the disposition to override automatic responses related to Type 1 processing and engage in Type 2

controlled thought, has a complex effect on divergent thinking. To some extent, cognitive reflection may be necessary to shift between the generative and evaluative processes involved in the production of new ideas. However, individuals characterized by high levels of reflection may be less able to rely on their intuitive, autonomous mind which can also be needed for unleashing one's creative power (e.g., Dorfman et al., 1996; Norris and Epstein, 2011; Jarosz et al., 2012).

The finding of an inverted U-shape relationship between cognitive reflection (and, analogously, intuitive processing) and creativity is consistent with recent advances on the "mad genius hypothesis": mild levels of top-down control dysfunction may be beneficial for creativity but severe impairment leads to poor creative performance (for a review, see Abraham, 2014).

Relatedly, neuropsychological research has shown an inverted-U shape relationship between spontaneous eye blink rates and *flexibility* in divergent creative thinking tasks (Chermahini and Hommel, 2010). To the extent that eye blink rates reflect dopaminergic activity (Karson, 1983), which is in turn linked to inhibitory control (Cohen and Servan-Schreiber, 1992), our results are in line with the finding of Chermahini and Hommel (2010).

Beyond its connection to basic cognitive research, our findings offer insights to managers in search for the creative talent of millennials. One essential implication of our study is that thinking too much may hamper important aspects of divergent creative thinking. This result is of primary relevance to hiring managers who may want to rely on cognitive reflection as the main criterion to recruit diligent (Corgnet et al., 2015b) and creative millennials. Our findings suggest that the cognitive tests used to recruit workers have to be adapted to the nature of the job offered. For example, recruiting for jobs that fundamentally require finding well-defined solutions to problems (such as accounting or actuarial jobs) can rely on a mix of cognitive ability and reflection tests which are good predictors of convergent creative thinking and diligence. However, recruiting for jobs that mainly require divergent creative thinking (such as marketing, industrial design, or psychology jobs) should not solely rely on cognitive measures. Recruiting based on cognitive reflection skills may actually prevent the hire of highly creative workers. These recommendations are becoming increasingly relevant as a growing number of jobs in modern economies require divergent creative thinking (Pink, 2005).

The current research has some necessary limitations that future research might remedy. To keep focus our study uses only one measure of fluid intelligence (Raven) and a single measure of cognitive style (CRT). Future research may assess the robustness of our findings to other measures of fluid intelligence and cognitive style, possibly extending the analysis to include crystallized intelligence. Also, our sample consisted entirely of undergraduates, with a limited age, education, and

income range. Although this was a methodological choice that allowed us to study the workforce of the future, further studies may assess the robustness of our findings to different populations. Regarding our creativity measures, future research may attempt to extend our analysis to the case of practical creative tasks that are commonly encountered, for example, at the workplace. To that end, future research may embed the study of creativity in an organizational setting that allows for studying the relationship between workplace problem solving and cognitive skills.

On a methodological note, we used a fixed ordering of which may have influenced the results as, among other factors, fatigue may interfere with test results. While the 2-min break in the middle of the experiment might have mitigated spillover effects between the first and the second part of the experiment, concerns still remain. We encourage future research to explore possible ordering effects. In addition, future research focusing on state-level analyses of the role of intuition vs. reflection in creative performance is necessary to assess the robustness (and causality) of our trait-level findings as well as deepen our understanding of the cognitive basis of creativity. Along these lines, it would be interesting for future research to test the effect of cognitive manipulations such as cognitive load, ego depletion, priming, or time pressure/delay on creative performance. Our findings suggest that future research on the topic should attempt to capture potentially non-linear effects thus elaborating experimental designs that allow such effects to materialize. This can be done, for example, by considering at least three levels per treatment condition.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

The authors acknowledge financial support from the International Foundation for Research in Experimental Economics, the Argyros School of Business and Economics at Chapman University, the Spanish Ministry of Education [Grant 2012/00103/001], Ministry of Economy and Competence [2016/00122/001], Spanish Plan Nacional I+D MCI [ECO2013-44879-R], 2014-17, and Proyectos de Excelencia de la Junta Andalucía [P12.SEJ.1436], 2014-18.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01626>

Abraham, A., Windmann, S., McKenna, P., and Güntürkün, O. (2007). Creative thinking in schizophrenia: the role of executive dysfunction and symptom severity. *Cogn. Neuropsychiatry* 12, 235–258. doi: 10.1080/13546800601046714

## REFERENCES

- Abraham, A. (2014). Is there an inverted-U relationship between creativity and psychopathology? *Front. Psychol.* 5:750. doi: 10.3389/fpsyg.2014.00750

- Acar, S., and Sen, S. (2013). A multilevel meta-analysis of the relationship between creativity and schizotypy. *Psychol. Aesthetics Creativity Arts* 7:214. doi: 10.1037/a0031975
- Allen, A. P., and Thomas, K. E. (2011). A dual process account of creative thinking. *Creativity Res. J.* 23, 109–118. doi: 10.1080/10400419.2011.571183
- Ball, L. J., Marsh, J. E., Litchfield, D., Cook, R. L., and Booth, N. (2015). When distraction helps: evidence that concurrent articulation and irrelevant speech can facilitate insight problem solving. *Thinking Reasoning* 21, 76–96. doi: 10.1080/13546783.2014.934399
- Barr, N., Pennycook, G., Stoltz, J. A., and Fugelsang, J. A. (2015). Reasoned connections: a dual-process perspective on creative thought. *Thinking Reasoning* 21, 61–75. doi: 10.1080/13546783.2014.895915
- Basadur, M. S. (1995). Optimal ideation-evaluation ratios. *Creativity Res. J.* 8, 63–75. doi: 10.1207/s15326934crj0801\_5
- Batey, M., and Furnham, A. (2006). Creativity, intelligence, and personality: a critical review of the scattered literature. *Genet. Soc. Gen. Psychol. Monogr.* 132, 355–429. doi: 10.3200/MONO.132.4.355-430
- Beatty, R. E., and Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychol. Aesthetics Creativity Arts* 6:309. doi: 10.1037/a0029171
- Bosch-Domènec, A., Brañas-Garza, P., and Espín, A. M. (2014). Can exposure to prenatal sex hormones (2D: 4D) predict cognitive reflection? *Psychoneuroendocrinology* 43, 1–10. doi: 10.1016/j.psyneuen.2014.01.023
- Brañas-Garza, P., García-Muñoz, T., and Hernán-González, R. (2012). Cognitive effort in the beauty contest game. *J. Econ. Behav. Organ.* 83, 254–260. doi: 10.1016/j.jebo.2012.05.018
- Capraro, V., Corgnet, B., Espín, A. M., and Hernán-González, R. (2016). Deliberation favors social efficiency by helping people disregard their relative shares: evidence from US and India. Available online at: SSRN: <http://ssrn.com/abstract=2799850>
- Chermahini, S. A., and Hommel, B. (2010). The (b) link between creativity and dopamine: spontaneous eye blink rates predict and dissociate divergent and convergent thinking. *Cognition* 115, 458–465. doi: 10.1016/j.cognition.2010.03.007
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.* 74, 829–836. doi: 10.1080/01621459.1979.10481038
- Cleveland, W. S., and McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science* 229, 828–833. doi: 10.1126/science.229.4716.828
- Cohen, J. D., and Servan-Schreiber, D. (1992). Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. *Psychol. Rev.* 99, 45–77.
- Cokely, E. T., and Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: a protocol analysis and process model evaluation. *Judgment Decision Making* 4:20.
- Corgnet, B., Espín, A. M., and Hernán-González, R. (2015a). The cognitive basis of social behavior: cognitive reflection overrides antisocial but not always prosocial motives. *Front. Behav. Neurosci.* 9:287. doi: 10.3389/fnbeh.2015.00287
- Corgnet, B., Espín, A. M., Hernán-González, R., Kujal, P., and Rassenti, S. (2016). To trust, or not to trust: cognitive reflection in trust games. *J. Behav. Exp. Econ.* 64, 20–27. doi: 10.1016/j.soec.2015.09.008
- Corgnet, B., Hernán-González, R., and Mateo, R. (2015b). Cognitive reflection and the diligent worker: an experimental study of millennials. *PloS ONE* 10:e0141243. doi: 10.1371/journal.pone.0141243
- Dilchert, S., Ones, D. S., Davis, R. D., and Rostow, C. D. (2007). Cognitive ability predicts objectively measured counterproductive work behaviors. *J. Appl. Psychol.* 92:616. doi: 10.1037/0021-9010.92.3.616
- Dorfman, J., Shames, V. A., and Kihlstrom, J. F. (1996). “Intuition, incubation, and insight: implicit cognition in problem solving,” in *Implicit Cognition*, ed G. Underwood (Oxford: Oxford University Press), 257–296.
- Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.* 59, 255–278. doi: 10.1146/annurev.psych.59.103006.093629
- Evans, J. S. B. T. (2009). “How many dual-process theories do we need? One, two, or many?” in *In Two Minds: Dual Processes and Beyond*, eds Evans, J. S. B. T. and K. Frankish (New York, NY: Oxford University Press), 33–55.
- Evans, J. S. B. T., and Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspect. Psychol. Sci.* 8, 223–241. doi: 10.1177/1745691612460685
- Eysenck, H. J. (1993). Creativity and personality: suggestions for a theory. *Psychol. Inquiry* 4, 147–178. doi: 10.1207/s15327965pli0403\_1
- Finke, R. A., Ward, T. B., and Smith, S. M. (1992). *Creative Cognition: Theory, Research and Applications*. Cambridge, MA: MIT Press.
- Frederick, S. (2005). Cognitive reflection and decision making. *J. Econ. Perspect.* 19, 25–42. doi: 10.1257/089533005775196732
- Gabora, L. (2005). Creative thought as a non-Darwinian evolutionary process. *J. Creat. Behav.* 39, 262–283. doi: 10.1002/j.2162-6057.2005.tb01261.x
- Gabora, L., and Ranjan, A. (2013). “How insight emerges in distributed, content-addressable memory,” in *The Neuroscience of Creativity*, eds A. Bristol, O. Vartanian, and J. Kaufman (New York, NY: Oxford University Press), 19–44.
- Getzels, J. W., and Jackson, P. W. (1962). *Creativity and Intelligence: Explorations with Gifted Students*. New York, NY: Wiley.
- Gilhooly, K. J., Fioratou, E., Anthony, S. H., and Wynn, V. (2007). Divergent thinking: strategies and executive involvement in generating novel uses for familiar objects. *Br. J. Psychol.* 98, 611–625. doi: 10.1111/j.2044-8295.2007.tb00467.x
- Guilford, J. P. (1967). *The Nature of Human Intelligence*. New York, NY: McGraw-Hill.
- Hocevar, D. (1979). Ideational fluency as a confounding factor in the measurement of originality. *J. Educ. Psychol.* 71:191. doi: 10.1037/0022-0663.71.2.191
- Howard-Jones, P. A. (2002). A dual-state model of creative cognition for supporting strategies that foster creativity in the classroom. *Int. J. Technol. Design Educ.* 12, 215–226. doi: 10.1023/A:1020243429353
- Hunter, J. E., and Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychol. Bull.* 96, 72–98. doi: 10.1037/0033-2909.96.1.72
- Jaeggi, S. M., Studer-Luethi, B., Buschkuhl, M., Su, Y. F., Jonides, J., and Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning – implications for training and transfer. *Intelligence* 38, 625–635. doi: 10.1016/j.intell.2010.09.001
- Jaracz, J., Patrzala, A., and Rybakowski, J. K. (2012). Creative thinking deficits in patients with schizophrenia: neurocognitive correlates. *J. Nerv. Ment. Dis.* 200, 588–593. doi: 10.1097/NMD.0b013e31825fc49
- Jarosz, A. F., Colflesh, G. J. H., and Wiley, J. (2012). Uncorking the muse: alcohol intoxication facilitates creative problem solving. *Conscious. Cogn.* 21, 487–493. doi: 10.1016/j.concog.2012.01.002
- Jauk, E., Benedek, M., Dunst, B., and Neubauer, A. C. (2013). The relationship between intelligence and creativity: new support for the threshold hypothesis by means of empirical breakpoint detection. *Intelligence* 41, 212–221. doi: 10.1016/j.intell.2013.03.003
- Kahneman, D., and Frederick, S. (2007). Frames and brains: elicitation and control of response tendencies. *Trends Cogn. Sci.* 11, 45–46. doi: 10.1016/j.tics.2006.11.007
- Karson, C. N. (1983). Spontaneous eye-blink rates and dopaminergic systems. *Brain* 106, 643–653.
- Kaufman, J. C. (2009). *Creativity 101*. New York, NY: Springer Publishing Company.
- Kim, K. H. (2005). Can only intelligent people be creative? A meta-analysis. *J. Secondary Gifted Educ.* 16, 57–66. doi: 10.4219/jsgc-2005-473
- Kim, K. H., Cramond, B., and VanTassel-Baska, J. (2010). “The relationship between creativity and intelligence,” in *The Cambridge Handbook of Creativity*, eds J. C. Kaufman and R. J. Sternberg (New York, NY: Cambridge University Press), 395–412.
- Kim, S., Hasher, L., and Zacks, R. T. (2007). Aging and a benefit of distractibility. *Psychonomic Bull. Rev.* 14, 301–305. doi: 10.3758/BF03194068
- Kovacevic, S., Azma, S., Irimia, A., Sherfey, J., Halgren, E., and Marinkovic, K. (2012). Theta oscillations are sensitive to both early and late conflict processing stages: effects of alcohol intoxication. *PloS ONE* 7:e43957. doi: 10.1371/journal.pone.0043957
- Lohse, J. (2016). Smart or selfish—when smart guys finish nice. *J. Behav. Exp. Econ.* 64, 28–40. doi: 10.1016/j.soec.2016.04.002
- Marinkovic, K., Rickenbacher, E., Azma, S., and Artsy, E. (2012). Acute alcohol intoxication impairs top-down regulation of stroop incongruity as revealed by blood oxygen level-dependent functional magnetic

- resonance imaging. *Human Brain Map.* 33, 319–333. doi: 10.1002/hbm.21213
- Mednick, S. A. (1962). The associative basis of the creative process. *Psychol. Rev.* 69, 220. doi: 10.1037/h0048850
- Murphy, K. R. (1989). Is the relationship between cognitive ability and job performance stable over time? *Human Perform.* 2, 183–200. doi: 10.1207/s15327043hup0203\_3
- Nijstad, B. A., De Dreu, C. K. W., Rietzschel, E. F., and Baas, M. (2010). The dual pathway to creativity model: creative ideation as a function of flexibility and persistence. *Eur. Rev. Soc. Psychol.* 21, 34–77. doi: 10.1080/10463281003765323
- Norris, P., and Epstein, S. (2011). An experiential thinking style: its facets and relations with objective and subjective criterion measures. *J. Personal.* 79, 1043–1080. doi: 10.1111/j.1467-6494.2011.00718.x
- Nusbaum, E. C., and Silvia, P. J. (2011). Are intelligence and creativity really so different? Fluid intelligence, executive processes, and strategy use in divergent thinking. *Intelligence* 39, 36–45. doi: 10.1016/j.intell.2010.11.002
- Oechssler, J., Roider, A., and Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *J. Econ. Behav. Organ.* 72, 147–152. doi: 10.1016/j.jebo.2009.04.018
- Olea, M. M., and Ree, M. J. (1994). Predicting pilot and navigator criteria: not much more than g. *J. Appl. Psychol.* 79, 845–851. doi: 10.1037/0021-9010.79.6.845
- Paxton, J. M., Unger, L., and Greene, J. D. (2012). Reflection and reasoning in moral judgement. *Cogn. Sci.* 36, 163–177. doi: 10.1111/j.1551-6709.2011.01210.x
- Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., and Fugelsang, J. A. (2014). The role of analytic thinking in moral judgements and values. *Thinking Reasoning* 20, 188–214. doi: 10.1080/13546783.2013.865000
- Peterson, J. B., Rothfleisch, J., Zelazo, P. D., and Pihl, R. O. (1990). Acute alcohol intoxication and cognitive functioning. *J. Stud. Alcohol* 51, 114–122. doi: 10.15288/jsa.1990.51.114
- Pink, D. (2005). *A Whole New Mind.* New York City, NY: Riverhead Books.
- Rainer, T. S., and Rainer, J. W. (2011). *The Millennials: Connecting to America's Largest Generation.* Nashville, TN: BandH Publishing Group.
- Raven, J. C. (1936). *Mental Tests used in Genetic Studies: The Performances of Related Individuals in Tests Mainly Educative and Mainly Reproductive.* Unpublished master's thesis, University of London.
- Ricks, T. R., Turley-Ames, K. J., and Wiley, J. (2007). Effects of working memory capacity on mental set due to domain knowledge. *Mem. Cogn.* 35, 1456–1462. doi: 10.3758/BF03193615
- Runcio, M. A. (2007). *Creativity.* San Diego, CA: Acade.
- Sawyer, R. K. (2006). *Explaining Creativity: The Science of Human Innovation.* New York, NY: Oxford University Press.
- Schmidt, F. L. (2009). “Select on intelligence,” in *Handbook of Principles of Organizational Behavior: Indispensable Knowledge for Evidence-Based Management*, ed E. A. Locke (New York, NY: John Wiley and Sons, Ltd.), 3–18.
- Schmidt, F. L., Hunter, J. E., and Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *J. Appl. Psychol.* 71:432. doi: 10.1037/0021-9010.71.3.432
- Schuldberg, D. (2005). Eysenck Personality Questionnaire scales and paper-and-pencil tests related to creativity. *Psychol. Rep.* 97, 180–182. doi: 10.2466/pr0.97.5.180-182
- Silvia, P. J. (2008a). Another look at creativity and intelligence: exploring higher-order models and probable confounds. *Pers. Individ. Diff.* 44, 1012–1021. doi: 10.1016/j.paid.2007.10.027
- Silvia, P. J. (2008b). Creativity and intelligence revisited: a latent variable analysis of Wallach and Kogan (1965). *Creativity Res. J.* 20, 34–39. doi: 10.1080/10400410701841807
- Silvia, P. J. (2015). Intelligence and creativity are pretty similar after all. *Educ. Psychol. Rev.* 27, 599–606. doi: 10.1007/s10648-015-9299-1
- Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., et al. (2008). Assessing creativity with divergent thinking tasks: exploring the reliability and validity of new subjective scoring methods. *Psychol. Aesthetics Creativity Arts* 2, 68–85. doi: 10.1037/1931-3896.2.2.68
- Sowden, P. T., Pringle, A., and Gabora, L. (2015). The shifting sands of creative thinking: connections to dual-process theory. *Thinking Reasoning* 21, 40–60. doi: 10.1080/13546783.2014.885464
- Stanovich, K. E. (2009). “Distinguishing the reflective, algorithmic, and autonomous minds: is it time for a tri-process theory,” in *In Two Minds: Dual Processes and Beyond*, eds J. S. B. T. Evans and K. Frankish (New York, NY: Oxford University Press), 55–88.
- Stanovich, K. E. (2010). *Rationality and the Reflective Mind.* Oxford University Press.
- Toplak, M. E., West, R. F., and Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics and biases tasks. *Mem. Cogn.* 39, 1275–1289. doi: 10.3758/s13421-011-0104-1
- Toplak, M. E., West, R. F., and Stanovich, K. E. (2014). Assessing miserly information processing: an expansion of the cognitive reflection test. *Thinking Reasoning* 20, 147–168. doi: 10.1080/13546783.2013.844729
- Troyer, A. K., and Moscovitch, M. (2006). “Cognitive processes of verbal fluency tasks,” in *The Quantified Process Approach to Neuropsychological Assessment*, ed A. M. Poreh (New York, NY: Taylor and Francis Group), 143–157.
- Wallach, M. A., and Kogan, N. (1965). *Modes of Thinking in Young Children: A Study of the Creativity-Intelligence Distinction.* New York, NY: Holt, Rinehart and Winston.
- Weisberg, R. W. (2006). *Creativity: Understanding Innovation in Problem Solving, Science, Invention, and the Arts.* Hoboken, NJ: John Wiley and Sons.
- Wiley, J., and Jarosz, A. F. (2012a). “How working memory capacity affects problem solving,” in *Psychology of Learning and Motivation*, Vol. 56, ed B. H. Ross (Burlington, MA: Academic Press), 185–228.
- Wiley, J., and Jarosz, A. F. (2012b). Working memory capacity, attentional focus, and problem solving. *Curr. Direct. Psychol. Sci.* 21, 258–262. doi: 10.1177/0963721412447622

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Corngot, Espín and Hernán-González. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Gender Differences in Performance Predictions: Evidence from the Cognitive Reflection Test

Patrick Ring<sup>1</sup>, Levent Neyse<sup>1\*</sup>, Tamas David-Barett<sup>2</sup> and Ulrich Schmidt<sup>1,3,4</sup>

<sup>1</sup> Social and Behavioral Approaches to Global Problems, Kiel Institute for the World Economy, Kiel, Germany, <sup>2</sup> Medical Sciences Division, Department of Experimental Psychology, University of Oxford, Oxford, UK, <sup>3</sup> Department of Economics, University of Kiel, Kiel, Germany, <sup>4</sup> Department of Economics and Econometrics, University of Johannesburg, Johannesburg, South Africa

This paper studies performance predictions in the 7-item Cognitive Reflection Test (CRT) and whether they differ by gender. After participants completed the CRT, they predicted their own (i), the other participants' (ii), men's (iii), and women's (iv) number of correct answers. In keeping with existing literature, men scored higher on the CRT than women and both men and women were too optimistic about their own performance. When we compare gender-specific predictions, we observe that men think they perform significantly better than other men and do so significantly more than women. The equality between women's predictions about their own performance and their female peers cannot be rejected. Our findings contribute to the growing literature on the underpinnings of behavior in economics and in psychology by uncovering gender differences in confidence about one's ability relative to same and opposite sex peers.

**Keywords:** overconfidence, Cognitive Reflection Test, gender difference, performance prediction, competition, intra-gender competition

## OPEN ACCESS

### Edited by:

Nikolaos Georgantzis,  
University of Reading, UK

### Reviewed by:

Marcello Sartorelli,  
University of Alicante, Spain  
Zahra Murad,  
University of Surrey, UK

### \*Correspondence:

Levent Neyse  
levent.neyse@fwi-kiel.de

### Specialty section:

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

**Received:** 30 June 2016

**Accepted:** 13 October 2016

**Published:** 01 November 2016

### Citation:

Ring P, Neyse L, David-Barett T and Schmidt U (2016) Gender Differences in Performance Predictions: Evidence from the Cognitive Reflection Test.  
*Front. Psychol.* 7:1680.

doi: 10.3389/fpsyg.2016.01680

## INTRODUCTION

Confidence is an essential personality trait with a positive impact in numerous contexts, such as subjective well-being (Taylor and Brown, 1988, 1994), professional success (Kanter, 2004), or mental health (Taylor, 1989). Overconfidence, on the other hand, is a psychological bias by definition, since it is an inaccurate judgment of one's own abilities. Typical examples for this type of bias are overly optimistic beliefs in one's professional abilities (Meyer et al., 2013) or physical fitness (Obling et al., 2015). This overly optimistic belief may be both absolute (i.e., individuals predict that their performance is better than it actually is) or relative (i.e., individuals predict that their performance is better than their peers' when it actually is not). In the literature, the first type of overconfidence is referred to as overestimation and the latter as overplacement (Moore and Healy, 2008).

Overconfident beliefs appear to have positive consequences in some contexts, while they can be detrimental in others. Among other things, it has been suggested that overconfidence has positive psychological benefits, for example, on ambition, morale, and persistence (Pajares, 1996; Johnson and Fowler, 2011). Besides potential positive psychological benefits, overconfidence seems to help individuals in a social setting by convincing others that they have better skills or knowledge than they actually have (von Hippel and Trivers, 2011). Anderson et al. (2012) have shown empirically that individuals with high levels of overconfidence are perceived as more competent by their peers. This overstatement of one's abilities could be an advantage in hiring decisions (Reuben et al., 2014).

Besides potential positive aspects, several empirical studies display the negative economic consequences of overconfidence. Camerer and Lovallo (1999), for example, have found that in a laboratory setting individuals tend to overestimate their chances of relative success and therefore excessively enter a competitive game. In a trading experiment, highly overconfident investors show less reaction to bad news, which results in lower profits for them compared to low overconfidence investors (Trinugroho and Sembel, 2011). Similarly, Barber and Odean (2001) have reported that overconfident investors reduce their net earnings by excessive trading; i.e., the expected gains from a trade do not exceed its transaction costs. Moreover, managerial overconfidence seems to explain investment distortion (Malmendier and Tate, 2005).

Despite the potential costs associated with overconfident beliefs in some settings, overconfident judgments are an integral part of various aspects of human decision making (De Bondt and Thaler, 1995). They are common in many professional fields such as investment banking (Stael von Holstein, 1972), economic negotiations (Neale and Bazerman, 1991), the law (Wagenaar and Keren, 1986), and even in clinical psychology (Oskamp, 1965). One typically observed pattern is that while both men and women are overconfident, men are more frequently prone to this bias than women (Lichtenstein et al., 1982; Lundeberg et al., 1994) and this seems to have important economic implications which will be discussed in the next subsection.

Barber and Odean (2001), for instance, have investigated the common stock investments of men and women separately. They have shown that men trade 45% more than women and this trading behavior actually reduces their earnings. They have concluded that this is likely due to greater overconfidence in men. Among other things, lower risk aversion in men can be attributed to higher overconfidence (Soll and Klayman, 2004). Furthermore, it has been shown in laboratory experiments that women are less likely to enter competition than men and lower levels of overconfidence are one explanation for this behavior (Niederle and Vesterlund, 2007; Reuben et al., 2012). It seems that women are disadvantaged in hiring decisions, because underconfident women may not appear as competent as their male peers (Reuben et al., 2014).

While the general tendency of men being more overconfident than women has been reported in several studies, less is known about the causes of this difference. This paper presents an experimental assessment of the extent to which this bias is driven by gender differences in confidence about one's ability relative to same and opposite sex peers. Thereby, the paper extends the current literature (e.g., Dean and Ortoleva, 2015) on overconfidence by using gender-specific questions. This appears relevant from an economic perspective, as the composition of one's potential competitors is important for individual decisions on whether to enter a competitive game (Datta Gupta et al., 2013). Beliefs about one's self and the others have been shown to be important drivers for this decision (Camerer and Lovallo, 1999). Similar findings have been reported in evolutionary biology. In the course of human evolution, competition among men typically took place as direct and aggressive contests. Competition among women, by contrast, was typically more indirect and subtle

(Stockley and Campbell, 2013). One potential explanation for these different types of behavior could be that the attractiveness of direct intra-gender competition is different for men and women, as they have different perceptions about their same-sex peers. Recent studies have applied evolutionary theory to explain decision-making patterns and this paper extends the literature to overconfidence. For example, it has been hypothesized that men, who face a higher sexual selection pressure than women (Trivers, 1972), should be more concerned about relative outcomes. Women, by contrast, should be more concerned about absolute outcomes, i.e., about resources for themselves and their offspring (Buss, 1989; Ermer et al., 2008). Following the predictions of this hypothesis, Schmidt et al. (2015) and Friedl et al. (2016) have shown that social comparison has a greater effect on men than on women in decision-making under risk and ambiguity.

In order to study gender differences in confidence about one's ability relative to same and opposite sex peers, participants of this study first solved the 7-item Cognitive Reflection Test (CRT). Then they predicted their own (i), the other participants' (ii), men's (iii), and women's (iv) number of correct answers in this task<sup>1</sup>. It was found that men perform better than women on this particular task, a result that has been previously reported (Kahneman and Frederick, 2002; Frederick, 2005). Moreover, it was observed that both men and women overestimate their performance; yet no significant gender effects in overestimation were found. When gender-specific predictions were compared, it emerged that men think they perform significantly better than other men. The equality between women's predictions about their own performance and their female peers cannot be rejected.

## MATERIALS AND METHODS

### Participants

Participants of the study were undergraduate students at Kiel University ( $N = 131$ ; 72 women; mean age = 24.7). The experiment was organized and recruited with the software hroot (Bock et al., 2014). The participants were randomly seated in a classroom in groups of 15. They first read the general instructions for the experiment themselves; then the instructions were read out loud. After the protocol was completed, they were invited to a separate room to get paid anonymously. The protocol also included a short questionnaire on life satisfaction questions and digit ratio measurement. Evidence obtained on the relation between overconfidence scores and digit ratios from this experiment is reported in Neyse et al. (2016).

### 7-item CRT

The 7-item CRT (Toplak et al., 2014) is an extended version of the original 3-item CRT (Frederick, 2005) that includes four additional questions. The CRT is designed to observe participants' ability to activate the Type 2 cognitive process instead of giving intuitive and effortless answers through the

<sup>1</sup>Note that we use the word *predict* to refer to participants' guesses throughout the article, which should not be confused with the term *prediction* in econometrics.

Type 1 cognitive process. According to the dual process theories of cognition (Kahneman and Frederick, 2002), the Type 1 cognitive process yields to intuitive and automatic reasoning, while the Type 2 process requires more thorough thinking and conscientiousness. The first question of the CRT is as follows:

*"A bat and a ball together cost \$1.10. A bat costs \$1 more than a ball. How much does a ball cost?"*

The intuitive, but incorrect, answer is 10 cents. The correct answer is 5 cents.

## Performance Predictions

Participants first received the 7-item CRT, which they had to complete within 10 min. After 10 min, the answer sheets were collected. This way, participants were prevented from making any changes on the answer sheets, since their predictions were also incentivized. Following the CRT, they were given another sheet on which they were asked to predict their own number of correct answers (i), the average number of correct answers of other participants in their group (ii), the men in their group (iii), and the women in their group (iv). For each correct answer in the CRT, the participants were paid €0.5. Correct predictions about their own score and others' scores were rewarded with €2 and false predictions with nothing. Gender-specific predictions were not incentivized<sup>2</sup>. The prediction task was not announced beforehand in order to avoid strategic behavior in answering the 7-item CRT itself. Participants used pen and paper to answer both the CRT-questions and the prediction task. Instructions for the experiment can be found in the Supplementary Material.

## Ethics Statement

All participants of the experiment were informed about the content and the protocol of the study before participation. Their anonymity was preserved by assigning them a randomly generated code that cannot be associated with any personal information or decision. As is standard in economics experiments, no ethical concerns were involved other than preserving the anonymity of the participants. The whole protocol was performed in accordance with the Declaration of Helsinki and conformed to the ethical guidelines of the Kiel University Experimental Economics Lab, where it was approved by the lab manager.

## RESULTS

### Summary Statistics

**Figure 1** presents the means of actual scores and predictions by gender. The participants scored 4.44 ( $SD = 1.836$ ) correct answers on average regardless of gender<sup>3</sup>. Mean number of

correct answers for men is 4.98 ( $SD = 1.892$ ) and for women 4.00 ( $SD = 1.914$ ). A two-sample Wilcoxon rank-sum test confirms that the average score of men is significantly higher than of women ( $z = -2.847, p = 0.004$ ). This is in line with previous findings in the literature (Kahneman and Frederick, 2002; Frederick, 2005; Cueva et al., 2016).

Participants predicted that they themselves had answered 5.72 ( $SD = 1.416$ ) questions correctly on average. Men predicted their own scores as 6.24 ( $SD = 1.165$ ), while women predicted their own scores as 5.29 ( $SD = 0.173$ ). A Wilcoxon rank-sum test confirms that this gender difference is significant ( $z = -4.144, p < 0.001$ ).

The overall predictions about the other participants' number of correct answers is 5.12 ( $SD = 0.966$ ) on average. Men's mean prediction is 5.22 ( $SD = 1.115$ ), while women's is 5.03 ( $SD = 0.822$ ). The two-sample Wilcoxon rank-sum test does not reject the null hypothesis of no difference ( $z = -1.564, p = 0.118$ ).

In addition to predictions about their own and other participants' performance, participants were also asked to predict the average scores of men and women in their group separately. The prediction about men's mean scores is 4.89 ( $SD = 1.109$ ) and about women's is 5.47 ( $SD = 0.998$ ) for the whole sample. Men's prediction about other men is 5.08 ( $SD = 1.204$ ) and women's prediction about men is 4.74 ( $SD = 1.007$ ). The difference is statistically significant ( $z = -2.129, p = 0.033$ ). Men predicted women's score as 5.38 ( $SD = 1.117$ ) and women's average prediction about women was 5.55 ( $SD = 0.885$ ). Non-parametric analysis does not confirm a statistically significant difference ( $z = 0.400, p = 0.690$ )<sup>4</sup>.

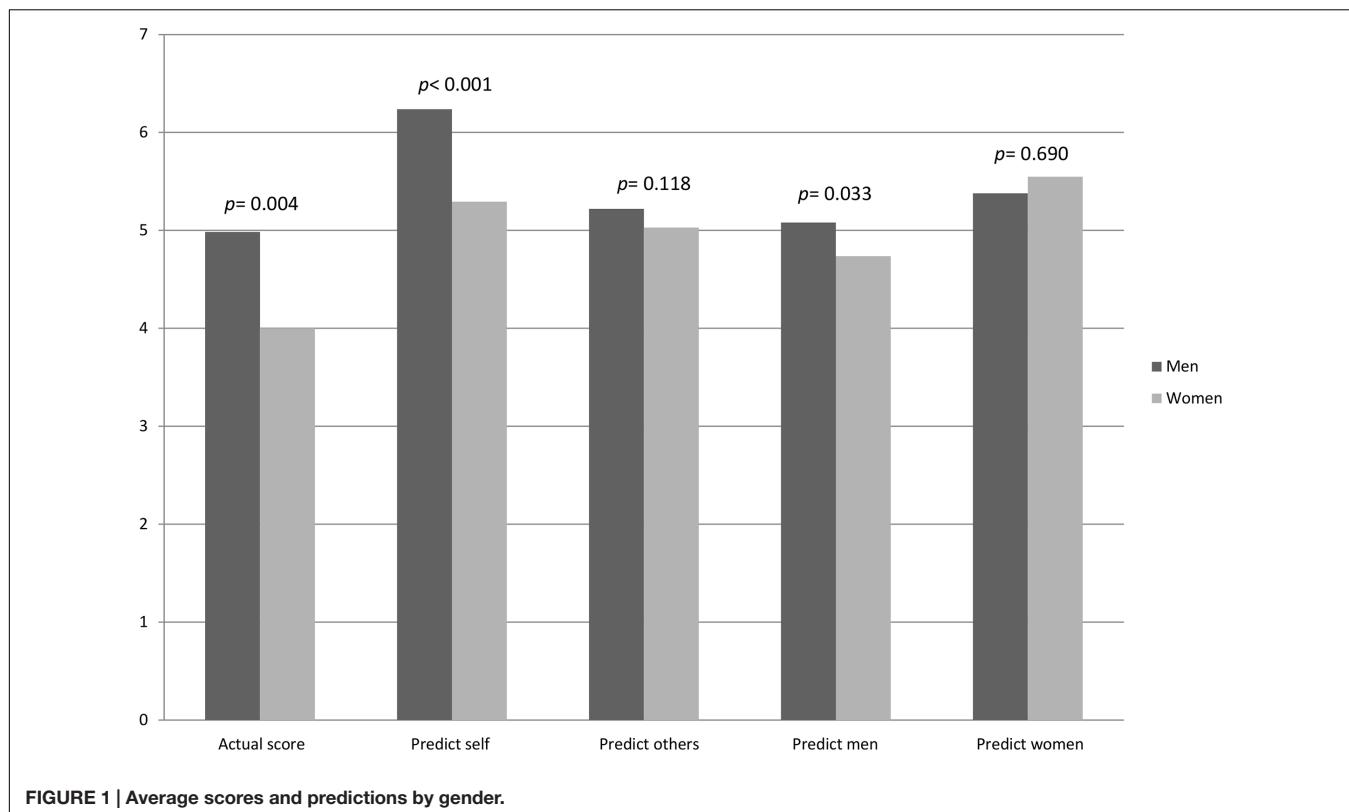
**Table 1** presents the comparison analysis of predictions. All results are gathered with Wilcoxon signed-rank tests.  $p$ -values are given for all participants as well as for men and women separately. Inequality signs to the right of each  $p$ -value indicate whether the value of the difference between the two predictions in the first column is positive, zero, or negative. Differences in means are not reported in **Table 1** as they are available in **Figure 1**. The first row shows that both men and women overestimate their scores. Their predictions about their own scores are significantly higher than their actual scores ( $p < 0.001$ ). This result is a clear indication of overestimation, which is the difference between one's actual score and prediction. The second row shows that both men and women predicted that they would do better than other participants ( $p < 0.001$  for men and  $p = 0.050$  for women). This result is an indication of overplacement. Gender-specific predictions indicate that both men and women thought they did better than men ( $p < 0.001$  for both). Yet, only men thought they did better than women ( $p < 0.001$  for men and  $p = 0.104$  for women). Finally gender-specific predictions are compared with each other. Row 5 shows that both men and women thought women would do better

<sup>2</sup>Since the gender information was gathered in a different sheet of paper at the end of the protocol and due to time constraints, we did not incentivize the gender-specific predictions. A two-sample variance comparison test did not reject the null hypothesis that the prediction variance is equal for unincentivized and incentivized predictions for both men and women at the 5% significance level.

<sup>3</sup>If we consider only the first three items as in Frederick (2005), the average number of correct answers is 1.76 ( $SD = 1.068$ ). In contrast to previous studies (Brañas-Garza et al., 2015; Cueva et al., 2016), the number of participants who

answered none of the questions correctly was rather low in our experiment. See Supplementary Figures A1 and A2 for histograms on the distribution of correct answers.

<sup>4</sup>Since there was no interaction between participants and since the performance of participants was not disclosed, group behavior cannot affect the individual behavior. Therefore, no possible reflection problem is anticipated (Manski, 1993).

**FIGURE 1 |** Average scores and predictions by gender.**TABLE 1 |** Comparisons of actual performance and predictions about others.

	All		Men		Women	
	p < 0.001	>	p < 0.001	>	p < 0.001	>
Self vs actual score						
<b>Comparisons of predictions</b>						
Self vs others	p < 0.001	>	p < 0.001	>	p = 0.050	>
Self vs men	p < 0.001	>	p < 0.001	>	p < 0.001	>
Self vs women	p = 0.023	>	p < 0.001	>	p = 0.104	-
Men vs women	p < 0.001	<	p = 0.004	<	p < 0.001	<

Self, others, men, and women denote predictions whereas actual score does not.

than men on the task ( $p < 0.001$  for women and  $p = 0.004$  for men).

## Gender-Specific Differences in Performance Predictions

The main research questions are whether there are gender-specific differences in overestimation and overplacement scores and whether such gender-specific differences can be related to participants' gender biases about other participants' performance. In order to answer them, four different variables based on participants' predictions and their actual performance were generated (Table 2).

Overestimation is the difference between one's self-prediction and actual score, and overplacement is the difference between one's self-prediction and the prediction about others regardless of gender. According to Moore and Healy (2008) overestimation

and overplacement are two aspects of overconfidence<sup>5</sup>. The intra-gender overplacement variable detects how much better or worse one thinks she/he is than the other participants with the same gender. Likewise, the inter-gender overplacement variable shows how much better or worse one thinks she/he is than participants of the other gender.

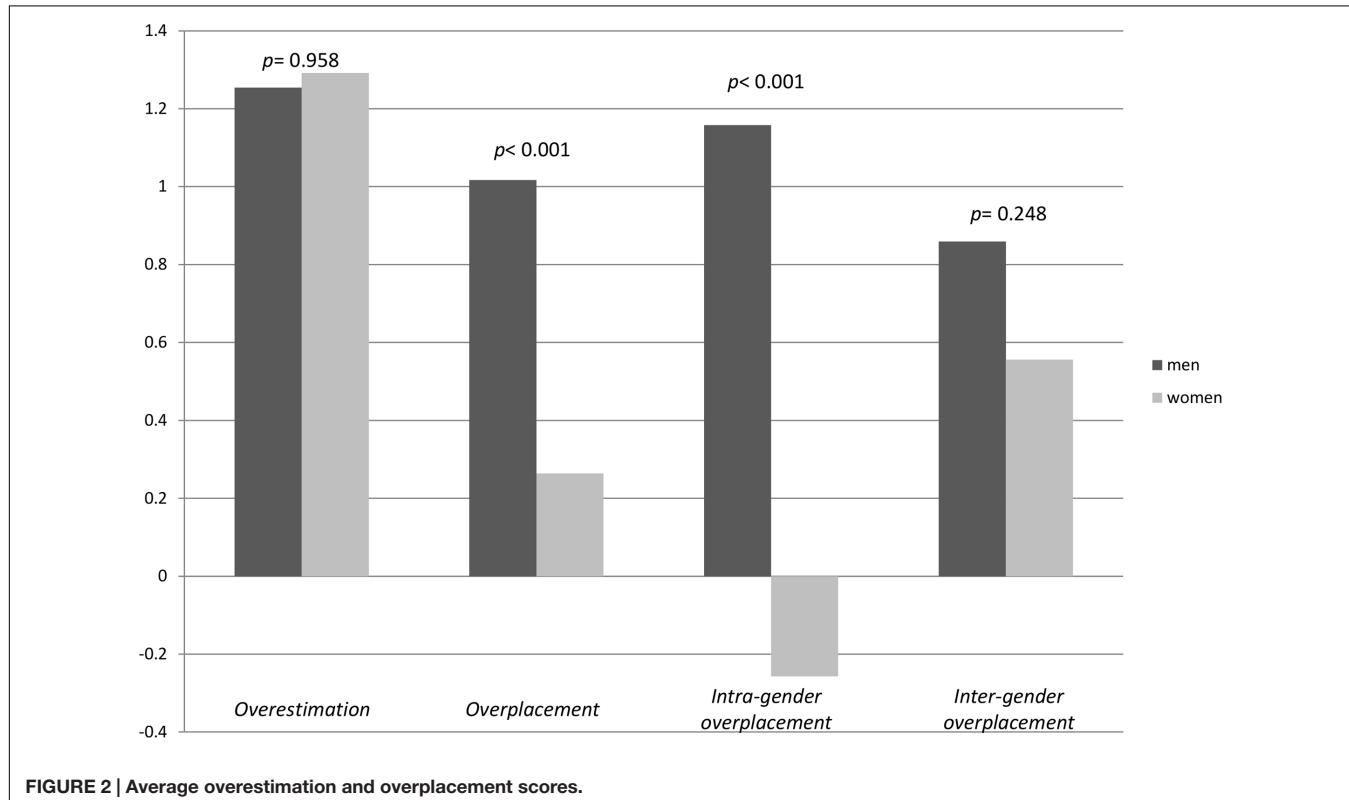
Both men and women in our sample overestimated their own scores (Figure 2). The average overestimation score for men is 1.25 ( $SD = 1.409$ ) and for women 1.29 ( $SD = 1.542$ ). A Wilcoxon rank-sum test does not detect any statistically significant gender difference in overestimation scores ( $z = 0.053$ ,  $p = 0.958$ ). Yet, men tend to overplace themselves significantly more than women ( $z = -3.737$ ,  $p < 0.001$ ). The average overplacement score is 1.02 ( $SD = 1.025$ ) for men and 0.26 ( $SD = 1.138$ ) for

<sup>5</sup>The third aspect is excessive precision, yet we only focus on the first two in the current study.

**TABLE 2 | Generated overestimation and overplacement variables.**

				Min	Max
Overestimation	Actual score	–	Prediction about one's performance	–2	6
Overplacement	Prediction about one's performance	–	Prediction about the others' performance (all)	–2	5
Intra-gender overplacement	Prediction about one's performance	–	Prediction about the others' performance (own gender)	–3	5
Inter-gender overplacement	Prediction about one's performance	–	Prediction about the others' performance (other gender)	–2	5

Min (max) is the smallest (largest) value of the respective variable in our data-set.

**FIGURE 2 | Average overestimation and overplacement scores.**

women. Intra-gender overplacement is significantly higher in men than women ( $z = -5.942, p < 0.001$ ). Men's average intra-gender overplacement score is 1.16 ( $SD = 1.142$ ) and women's is –0.26 ( $SD = 1.303$ ). However, significant gender differences in inter-gender overplacement were not observed ( $z = -1.155, p = 0.248$ ). The inter-gender overplacement scores are 0.86 ( $SD = 1.058$ ) for men and 0.56 ( $SD = 1.047$ ) for women.

In a nutshell, we observe that men think they perform significantly better than other men and do so significantly more than women. The equality between women's predictions about their own performance and their female peers, however, cannot be rejected.

## DISCUSSION

The main outcome of the study is that men think that they would perform significantly better on the 7-item CRT than their male peers, while women made comparable predictions about their own performance and their female peers. This gender-specific

overplacement variable is significantly different between men and women with men overplacing their performance more than women.

A large body of literature in economics and psychology suggests that women, on average, are less confident and competitive than men (see Croson and Gneezy, 2009 for an overview). We contribute to this literature by uncovering gender differences in confidence about one's performance relative to same and opposite sex peers. Previous research has indicated that social components in a choice situation have an impact on gender differences in confidence and competitive behavior. On the one hand, it has been shown that women are more confident in their group's performance than in their own performance, while men are less confident in their group's performance compared to their own (Healy and Pate, 2007). While this study indicates that predictions about one's own and other's performance might differ by gender in certain situations, it does not specifically assess whether differences in performance are due to gender distribution within the reference group. Therefore, it is not

directly comparable to the present study, where each participant was specifically asked about her/his prediction about men's and women's performance separately. On the other hand, it has been shown that men's decision to enter a tournament or a piece-rate pay scheme can depend on the co-participant's gender (Datta Gupta et al., 2013). In that study, men competed less against other men than against women, when the gender information was made sufficiently salient. While this result appears to be out of line with our findings, which might be due to the task type or the transition from beliefs to actions, it shows that competitive behavior might have intra- and inter-gender-specific components. This is a finding that is also often reported in the context of evolutionary-biology, which we refer to in the next part of the discussion.

In the course of human evolution, competition among males typically took place as direct and aggressive contests. Competition among females, by contrast, often occurred more indirectly and subtly. One potential explanation for these different types of behavior could be that the attractiveness of direct intra-gender competition is different for men and women, as they have different perceptions about their same-sex peers. It may be suggested that confidence in one's own abilities relative to one's competitors is an important drive underlying this observation. The link between beliefs about relative skill and the decision to enter competition has been established by several economics experiments (Camerer and Lovallo, 1999). If men think that they perform better than their peers (other men), it potentially makes direct competitions attractive for them. If women, by contrast, think that their peers (other women) perform similarly to them, direct competition appears less attractive and competition might take place on a more subtle level.

It appears important, however, to stress the possibility of reversed causality. It might be that evolutionary differences in competitiveness as an attitude may affect confidence beliefs due to self-enhancement. Self-enhancement refers to the fact that individuals gain positive utility by comparing themselves with lower ranked peers (Wood and Taylor, 1991). In particular, due to evolutionary differences between the levels of male and female competitiveness, confidence beliefs may give them different utility levels. Falk and Knell (2004) developed a social comparison model that includes self-enhancement and self-improvement in the utility function. The model shows that people with higher abilities tend to compare themselves with people who also have high abilities. They also show that women have lower reference standards. This finding is in line with our results showing that women over-place themselves less than men.

## REFERENCES

- Anderson, C., Brion, S., Moore, D. A., and Kennedy, J. A. (2012). A status-enhancement account of overconfidence. *J. Pers. Soc. Psychol.* 103, 718–735. doi: 10.1037/a0029395  
 Barber, B. M., and Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Q. J. Econ.* 116, 261–292. doi: 10.1162/003355301556400

Some words of caution are in order. First, this study is on predictions about others relative to one's performance. In future studies whether the above outlined causal chain from beliefs about performance translates into actual competitive behavior should be addressed. Second, performance predictions for the CRT were studied. This is a special task that aims on impulsiveness of decision-making and to what extent our findings apply in a broader context deserves further investigation.<sup>6</sup> Dreber et al. (2014), for example, have shown that the type of task matters with respect to gender differences in competitive behavior. Third, confidence and competition are social notions that develop via countless interactions in distinct contexts. Due to its specific research questions, the design of the current study does not involve any social interaction between participants. Yet, it may be the case that overconfident behavior can alter with human interaction or social motives. For example, Burks et al. (2013) have shown that overconfidence can be induced by the desire to send positive signals to others about one's own skill. Therefore, attention needs to be paid to the role of social interaction and motives on overconfidence in future studies.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

The study is part of the project "Neurobiological Foundations of Economic Decision Making under Uncertainty and Excessive Risk Taking," which is supported by the Leibniz Association (SAW-2013-IfW-2). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01680/full#supplementary-material>

<sup>6</sup> See Cueva et al. (2016) for a review of reflective and non-reflective responses in CRT.

- Bock, O., Baetge, I., and Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *Eur. Econ. Rev.* 71, 117–120. doi: 10.1016/j.euroecorev.2014.07.003  
 Brañas-Garza, P., Kujal, P., and Lenkei, B. (2015). *Cognitive Reflection Test: Whom, How and When*. Available at: <https://mpra.ub.uni-muenchen.de/68049/>  
 Burks, S. V., Carpenter, J. P., Goette, L., and Rustichini, A. (2013). Overconfidence and social signalling. *Rev. Econ. Stud.* 80, 949–983. doi: 10.1093/restud/rds046

- Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behav. Brain Sci.* 12, 1–14. doi: 10.1017/S0140525X00023992
- Camerer, C., and Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *Am. Econ. Rev.* 89, 306–318. doi: 10.1257/aer.89.1.306
- Croson, R., and Gneezy, U. (2009). Gender differences in preferences. *J. Econ. Lit.* 47, 448–474. doi: 10.1257/jel.47.2.448
- Cueva, C., Iturbe-Ormaetxe, I., Mata-Pérez, E., Ponti, G., Sartarelli, M., Yu, H., et al. (2016). Cognitive (ir) reflection: New experimental evidence. *J. Behav. Exp. Econ.* 64, 81–93. doi: 10.1016/j.soec.2015.09.002
- Datta Gupta, N., Poulsen, A., and Villevall, M. C. (2013). Gender matching and competitiveness: Experimental evidence. *Econ. Inq.* 51, 816–835. doi: 10.1111/j.1465-7295.2011.00378.x
- De Bondt, W. F. M., and Thaler, R. H. (1995). “Financial decision-making in markets and firms: A behavioral perspective,” in *Handbooks in Operations Research and Management Science*, Vol. 9, eds R. A. Jarrow, V. Maksimovic, and W. T. Ziemba (Amsterdam: Elsevier), 385–410.
- Dean, M., and Ortoleva, P. (2015). *Is It All Connected? A Testing Ground for Unified Theories of Behavioral Economics Phenomena*. Available at: <https://ssrn.com/abstract=2643355>
- Dreber, A., von Essen, E., and Ranehill, E. (2014). Gender and competition in adolescence: Task matters. *Exp. Econ.* 17, 154–172. doi: 10.1007/s10683-013-9361-0
- Ermer, E., Cosmides, L., and Tooby, J. (2008). Relative status regulates risky decision making about resources in men: Evidence for the co-evolution of motivation and cognition. *Evol. Hum. Behav.* 29, 106–118. doi: 10.1016/j.evolhumbehav.2007.11.002
- Falk, A., and Knell, M. (2004). Choosing the joneses: Endogenous goals and reference standards. *Scand. J. Econ.* 106, 417–435. doi: 10.1111/j.0347-0520.2004.00370.x
- Frederick, S. (2005). Cognitive reflection and decision making. *J. Econ. Perspect.* 19, 25–42. doi: 10.1257/089533005775196732
- Friedl, A., Ring, P., and Schmidt, U. (2016). Gender differences in ambiguity aversion under different outcome correlation structures. *Theory Decis.* 1–9.
- Healy, A., and Pate, J. G. (2007). *Overconfidence, Social Groups, and Gender: Evidence from the Lab and Field*. Available at: <https://ssrn.com/abstract=934320>.
- Johnson, D. D., and Fowler, J. H. (2011). The evolution of overconfidence. *Nature* 477, 317–320. doi: 10.1038/nature10384
- Kahneman, D., and Frederick, S. (2002). “Representativeness revisited: Attribute substitution in intuitive judgment,” in *Heuristics and Biases: The Psychology of Intuitive Judgment*, eds T. Gilovich, D. Griffin, and D. Kahneman (Cambridge: Cambridge University Press), 49–81.
- Kanter, R. M. (2004). *Confidence: How Winning Streaks and Losing Streaks Begin and End*. New York, NY: Crown Business.
- Lichtenstein, S., Fischhoff, B., and Phillips, L. D. (1982). “Calibration of probabilities: The state of the art to 1980,” in *Judgment under Uncertainty: Heuristics and Biases*, eds D. Kahneman, P. Slovic, and A. Tversky (Cambridge: Cambridge University Press), 306–334.
- Lundeberg, M. A., Fox, P. W., and Puncocoar, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *J. Educ. Psychol.* 86, 114–121. doi: 10.1037/0022-0663.86.1.114
- Malmendier, U., and Tate, G. (2005). CEO overconfidence and corporate investment. *J. Finance* 60, 2661–2700. doi: 10.1111/j.1540-6261.2005.00813.x
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Rev. Econ. Stud.* 60, 531–542. doi: 10.2307/2298123
- Meyer, A. D., Payne, V. L., Meeks, D. W., Rao, R., and Singh, H. (2013). Physicians’ diagnostic accuracy, confidence, and resource requests: A vignette study. *JAMA Intern. Med.* 173, 1952–1958. doi: 10.1001/jamainternmed.2013.10081
- Moore, D., and Healy, P. J. (2008). The trouble with overconfidence. *Psychol. Rev.* 115, 502–517. doi: 10.1037/0033-295X.115.2.502
- Neale, M. A., and Bazerman, M. H. (1991). *Cognition and Rationality in Negotiation*. New York, NY: The Free Press.
- Neyse, L., Bosworth, S., Ring, P., and Schmidt, U. (2016). Overconfidence, incentives and digit ratio. *Sci. Rep.* 6:23294. doi: 10.1038/srep23294
- Niederle, M., and Vesterlund, L. (2007). Do Women shy away from competition? Do men compete too much? *Q. J. Econ.* 122, 1067–1101.
- Oblig, K. H., Hansen, A. L. S., Overgaard, K., Normann, K., Sandbaek, A., and Maindal, H. T. (2015). Association between self-reported and objectively measured physical fitness level in a middle-aged population in primary care. *Prev. Med. Rep.* 2, 462–466. doi: 10.1016/j.pmedr.2015.05.010
- Oskamp, S. (1965). Overconfidence in case-study judgments. *J. Consult. Psychol.* 29, 261–265. doi: 10.1037/h0022125
- Pajares, F. (1996). Self-efficacy beliefs and mathematical problem-solving of gifted students. *Contemp. Educ. Psychol.* 21, 325–344. doi: 10.1006/ceps.1996.0025
- Reuben, E., Rey-Biel, P., Sapienza, P., and Zingales, L. (2012). The emergence of male leadership in competitive environments. *J. Econ. Behav. Organ.* 83, 111–117. doi: 10.1016/j.jebo.2011.06.016
- Reuben, E., Sapienza, P., and Zingales, L. (2014). How stereotypes impair women’s careers in science. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4403–4408. doi: 10.1073/pnas.1314788111
- Schmidt, U., Friedl, A., and Lima de Miranda, K. (2015). *Social Comparison and Gender Differences in Risk Taking*. Available at: <https://www.ifw-members.ifw-kiel.de/publications/social-comparison-and-gender-differences-in-risk-taking/kwp-2011>
- Soll, J. B., and Klayman, J. (2004). Overconfidence in interval estimates. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 299–314.
- Stael von Holstein, C. (1972). Probabilistic forecasting: An experiment related to the stock market. *Organ. Behav. Hum. Perform.* 8, 139–158. doi: 10.1016/0030-5073(72)90041-4
- Stockley, P., and Campbell, A. (2013). Female competition and aggression: Interdisciplinary perspectives. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368:20130073. doi: 10.1098/rstb.2013.0073
- Taylor, S. E. (1989). *Positive Illusions: Creative Self-deception and the Healthy Mind*. New York, NY: Basic Books.
- Taylor, S. E., and Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychol. Bull.* 103, 193–210. doi: 10.1037/0033-2950.103.2.193
- Taylor, S. E., and Brown, J. D. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychol. Bull.* 116, 21–27. doi: 10.1037/0033-2950.116.1.21
- Toplak, M. E., West, R. F., and Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Think. Reason.* 20, 147–168. doi: 10.1080/13546783.2013.844729
- Trinugroho, I., and Sembel, R. (2011). Overconfidence and excessive trading behavior: An experimental study. *Int. J. Bus. Manag.* 6, 147–152. doi: 10.5539/ijbm.v6n7p147
- Trivers, R. L. (1972). “Parental investment and sexual selection,” in *Sexual Selection and the Descent of Man*, ed. B. Campbell (Chicago, IL: Aldine Publishing Company), 136–179.
- von Hippel, W., and Trivers, R. (2011). The evolution and psychology of self-deception. *Behav. Brain Sci.* 34, 1–16. doi: 10.1017/S0140525X10003018
- Wagenaar, W., and Keren, G. B. (1986). “Does the expert know? The reliability of predictions and confidence ratings of experts,” in *Intelligent Decision Support in Process Environments*, eds E. Hollnagel, G. Mancini, and D. D. Woods (Berlin: Springer), 87–103.
- Wood, J. V., and Taylor, K. L. (1991). “Serving self-relevant goals through social comparison,” in *Social Comparison: Contemporary Theory and Research*, eds J. Suls and T. A. Wills (Hillsdale, MI: Erlbaum), 23–49.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2016 Ring, Neyse, David-Barett and Schmidt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Reference Point Heterogeneity

Ayse Terzi<sup>1</sup>, Kees Koedijk<sup>2</sup>, Charles N. Noussair<sup>3\*</sup> and Rachel Pownall<sup>4</sup>

<sup>1</sup> Department of Finance, Tilburg University, Tilburg, Netherlands, <sup>2</sup> Department of Finance, Tilburg University, Tilburg, Netherlands, <sup>3</sup> Department of Economics and Economic Science Laboratory, University of Arizona, Tucson, AZ, USA,

<sup>4</sup> Department of Finance, Tilburg University, Tilburg, Netherlands

It is well-established that, when confronted with a decision to be taken under risk, individuals use reference payoff levels as important inputs. The purpose of this paper is to study which reference points characterize decisions in a setting in which there are several plausible reference levels of payoff. We report an experiment, in which we investigate which of four potential reference points: (1) a population average payoff level, (2) the announced expected payoff of peers in a similar decision situation, (3) a historical average level of earnings that others have received in the same task, and (4) an announced anticipated individual payoff level, best describes decisions in a decontextualized risky decision making task. We find heterogeneity among individuals in the reference points they employ. The population average payoff level is the modal reference point, followed by experimenter's stated expectation of a participant's individual earnings, followed in turn by the average earnings of other participants in previous sessions of the same experiment. A sizeable share of individuals show multiple reference points simultaneously. The reference point that best fits the choices of the individual is not affected by a shock to her income.

## OPEN ACCESS

### Edited by:

Nikolaos Georgantzis,  
University of Reading, UK

### Reviewed by:

Liudmila Liutsko,  
Park of Biomedical Research in  
Barcelona, Spain  
Ismael Rodriguez-Lara,  
Middlesex University, UK  
Iván Barreda Tarazona,  
Jaume I University, Spain

### \*Correspondence:

Charles N. Noussair  
cnoussair@email.arizona.edu

### Specialty section:

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

Received: 30 May 2016

Accepted: 23 August 2016

Published: 12 September 2016

### Citation:

Terzi A, Koedijk K, Noussair CN and  
Pownall R (2016) Reference Point  
Heterogeneity. *Front. Psychol.* 7:1347.  
doi: 10.3389/fpsyg.2016.01347

## 1. INTRODUCTION

Economic decision making under risk involves the consideration of the probabilities of various outcomes, as well as the perceived utilities of these outcomes. However, empirical work has suggested that when judging and evaluating a risky lottery, reference payoff levels are also critical. A payoff appears to be evaluated based on how it compares to a reference level, with a reference point serving to separate desirable from undesirable outcomes, according to some criterion. Thus, understanding how payoff levels come to be viewed as reference points is a key step in uncovering the cognitive process that generates decisions taken under risk.

Indeed, reference dependence, an asymmetry in the treatment of payoffs above vs. below a benchmark payoff level, has been a robust finding in both economics and psychology, since it was first proposed and documented by Kahneman and Tversky (1979). Reference dependence is a cornerstone of prospect theory, the most influential behavioral model of decision making under risk. Reference points have been shown to characterize decision making in laboratory research, surveys, and in field data from numerous domains. These domains include household saving, labor market participation, consumer behavior, education, and investment decisions (see e.g., Hardie et al., 1993; Camerer, 1997, 2004; Starmer, 2000; Grinblatt and Han, 2005). Experimental studies have documented the effect of reference point formation on the provision of effort (Abeler et al., 2011), the pricing of securities (Tversky and Kahneman, 1992), and the exchange and valuation of consumer products (Ericson and Fuster, 2011).

However, while there is general agreement that reference points are important, little is known about which payoff levels will come to serve as reference points. Typically, in empirical work, the reference points of the decision maker are taken as evident given the decision context. This is reasonable in some settings, though less plausible in others. There are no widely-accepted, general accounts of how a particular payoff level emerges as a reference point.

Furthermore, it is not clear that in a particular given decision context, only one unique reference point is relevant. Kahneman (1992) raises the possibility of multiplicity of reference points and characterizes this as an important topic for future study. Sullivan and Kida (1995) demonstrate that corporate managers form multiple reference points, specifically the historical profit level, as well as profit and revenue targets. In an experimental study, Baucells et al. (2011) show that the reference trading price of a financial asset is a combination of multiple potential reference prices.

One class of prominent theories of reference point formation is based on the expectations of the decision maker herself (Bell, 1985; Loomes and Sugden, 1986; Kőszegi and Rabin, 2006, 2007; Heidhues and Kőszegi, 2008). Expectations-based reference points have been used to explain insurance choices (Barseghyan et al., 2011), and labor supply decisions (Farber, 2005, 2008; Crawford and Meng, 2011). However, the payoffs that peers receive are also relevant. Experimental work has largely supported the models of inequity aversion proposed by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), which assume that the average payoff of peers serves as a reference point. Furthermore, expectations can be formed through a history of social interaction, e.g., contracts, experiences, past trends, or the recommendations of others (Davies and Kandel, 1981; Abel, 1990; Gali, 1994; Carmeli and Schaubroeck, 2007; Vendrik and Woltjer, 2007; Post et al., 2008; Linde and Sonnemans, 2012). Kőszegi and Rabin (2006) point out that there are multiple candidates that can serve as expectation-based reference points. They emphasize that candidate reference points might also coincide. For example, the expectations of an individual about her own and her peers' payoffs may be the same in some instances. The reference point in effect is obviously consequential. For example, Kőszegi and Rabin (2007), argue that the implications of reference dependence differ depending on the specification of the reference point.

Thus, there are several candidate expectation-based reference levels that appear to be prominent. The purpose of the paper is to study which reference points characterize decisions in a setting in which there are several plausible reference levels of payoff. The question we consider here is individuals differ from each other in their propensity to use different reference points, when they make decisions in the same setting. We study which, if any, of four candidate reference points is most likely to emerge in a decontextualized setting. If the reference points that emerge vary greatly by individual, it can only be due to differences arising from the individuals themselves, rather than the task or the setting.

To investigate this, we conduct an experiment which allows a participant to use any or all of four competing reference points in a risky decision making task. The first is the payoff level for the individual anticipated by the experimenter (who may be interpreted as an authority figure or an employer). We abbreviate this reference point as IE, or Individual Expectation. This level, indicated on each subject's instructions, is a natural candidate for a reference point, since it directly ascribes a benchmark for the individual to attain. The second potential reference point is the anticipated average payoffs of peers in the same decision situation (PE, Peer Expectation). This is also indicated in writing on an individual's instructions, with equal prominence as IE. Note that expectations, as used here, do not refer to an individual's own beliefs or aspirations, or to a mathematical expectation of her payoff. The third is the historical average payoff of others in the same position in past sessions (HA, Historical Average), also indicated in the instructions, and the fourth is the average performance of a relatively large population (PA, Population Average), which is known to subjects at the time of recruitment to the session. PE, HA, and PA all represent payoffs of other individuals in the same or similar experiments, but vary in the social distance between the parties they apply to and the individual herself. Because there is no compelling rationale for believing that one reference point would dominate the others, we refrain from advancing hypotheses in advance about which reference points would be most consistent with the data.

In our experimental design, we present three of the reference points simultaneously, in order to conduct a horse race between the alternatives. In some session we presented PA, IE, and HA, while in others session the payoff levels displayed were PA, IE, and PE. We elicit the certainty equivalents of a large number of lotteries and obtain estimates of individual reference points. The design permits the detection of individuals who use none or one unique reference point, as well as those who employ multiple reference points concurrently. By using one fixed probability for gains and losses of 0.5 throughout the experiment, we attenuate the impact of probability weighting on our results.

It is also important to understand whether reference points change in response to shocks to wealth levels. Some studies have considered this topic. Arkes et al. (2008) show that subjects are more likely to adapt their reference points to gains in their wealth than to losses. Chen and Rao (2002) stress the importance of the order of presentation of two equally-sized gains and losses. They suggest that the first payoff that is presented leads to a more significant adaptation of the reference point than the second. In a financial market setting, Baucells et al. (2011) show that reference prices for a financial asset are a function of the first and the last trading price. Masatlioglu and Ok (2005) model the theory of choice in a static setting where the initial endowment or status quo plays a key role. They show that an agent with reference-dependent preferences prefers to stay at his status quo as long as another option does not dominate it in all dimensions. Post et al. (2008) find evidence of path dependence in reference levels in choices under risk. One of the treatments in our experiment is complementary to this strand of research, and allows us to study the adjustment of the reference point after a shock to one's income level.

Our results show that if all individuals are classified by the one reference point that they adhere to most closely, the population average (PA) is employed most frequently followed by the individual expectation (IE), and then by the historical average (HA). The social comparison group which is the most distant though also the largest, the population of experimental subjects, appears to be the most relevant. Multiple reference points are observed for a sizable share of individuals, while some others show no evidence of having any reference point. Many individuals use a heuristic, in which they value a lottery at a fixed percentage of its expected value. Finally, we find evidence that reference points do not change after a shock to income has occurred. Overall, these results reveal that there is individual-level heterogeneity in the use of reference points within a fixed decontextualized setting. Thus, reference point choice is driven in part by personal inclination.

The remainder of this paper is organized as follows. Section 2 describes the experimental design. In Section 3 we discuss the results, and Section 4 concludes the paper.

## 2. MATERIALS AND METHODS

### 2.1. Conduct of Sessions and Procedures

A total of 44 sessions were conducted at the Centerlab at Tilburg University in The Netherlands, between November 2013 and June 2014. Subjects were all Bachelor's and Master's students in Economics and Business Administration, and therefore were relatively homogeneous in their training. A total of 163 subjects participated. Fifty-five percent were male. The average age of member of the subject pool is 22. The experiment was executed with the z-Tree computer program (Fischbacher, 2007). There was a varying number of participants per session and each subject acted independently of others in this individual decision making experiment. Each session lasted 45 min, including the time during which the experimenter read the instructions. The payoffs in the experiment were expressed in terms of an experimental currency, which was converted to a Euro payment to subjects at the end of the sessions. The average earnings per subject were 16 Euros (1 Euro = \$1.30 approximately at the time the experiment was conducted).

A session consists of 60 periods. In each period  $t$ , subjects are presented with a binary prospect  $(1/2, y_t)$ , which results in outcome  $y_t$  with probability 0.5 and in outcome 0 with probability 0.5. This prospect is paired with eight different certain payment levels,  $x_{jt}, j = 1, \dots, 8$  in a price list format, during each of the 60 periods. In each period, each subject must make eight choices. Each choice in period  $t$  is between  $(1/2, y_t)$  and  $x_{jt}$ . The eight choices are displayed on the subject's computer screen simultaneously. The magnitude of  $x_{jt}$  ranges in value from 40 to 180% of  $y_t/2$ , the expected value of the prospect. The certain payments appear in ascending order of magnitude in the price list on the computer screen.

The sixty periods are divided into three 20-period segments. The certain payments  $x_{jt}$ , as well as the amount that the lottery can pay out  $y_t$ , increases in constant increments from one period to the next within each segment. The lowest certain amount  $x_{jt}$  chosen by the subject over  $(0.5, y_t)$  in period  $t$ , serves as our

measure of the certainty equivalent for the prospect  $(0.5, y_t)$  for that subject. The expected value of the prospects and the potential certainty equivalents span the four potential reference points. Thus, the expected values of  $(0.5, y_t)$ , as well as the value of  $x_{jt}$ , are in some instances in the domain of gains and at other times in the domain of losses relative to each of the four reference points we consider.

At the beginning of a session, the experimenter read the instructions for the experiment aloud. The instructions included key statements about earnings, which were intended to introduce the candidate reference points.

Subjects registered through an online system and at that time were informed of the average earnings in Euros for experiments of similar length conducted at the laboratory, 12 Euros. This is the overall average payoff of subjects participating in an experiment at Centerlab, and we interpret this level as the PA reference point.

At the start of the experiment, each subject was given information about his/her initial cash balance, which was hers to keep. This information remained on her computer screen for the duration of the session. The initial balance was always less than the PA reference level. Therefore, to reach the PA level, the subject had to earn the difference between this level and the initial balance.

The level of the IE reference point was indicated in bold font on the instructions that subjects received at the beginning of the session. It was also displayed on participants' computer screens for the entire session. It was emphasized that this individual expectation was not based on any specific knowledge about the realized final outcome, but only about what could be expected beforehand based on the way the experiment was designed.

In sessions 2–24, the historical average of earnings of participants from previous sessions of the experiment (the HA reference point) was also emphasized in the instructions and indicated on the computer screens. In sessions 25–44, the PE reference point was presented similarly.

We varied the magnitudes of the four reference points in different sessions. The values of each of the four candidate reference points are shown in **Table 1**. The first column of **Table 1** indicates the session, and each row groups together sessions conducted under identical parameters. The next three columns contain the monetary values, in terms of experimental currency, of each of the reference points. All four reference points are net of the initial endowment, which differs by individual. The PE and IE were adjusted to reflect the different parameters in effect in different sessions, and the HA differed because earnings of individuals in previous sessions varied. Each reference point was always a unique value for an individual subject, and the intervals in the table indicate the range of differing unique reference points among subjects in the session indicated. The ranges within each session are indicated in columns 2 and 3. Columns 5 and 6 give the exchange rate between experimental currency and Euros in effect, and whether there was an income shock after period 40. The payoffs were denominated in terms of an experimental currency that was convertible to Euro at the end of the session, at a conversion rate indicated in the second-to-last column of **Table 1**.

**TABLE 1 | Parameters used in the experiment.**

Session	Initial balance	Expectation of own earnings (IE)	Historical Average (HA)	Expectation of Peers (PE)	Population Average (PA)	Exchange rate	Treatment
1*	3500–6500	5500–8500	–	–	9100–12,100	1300	Baseline
2–3	4500–7000	7000–9500	15,600	–	8600–11,100	1300	Baseline
4–5	4500–7000	7000–9500	13,500	–	8600–11,100	1300	Baseline
6–7	4500–7000	7000–9500	12,700	–	8600–11,100	1300	Baseline
8–10	8500–10,500	15,500–17,000	13,100	–	7500–9500	1500	Baseline
13–24	35,000–45,000	45,000–60,000	28,500	–	33,000–43,000	6500	Baseline
25–32	50,000–60,000	70,000–85,000	–	100,000	42,000–52,000	8500	Shift
33–44	50,000–60,000	70,000–85,000	–	100,000	48,000–58,000	9000	Shift

\*Session 1 is excluded from the analyses due to the absence of a historical average. IE is the earnings level that the experimenter indicates to individual that is expected of her. PE is the earnings level that the experimenter indicates to an individual that he/she expects others participating in the same session to earn. HA is the average earnings of individuals in all prior sessions. PA is 12 Euros, the average earnings in all experimental studies conducted at the laboratory, minus the initial endowment. All reference points are similarly expressed net of the initial endowment and income shock. Within a session different individuals had different initial balances, IE and PA reference points. Thus, the indicated values are ranges. However, each individual himself had a unique initial balance, IE and PA level.

At the end of the session, the computer randomly chose one period  $t$  and one of the decisions within that period to count as each subject's earnings. Depending on the choice of the subject, the subject either played the lottery and received one of the outcomes of the prospect, 0 or  $y_t$ , or obtained the certain amount  $x_{jt}$ .<sup>1</sup>

## 2.2. Treatments

There were two treatments in the experiment, called Baseline and Shift. The last subsection described the Baseline treatment. In the sessions of the Shift treatment, we induced an exogenous shock to income after the 40th period by paying a bonus that was unanticipated by subjects. The bonus for each individual was equal to 50% of the initial endowment. It was emphasized that the shock was independent of the earlier choices participants made. The shock was described to participants by the following announcement made by the experimenter before period 1. "If during the course of the experiment any new information will be shown to you on the screen, please note that this is not due to the decisions you have previously made in the experiment. The computer does not do anything with your decisions until the experiment finishes."

## 3. RESULTS

This section is organized in the following manner. We first informally describe the data from two typical subjects. Section 3.1 describes and documents the widespread use of a rule, called the Proportional Discounting Heuristic, employed by 38% of our participants. Section 3.2 contains our analysis of the prevalence of the four different reference points.

**Figures 1, 2** illustrate two of the typical decision profiles in our data. The horizontal axis gives the period number, while the vertical axis shows monetary amounts expressed in terms of experimental currency. The points displayed in black are the expected values of the prospects presented in the period

<sup>1</sup>Paying only one period removes wealth effects. Starmer and Sugden (1991) have shown that this procedure generates behavior that is similar to that when all periods are paid.

indicated. The certainty equivalents elicited from the subject in the period are given by the gray points. The leftmost panel shows the expected values of the prospects and the certainty equivalents elicited in the first 20 periods. The expected values of these prospects include values both above and below a candidate reference level. The figure shows that the certainty equivalents of subject 16, who is depicted in the figure, are greater than the expected value of the prospects, whenever the expected value lies in the domain of losses relative to the PA reference point. Thus, the subject exhibits risk seeking behavior in this domain. When the expected value of the lottery lies above the PA, the observed certainty equivalents are less than the expected value of the prospects, which is consistent with risk averse preferences. Thus, we observe here that the subject changes her attitude toward risk at the PA payoff level<sup>2</sup>.

Another example, for subject 13, is presented in **Figure 2**. The certainty equivalents of this subject are all equal to the expected value of the prospect, whenever the expected value of the prospect is less than the Historical Average. This indicates that the individual is risk neutral in the domain of losses, relative to the HA reference point. When the expected value of the prospect is greater than HA, the individual becomes risk averse.

### 3.1. The Proportional Discounting Heuristic

A very common decision rule, employed by 38% of individuals, is the Proportional Discounting Heuristic. This rule involves setting a certainty equivalent equal to a constant fraction of the expected value of the lottery (or alternatively to a constant fraction of the maximum possible outcome of the lottery), as is depicted in **Figure 3**. The agent depicted in this figure has no reference point in the range spanned by the possible certain payments offered in the experiment (although we cannot rule out the possibility that the agent has a reference point at 0, for example). The certainty equivalent of individuals who proportionally discount is given by:

<sup>2</sup>One measure of consistency of choices that can be applied to the data is whether subjects' certainty equivalent increases from one period to the next within each 20-period segment. By this criterion, 14 subjects are consistent for all 60 decisions, 46 have fewer than 5 inconsistencies, and 98 have fewer than 10.

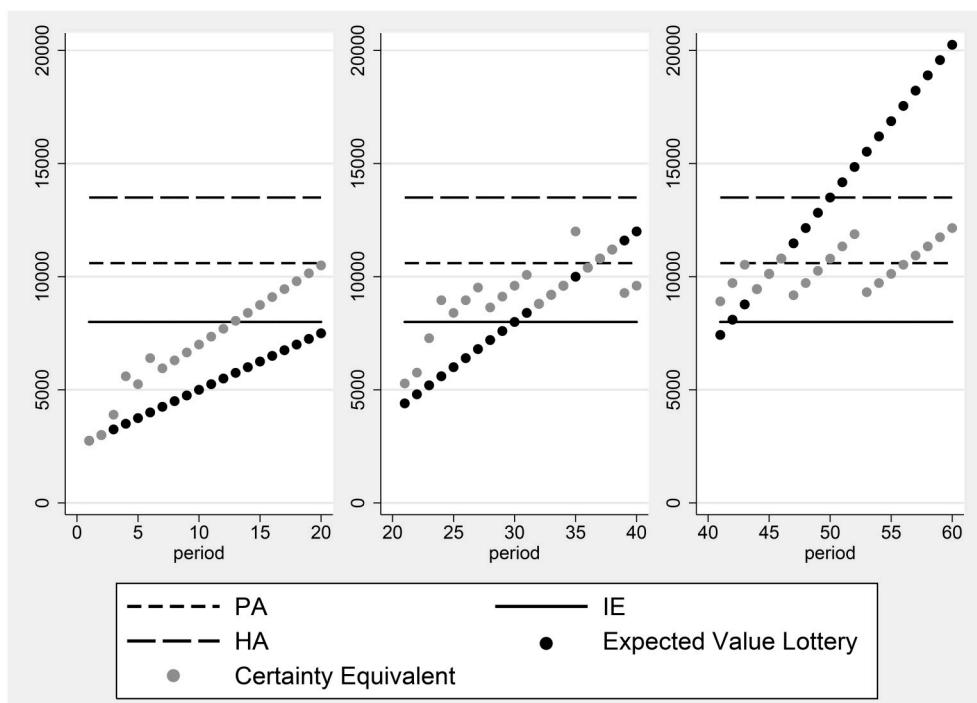


FIGURE 1 | Certainty equivalents of subject 16, who participated in session 5.

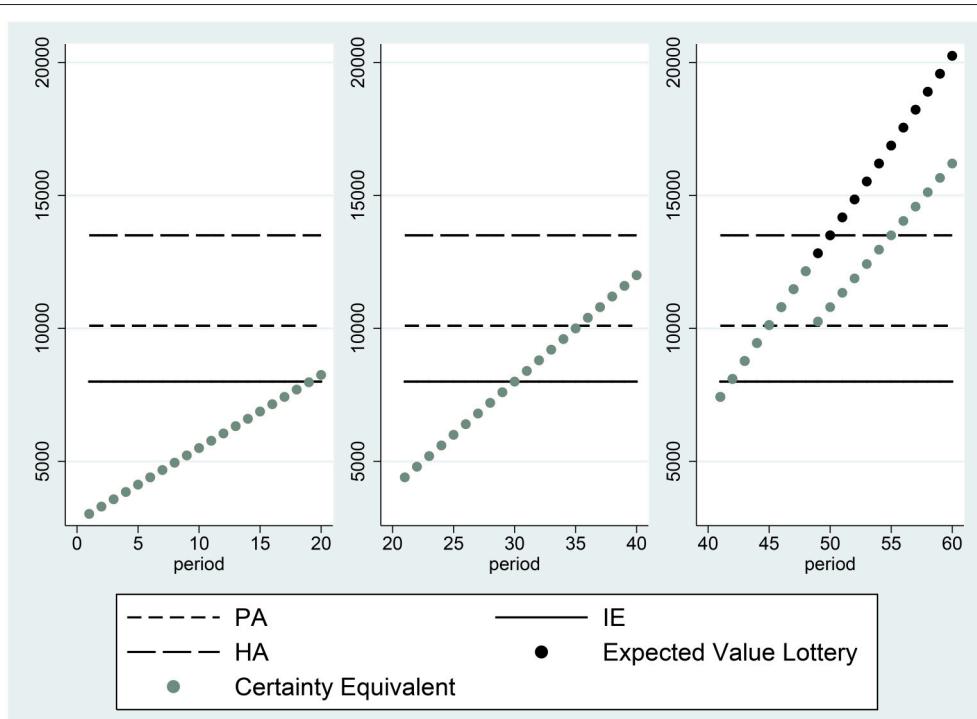
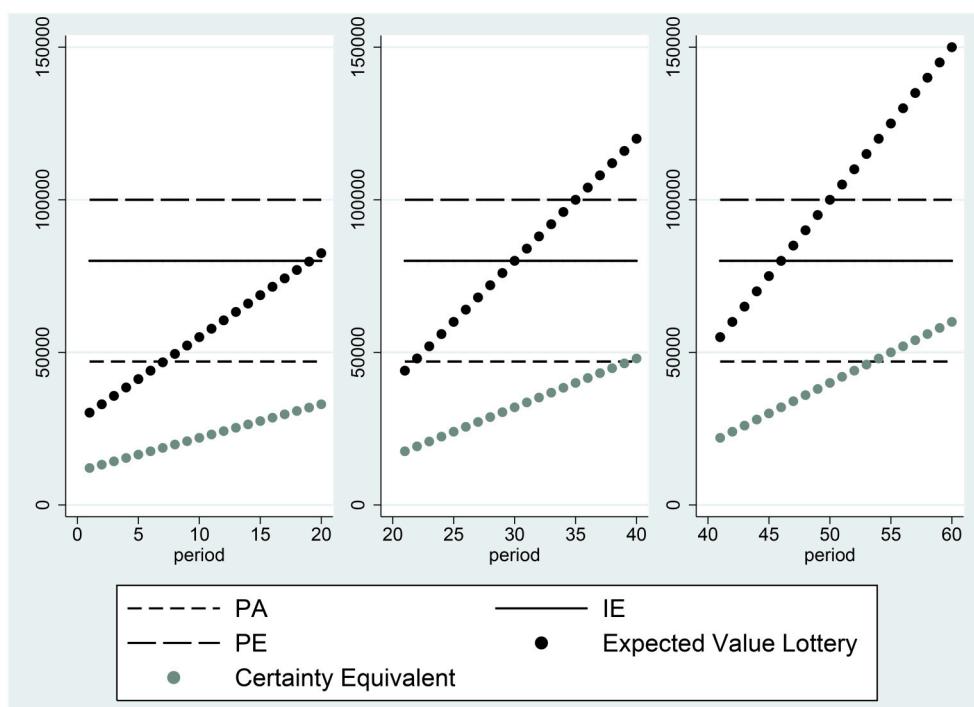


FIGURE 2 | Certainty equivalents of subject 13, who participated in session 3.



**FIGURE 3 |** Certainty equivalents of subject 100, who participated in session 32 and did not employ a reference point.

$$\text{Certainty equivalent} = \alpha * \text{Expected value of lottery} = \alpha * y_t / 2 \quad (1)$$

If  $\alpha = 1$ , the individual is risk neutral. Another heuristic which is observationally equivalent is the rule that  $\text{Certainty equivalent} = \theta * y_t$ , where  $\theta = \alpha/2$ . Our setting is conducive to observing the proportional discounting heuristic, because of the price list format and the sequence of presentation of the choices. This is because if a subject switches from the safe choice  $x_{jt}$  to the risky choice  $y_t$  at the same row on the table in all periods, his behavior is consistent with the heuristic. Thus, an individual who wishes to apply the heuristic would not find it excessively cognitively demanding to do so. The average  $\alpha$  parameter for this subsample is 0.92, equalling 0.96 for male and 0.90 for female subjects.

It is possible, if individuals have reference-dependent preferences, that  $\alpha$  can differ between the domains of losses and gains, as proposed by Iturbe-Ormaetxe et al. (2011), Iturbe-Ormaetxe Kortajarena et al. (2015). Such a shift in the discount proportion can be seen in the right panel of Figure 2. This behavior reveals a discrete change in attitude toward risk above vs. below the reference point. However, in data such as ours, a classification of individuals according to the behavioral rules they employ, such as the Proportional Discounting Heuristic, must allow for some trials to exhibit deviations from the exact decision consistent with the heuristic. To classify individuals as users of the Proportional Discounting Heuristic, we calculate the following:

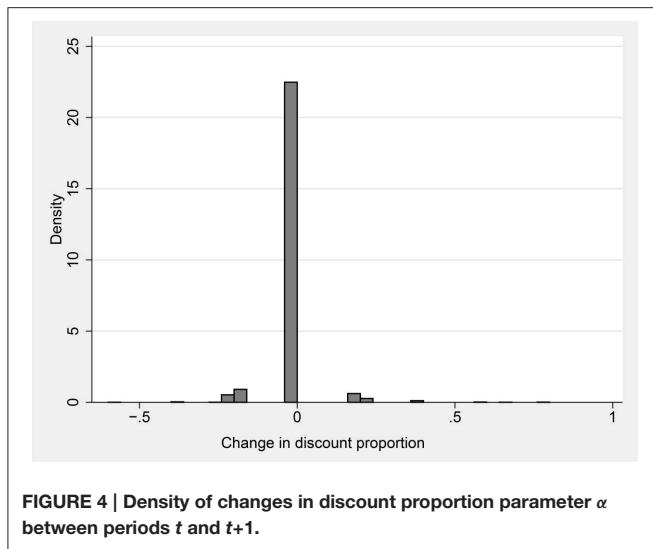
$$\Delta \text{proportional valuation} = (\text{certainty equivalent}/\text{expected value lottery})_t - (\text{certainty equivalent}/\text{expected value lottery})_{t-1} \quad (1)$$

$$x_{jt}^*/(0.5 * y_t) - x_{j,t-1}^*/(0.5 * y_{t-1}), x_{jt}^* = \min_j \{x_{jt} | x_{jt} \geq 0.5 * y_t\} \quad (2)$$

If the agent uses the proportional valuation heuristic, valuing every lottery at the same constant fraction of its expected value, then  $\Delta \text{proportional valuation}$  always equals zero. We classify an individual as a proportional discounter if she exhibits no more than six instances over the 60-period session, in which Equation (2) does not equal 0. Figure 4 illustrates the stability of the strategy employed on the part of users of the heuristic. The figure is a histogram of ( $\Delta \text{proportional valuation}$ ) for the 38% of the sample that are proportional discounters. The change in proportional valuation is zero in the great majority of cases.

### 3.2. Reference Points Employed

To identify the reference points subjects are using, we focus on the manner whereby a reference point influences decisions. We test for the presence of a target payoff level by investigating the choice between playing the lottery and receiving the certain payment. We expect that the presence of a reference point will influence decisions when the certain payment is just above the reference level. In such cases, agents might forego some expected payoff and choose the certain payment, in order to reach their reference payoff. To test for this pattern, we model the choice between the certainty equivalent and the lottery of each



individual as a function of the value of the certainty equivalent, the expected value of the lottery and a dummy variable indicating whether the safe option  $x_{jt}$  exceeds the reference point.

$$Z_{ijt} = \alpha_i + \beta_{1,i} 0.5 * y_t + \beta_{2,i} x_{jt} + \gamma_k D_k + \epsilon \quad (3)$$

where

$$D_k = \begin{cases} 1; & \text{if Certain amount} > \text{reference point } k \\ 0; & \text{if Certain amount} \leq \text{reference point } k \end{cases}$$

$Z_{ijt}$  is a binary variable which represents the choice of individual  $i$  between the prospect  $(0.5, y_t)$ , and the certain amount on offer,  $x_{jt}$ , in period  $t$ .  $Z_{ijt}$  takes the value 1 if the individual chooses the prospect, and 0 otherwise. Recall that all reference points are net of the initial endowment. A significant coefficient for the  $\gamma_k$  term would indicate the use of reference point  $k$ , as it reveals a change in the likelihood of choosing the lottery when the certain payment it is paired with exceeds the reference level. In the regression, we control for the expected value of the lottery and the level of the certain payment.

The model is estimated for each individual  $i$  and each reference point  $k$  separately. An  $F$ -test is performed to test for the significance of the restriction  $D_k = 0$ . If the resulting  $F$ -statistic is above the critical level, and the estimated gamma coefficient is negative, we will say that  $k$  is a reference point for the individual. When this test is significant for candidate reference point  $k$ , we say that the individual is using  $k$  as a reference point. Based on the result of this test, we assign an individual to either none, one, or multiple reference points. For each individual, the regression is estimated for each of the potential reference points. **Table 2** shows the incidence of each possible reference point profile in the sample.

The table shows that the PA is the most common reference point for individuals who used only one reference level, followed by IE and HA. PE does not seem to serve as a reference point. A sizable portion of subjects use multiple reference points, and most of these individuals use PA paired with HA. Lastly, a

**TABLE 2 | Reference point use by subjects.**

Session	All sample (%)	Female (%)	Male* (%)
2–24			
None	17.83	16.66	20.57
Population Average (PA)	15.05	23.29	10.26
Individual Expectation (IE)	21.93	26.69	20.52
Historical Average (HA)	8.23	6.69	7.69
PA and IE	2.75	3.34	2.58
PA and HA	34.21	23.34	38.39
IE and HA	0	0	0
All	0	0	0
25–44			
None	26.61	37.42	17.79
Population Average (PA)	62.27	52.53	73.38
Individual Expectation (IE)	2.23	0	2.23
Peer Expectation (PE)	0	0	0
PA and IE	8.88	10.05	6.61
PA and PE	0	0	0
IE and PE	0	0	0
All	0	0	0

\*The gender variable contains 5 missing values.

non-negligible portion of individuals do not appear to employ any of the candidate reference points. Gender differences are not significant, with Fisher exacts tests resulting in  $p$ -values of 0.61 for sessions 2–24, and 0.097 for sessions 25–44.

Regressions with the specification in Equation (3) on the aggregate pooled data from all individuals classified as using each reference point provide an overall picture of the estimated parameters, and of the strength of the attraction of each reference point. Recall that each reference point, other than PA, is specified as in addition to the initial endowment. The estimates are shown in **Tables 3, 4**. The results show that an increase in the expected value of the lottery increases the probability of choosing the lottery. On the other hand, increasing the value of the certain alternative decreases the probability of choosing the lottery. Each of the reference points is negative and significant in both tables. This indicates that for each of the reference points PA, HA, and IE, a subset of subjects exhibits changes in behavior for payoff levels above vs. below the reference point. When the certain payoff exceeds the reference point, it is more likely to be chosen.

### 3.3. Income Shock

In the Shift treatment, we study the effect of a shock to an individual's income level and investigate whether it changes the likelihood of choosing a particular reference point. In this treatment, at the end of period 40, subjects experience a change in their wealth. We increase their cash balance by 50% of their initial endowment, an amount which differs among subjects. Then, in the last 20 periods of the session, the same set of choices as in the first 20 periods are presented to the subjects again. We consider the effect of the shock on the choices of individuals in the last 20 periods of the experiment and compare these to the

**TABLE 3 | Estimated effect of reference point in sessions 2–24.**

	(1) choice	(2) choice	(3) choice
EV Lottery ( $0.5 * y_t$ )	0.05*** (7.48)	0.09*** (14.04)	0.06*** (10.39)
$x_{jt}$	-0.05*** (-9.55)	-0.07*** (-16.09)	-0.06*** (-13.61)
$D_{PA}$	-0.42*** (-16.01)		
$D_{IE}$		-0.37*** (-15.69)	
$D_{HA}$			-0.35*** (-11.04)
Gender	-0.05 (-1.51)	-0.02 (-0.54)	-0.03 (-0.84)
Constant	0.61*** (19.95)	0.49*** (13.65)	0.64*** (18.02)
Observations	16,720	8616	12,896
$(R^2)$	0.514	0.544	0.538

*t* statistics in parentheses.

Robust standard errors.

\*\*\* ( $p < 0.01$ ).**TABLE 4 | Estimated effect of reference point in sessions 25–44.**

	(1) choice	(2) choice
EV Lottery ( $0.5 * y_t$ )	0.06*** (11.63)	0.07*** (11.00)
$x_{jt}$	-0.05*** (-13.85)	-0.05*** (-14.01)
$D_{PA}$	-0.46*** (-26.64)	
$D_{IE}$		-0.33*** (-16.01)
Gender	0.02 (0.81)	-0.01 (-0.32)
Constant	0.56*** (22.18)	0.43*** (9.91)
Observations	29,192	3824
$(R^2)$	0.552	0.579

*t* statistics in parentheses.

Robust standard errors.

\*\*\* ( $p < 0.01$ ).

choices elicited in the first segment of 20 periods, with respect to which reference points most accurately characterize the decision pattern.

We report the proportions of reference points that fit best the decisions of these individuals in **Table 5**. The first column reports a classification of individuals in relation to reference points in periods 1–20 in the Shift treatment. The second column contains analogous data from periods 41–60. The results show no

**TABLE 5 | Reference points of subjects in Shift treatment before and after the income shock.**

	Period 1–20 (%)	Period 41–60 (%)
None	36.07	41.00
Population Average (PA)	59.00	57.35
Individual Expectation (IE)	1.64	0
Peer Expectation (PE)	3.29	0
PA and IE	0	0
PA and PE	0	0
IE and PE	0	1.64
All	0	0

significant difference in the incidence of the use of each reference point before, compared to after, the shock. A Fisher exact test of the equality of the distribution of reference points between periods 1–20 and 41–60 results in  $p = 0.481$ . This may reflect the fact that the shock, like initial income, is treated as a separate source of wealth than the earnings from the experimental task.

## 4. DISCUSSION

In this paper, we document heterogeneity among individuals in their personal inclination to use particular reference points. It is known from previous work that the reference point that characterizes a set of data best differs, depending on the setting in which the decision is taking place. However, we show here that the reference point that best fits the decision pattern of an individual also differs by individual, keeping the decision setting constant.

Our results do indicate that when individuals use a single reference point, the population average payoff level is the most frequently employed. This is followed by the anticipated payoff level indicated for the individual, and in turn by the average that comparable individuals have earned in past similar tasks. No participant used the earnings of peers in the same session as a reference point. The results are similar for men and women and we observe no significant gender differences in the use of reference points.

We also observe that a sizable fraction of individuals employs multiple reference points. The most common combinations of reference points are the population average with the historical average, and the population average with the individual expectation. It is striking to us that PA is such a strong attractor, in light of the fact that the social distance between an individual and the population average is arguably the greatest among all of the reference points that we have considered. The experimental design we have does not allow us to isolate the precise reason that PA is more prominent than the others. However, it does have the feature that it, along with HA, is historical and therefore certain, while IE and PE are anticipated future payoff levels. Furthermore, PA is always constant and known to be the same for all individuals, while the three other reference points can vary among individuals. Perhaps a reference payoff is more compelling when it is common knowledge that it is the same for everyone.

We also find that a considerable share of subjects tend to proportionally discount their certainty equivalent by a constant percentage of the expected payoff of the risky lottery. Some of these individuals also discount by a different fraction, depending on whether payoffs are above or below one or more of the reference points. The widespread use of the Proportional Discounting Heuristic seems intuitive as a behavioral rule, because it is simple to calculate and apply, though to our knowledge its use has not been documented in previous research.

Thus, our experiment illustrates two types of heterogeneity in how individuals perceive risky decision making tasks. The first is that some individuals differ in whether or not they apply a simple heuristic, proportional discounting, to value the lottery, while others adopt more complex or inconsistent valuation methods. The second is that the reference level of earnings that individuals use is idiosyncratic, with some individuals targeting one or more from among a set of prominent reference points, while others do not.

While a number of studies have focused on estimating the mean and median loss aversion parameters of a particular sample, a growing number of studies have documented heterogeneity

in the loss aversion level of individuals (Fehr and Goette, 2007; Gächter et al., 2007; Von Gaudecker et al., 2011). Building on this, other studies have investigated factors affecting the degree of individual loss aversion and have found that demographic characteristics play an important role (Hjorth and Fosgerau, 2009; Payne et al., 2015). Loss aversion only has meaning relative to a reference point. Our results complement this line of research by providing evidence that individuals exhibit different reference points in a similar task. Thus, in addition to having different levels of loss aversion, the reference points from which loss aversion is defined, are heterogeneous.

## AUTHOR CONTRIBUTIONS

AT conducted the experiment and analyzed the data. All four authors contributed to designing the experiment, guiding the analysis of the results, and writing the paper.

## FUNDING

The Center of Economic Research at Tilburg University provided the funds used to compensate participants in our experiment.

## REFERENCES

- Abel, A. B. (1990). Asset prices under habit formation and catching up with the joneses. *Am. Econ. Rev.* 80, 38–42.
- Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *Am. Econ. Rev.* 101, 470–492. doi: 10.1257/aer.101.2.470
- Arkes, H. R., Hirshleifer, D., Jiang, D., and Lim, S. (2008). Reference point adaptation: tests in the domain of security trading. *Organ. Behav. Hum. Decis. Process.* 105, 67–81. doi: 10.1016/j.obhdp.2007.04.005
- Barseghyan, L., Prince, J., and Teitelbaum, J. C. (2011). Are risk preferences stable across contexts? Evidence from insurance data. *Am. Econ. Rev.* 101, 591–631. doi: 10.1257/aer.101.2.591
- Baucells, M., Weber, M., and Welfens, F. (2011). Reference-point formation and updating. *Manage. Sci.* 57, 506–519. doi: 10.1287/mnsc.1100.1286
- Bell, D. E. (1985). Disappointment in decision making under uncertainty. *Oper. Res.* 33, 1–27. doi: 10.1287/opre.33.1.1
- Bolton, G. E., and Ockenfels, A. (2000). Erc: a theory of equity, reciprocity, and competition. *Am. Econ. Rev.* 90, 166–193. doi: 10.1257/aer.90.1.166
- Camerer, C. F. (1997). Progress in behavioral game theory. *J. Econ. Perspect.* 11, 167–188. doi: 10.1257/jep.11.4.167
- Camerer, C. F. (2004). “Prospect theory in the wild: evidence from the field,” in *Advances in Behavioral Economics*, eds D. Kahneman and A. Tversky (Cambridge, UK: Cambridge University Press) 148–161.
- Carmeli, A., and Schaubroeck, J. (2007). The influence of leaders’ and other referents’ normative expectations on individual involvement in creative work. *Leadersh. Q.* 18, 35–48. doi: 10.1016/j.lequa.2006.11.001
- Chen, H., and Rao, A. R. (2002). Close encounters of two kinds: false alarms and dashed hopes. *Mark. Sci.* 21, 178–196. doi: 10.1287/mksc.21.2.178.148
- Crawford, V. P., and Meng, J. (2011). New york city cab drivers’ labor supply revisited: reference-dependent preferences with rational expectations targets for hours and income. *Am. Econ. Rev.* 101, 1912–1932. doi: 10.1257/aer.101.5.1912
- Davies, M., and Kandel, D. B. (1981). Parental and peer influences on adolescents’ educational plans: some further evidence. *Am. J. Sociol.* 87, 363–387. doi: 10.1086/227462
- Ericson, K. M. M., and Fuster, A. (2011). Expectations as endowments: evidence on reference-dependent preferences from exchange and valuation experiments. *Q. J. Econ.* 126, 1879–1907. doi: 10.1093/qje/qjr034
- Farber, H. S. (2005). Is tomorrow another day? The labor supply of New York city cabdrivers. *J. Polit. Econ.* 113, 46–82. doi: 10.1086/426040
- Farber, H. S. (2008). Reference-dependent preferences and labor supply: the case of New York city taxi drivers. *Am. Econ. Rev.* 98, 1069–1082. doi: 10.1257/aer.98.3.1069
- Fehr, E., and Goette, L. (2007). Do workers work more if wages are high? Evidence from a randomized field experiment. *Am. Econ. Rev.* 97, 298–317. doi: 10.1257/aer.97.1.298
- Fehr, E., and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Q. J. Econ.* 114, 817–868. doi: 10.1162/003355399556151
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* 10, 171–178. doi: 10.1007/s10683-006-9159-4
- Gächter, S., Johnson, E. J., and Herrmann, A. (2007). *Individual-level Loss Aversion in Riskless and Risky Choices*. IZA Discussion Paper.
- Gali, J. (1994). Keeping up with the joneses: consumption externalities, portfolio choice, and asset prices. *J. Money Credit Bank.* 26, 1–8. doi: 10.2307/2078030
- Grinblatt, M., and Han, B. (2005). Prospect theory, mental accounting, and momentum. *J. Financ. Econ.* 78, 311–339. doi: 10.1016/j.jfineco.2004.10.006
- Hardie, B. G., Johnson, E. J., and Fader, P. S. (1993). Modeling loss aversion and reference dependence effects on brand choice. *Mark. Sci.* 12, 378–394. doi: 10.1287/mksc.12.4.378
- Heidhues, P., and Kőszegi, B. (2008). Competition and price variation when consumers are loss averse. *Am. Econ. Rev.* 98, 1245–1268. doi: 10.1257/aer.98.4.1245
- Hjorth, K., and Fosgerau, M. (2009). “Determinants of the degree of loss aversion,” in *Paper Selected for Presentation at the International Choice Modelling Conference* (Harrogate, UK).
- Iturbe-Ormaetxe, I., Ponti, G., Tomás, J., and Ubeda, L. (2011). Framing effects in public goods: Prospect theory and experimental evidence. *Games Econ. Behav.* 72, 439–447. doi: 10.1016/j.geb.2010.10.004
- Iturbe-Ormaetxe Kortajarena, I., Ponti, G., and Tomás Lucas, J. (2015). *Some (mis) Facts about Myopic Loss Aversion*. Technical report, Instituto Valenciano de Investigaciones Económicas, SA.
- Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organ. Behav. Hum. Decis. Process.* 51, 296–312. doi: 10.1016/0749-5978(92)90015-Y
- Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–292. doi: 10.2307/1914185
- Kőszegi, B., and Rabin, M. (2006). A model of reference-dependent preferences. *Q. J. Econ.* 121, 1133–1165. doi: 10.1093/qje/121.4.1133

- Kőszegi, B., and Rabin, M. (2007). Reference-dependent risk attitudes. *Am. Econ. Rev.* 97, 1047–1073. doi: 10.1257/aer.97.4.1047
- Linde, J., and Sonnemans, J. (2012). Social comparison and risky choices. *J. Risk Uncertain.* 44, 45–72. doi: 10.1007/s11166-011-9135-z
- Loomes, G., and Sugden, R. (1986). Disappointment and dynamic consistency in choice under uncertainty. *Rev. Econ. Stud.* 53, 271–282. doi: 10.2307/2297651
- Masatlioglu, Y., and Ok, E. A. (2005). Rational choice with status quo bias. *J. Econ. Theory* 121, 1–29. doi: 10.1016/j.jet.2004.03.007
- Payne, J. W., Shu, S., Webb, E., and Sagara, N. (2015). *Individual Heterogeneity in Loss Aversion and Its Impact on Social Security Claiming Decisions*. Technical report, National Bureau of Economic Research.
- Post, T., van den Assem, M. J., Baltussen, G., and Thaler, R. H. (2008). Deal or no deal? Decision making under risk in a large-payoff game show. *Am. Econ. Rev.* 98, 38–71. doi: 10.1257/aer.98.1.38
- Starmer, C. (2000). Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. *J. Econ. Lit.* 38, 332–382. doi: 10.1257/jel.38.2.332
- Starmer, C., and Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? An experimental investigation. *Am. Econ. Rev.* 81, 971–978.
- Sullivan, K., and Kida, T. (1995). The effect of multiple reference points and prior gains and losses on managers' risky decision making. *Organ. Behav. Hum. Decis. Process.* 64, 76–83. doi: 10.1006/obhd.1995.1091
- Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertain.* 5, 297–323. doi: 10.1007/BF00122574
- Vendrik, M., and Woltjer, G. B. (2007). Happiness and loss aversion: is utility concave or convex in relative income? *J. Public Econ.* 91, 1423–1448. doi: 10.1016/j.jpubeco.2007.02.008
- Von Gaudecker, H.-M., Van Soest, A., and Wengström, E. (2011). Heterogeneity in risky choice behavior in a broad population. *Am. Econ. Rev.* 101, 664–694. doi: 10.1257/aer.101.2.664

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Terzi, Koedijk, Noussair and Pownall. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Self-identified Obese People Request Less Money: A Field Experiment

Antonios Proestakis<sup>1\*</sup> and Pablo Brañas-Garza<sup>2</sup>

<sup>1</sup> Health, Consumers and Reference Materials, Joint Research Centre, European Commission, Ispra, Italy, <sup>2</sup> Economics Department, Business School, Middlesex University London, London, UK

Empirical evidence suggests that obese people are discriminated in different social environments, such as the work place. Yet, the degree to which obese people are internalizing and adjusting their own behavior as a result of this discriminatory behavior has not been thoroughly studied. We develop a proxy for measuring experimentally the “self-weight bias” by giving to both self-identified obese ( $n = 90$ ) and non-obese ( $n = 180$ ) individuals the opportunity to request a positive amount of money after having performed an identical task. Consistent with the System Justification Theory, we find that self-identified obese individuals, due to a preexisting *false consciousness*, request significantly lower amounts of money than non-obese ones. A within subject comparison between self-reports and external monitors’ evaluations reveals that the excessive weight felt by the “self” but not reported by evaluators captures the self-weight bias not only for obese but also for non-obese individuals. Linking our experimental results to the supply side of the labor market, we argue that self-weight bias, as expressed by lower salary requests, enhances discriminatory behavior against individuals who feel, but may not actually be, obese and consequently exacerbates the wage gap across weight.

## OPEN ACCESS

### Edited by:

Nikolaos Georgantzis,  
University of Reading, UK

### Reviewed by:

John Smith,  
Rutgers University Camden, USA  
Hernan Daniel Bejarano,  
Chapman University, USA

### \*Correspondence:

Antonios Proestakis  
antonios.proestakis@ec.europa.eu

### Specialty section:

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

Received: 29 June 2016

Accepted: 09 September 2016

Published: 23 September 2016

### Citation:

Proestakis A and Brañas-Garza P (2016) Self-identified Obese People Request Less Money: A Field Experiment. *Front. Psychol.* 7:1454.  
doi: 10.3389/fpsyg.2016.01454

## 1. INTRODUCTION

Obesity is a salient appearance characteristic, which can severely stigmatize individuals and provoke various forms of prejudice and discrimination in several areas, including the workplace, school, interactions with health professionals and other social settings (see Puhl and Heuer, 2009 for an extensive review). Numerous empirical studies have reported the negative effects of obesity on wages and employment rates (Cawley, 2004, 2007; Cawley and Danziger, 2005; Garcia and Quintana-Domeque, 2006; Brunello and D’Hombres, 2007; Han et al., 2009). For instance, Cawley (2004) estimated that for white females, an increase of 64 pounds above average weight was associated with a 9% decrease in wages. Han et al. (2009) found that the negative relationship between the BMI and wages is larger in occupations requiring social interactions and across older people. Brunello and D’Hombres (2007) observed that a 10% increase in the average BMI reduces the hourly wages of males by 1.9% and of females by 3.3% while Garcia and Quintana-Domeque (2006) reported a negative correlation between wages and obesity, ranging from -2 to -10% but only for women. Although weaker, the negative effects of obesity hold even when more complex measures (which are based on bioelectrical impedance analysis, e.g., total or percent body fat, fat-free mass, etc.) of obesity are employed (Burkhauser and Cawley, 2008; Johansson et al., 2009; Wada and Tekin, 2010). Evidence on discrimination attributed to obesity can also be found in experimental psychology studies (see Roehling et al., 2008 meta-analysis

on weight-discrimination). All in all, those papers showed that overweight job applicants and employees were evaluated more negatively and had worse employment outcomes compared to their non-overweight counterparts.

Unlike the bias against other minority groups (e.g., racial, ethnic, religious, etc.), negative attitudes toward overweight individuals are somehow socially accepted and even encouraged, making the stigma of obesity one of the most pervasive and persistent (Wang et al., 2004). Social Identity Theory (Tajfel and Turner, 1979) gives a plausible explanation about inter-group discrimination; distinct groups are more likely to view in-group members in a more positive light and out-group people more negatively [a result which is also experimentally confirmed, e.g., (Bernhard et al., 2006; Chen and Li, 2009); etc. and also introduced into the economic analysis in the seminal study by Akerlof and Kranton (2000)]. However, this theory does not explain in-group, anti-fat attitudes (i.e., negative attitudes and stereotypes about obese people at both the explicit and implicit level) which were documented in several empirical studies (e.g., Rudman et al., 2002; Wang et al., 2004; Crandall and Reser, 2005; Schwartz et al., 2006) and described by the System Justification (Jost and Banaji, 1994) and Social Dominance (Sidanius and Pratto, 2004) theories. More recent studies also make the crucial distinction between intra-group anti-fat attitudes (overweight individuals toward other overweighted individuals) and the internalization of weight bias toward the “self” (Puhl et al., 2007; Durso and Latner, 2008). Along these lines, we use the term “self-weight bias” to describe the internalized weight bias of overweight people toward themselves.

We hypothesize that because of the self-weight bias, obese participants will respond differently to a stimulus related to a compensation for a given task by *claiming less money for themselves*. The concept of “false consciousness” (Elster, 1982; Cunningham, 1987; Eagleton, 1991)—also central in the System Justification Theory (Jost and Banaji, 1994; Jost, 2011)—provides good theoretical grounds for our hypothesis. Obese individuals, like other marginalized groups, may develop a differential attitude due to false consciousness: *the tendency on the part of marginalized group members to implicitly accept society’s negative orientations toward their group as justification for their subordinate status* (Rudman et al., 2002). As noted before, many studies have already documented society’s negative orientation toward obese individuals (Puhl and Heuer, 2009) and their subordinate status in the workplace, as evidenced by their lower salaries. However, little is known about obese people’s implicit acceptance of their subordinate status.

In our experimental setting, subjects were asked to reveal “the amount of money they would like to request as compensation for their effort and for the information they have provided for fulfilling a questionnaire.” We expect that due to the self-weight bias, obese individuals would make on average lower monetary requests. An open-ended question was used (based on Greig, 2008), in order to reflect the salary negotiation process in a job-interview environment, where the job-candidate is asked to reveal his aspirations first. On top of the well documented wage discrimination against obese people, we suggest that a fraction of the wage gap across weight can be attributed to the lower

initial salary requests (due to self-weight bias) between obese and non-obese individuals, as these can serve as anchors in the negotiation process and influence subsequent offers and final agreements (Tversky and Kahneman, 1974; Ritov, 1996; Galinsky and Mussweiler, 2001).

In this study, using data of 270 subjects who were invited to claim money for filling in a questionnaire, we find robust evidence in favor of our self-weight bias hypothesis, namely that self-identified obese individuals claim a lower amount of money because they have implicitly accepted that they deserve less.

The common task for all participants was the completion of a 30-min (including instructions) questionnaire. Subjects were asked to self-report weight status and other appearance and personality characteristics together with other socioeconomic questions and a psychological test which was designed to distract subjects’ attention from the real focus of this study. We use this self-identified weight status (*self-weight* henceforth) to categorize participants and test our primary hypothesis. Such self-reported measure has been also used before in Bosch-Domènech et al. (2014) and was found to be highly correlated with self-reported BMI. In a similar (to ours) experimental setting including self-reported questionnaires and monetary incentives, Brañas-Garza et al. (2016) found that self-reported BMI is not related to social preferences (altruism, fairness and trust). In a study closer to our self-weight bias hypothesis (Durso and Latner, 2008), internalized weight bias (measured by Weight Bias Internalization Scale) was found to be significantly correlated with antifat attitudes, lower self-esteem, body image concern, drive for thinness and measures of mood and eating disturbance. However, in the study by Puhl et al. (2007), internalized weight bias (measured by the degree to which participants believed stereotypes to be true or false) was not related to types or amount of stigma experiences reported, self-esteem, depression, or attitudes toward obese persons.

Additionally to *self-weight*, we have asked the 27 monitors who conducted the experiment to evaluate participants’ weight (henceforth *monitors’ weight*) using the same Likert scale. A replication of the analysis using *monitors’ weight* instead of *self-weight* do not produce any significant result related to self-weight bias. Like in other studies on internalized weight bias (Puhl et al., 2007; Durso and Latner, 2008), using a self-reported measure of obesity is more relevant for approximating self-weight bias. As an additional test we compute the difference between the two measures (*self-weight* vs. *monitors’ weight*) to generate a new measure, the self-weight overstatement, which was found to be the key factor for the self-weight bias; *the excessive weight felt by the “self” but not reported by others (monitors) is a good predictor of lower money requests, not only for obese but also for non-obese individuals*.

In this study, we also attempt to shed light to the mixed findings in the literature related to the interaction between gender and weight bias. Starting from the gender literature, several studies (among others, Rosenbaum, 1984; Gerhart, 1990; Barron, 2003; Greig, 2008) have shown that women make significantly lower salary requests than men. However, when focusing on the obesity literature, the gender effect is ambiguous. While the meta-analysis by Roehling et al. (2008) showed that

both overweight men and women were equally susceptible to weight discrimination, other earlier empirical studies have found that the obesity wage penalty is stronger for Baum and Ford (2004) and Averett and Korenman (1996) or only applies to females (Register and Williams, 1990; Pagan and Davila, 1997). Similarly, contradicting gender effects have been evidenced in the *internalization weight bias* literature with a study identifying a positive effect (Lillis et al., 2010) while others no association between females and weight bias (Puhl et al., 2007; Durso and Latner, 2008). In our study, we find only “weak” evidence for gender differences in self-weight bias, in the sense that the difference in money requests between self-identified obese and non-obese females are more significant than the respective differences between self-identified obese and non-obese males.

This paper adds to the literature in number of ways: First, we develop a genuine implicit proxy for experimentally eliciting the weight bias internalization. Second, we find that only self-reported measures of obesity are relevant to self-weight bias since they capture how people feel rather than how people really are or how they look to others. Third, we find that self-identified obese individuals experience larger self-weight bias (expressed by lower money requests). Finally, we find that the self-weight overstatement, which is the difference between self-reported and external interviewer’s evaluation on subject’s weight status, is the key factor behind individuals’ self-weight bias.

After this introduction, the remainder of this study is organized as follows: the experimental methods are described in Section 2, while results are presented in Section 3. Section 4 concludes with a discussion of the results.

## 2. MATERIALS AND METHODS

We conducted an economic field experiment with 270 subjects from different socioeconomic backgrounds. Twenty seven monitors, aged between 20 and 60 years and from varying socioeconomic backgrounds were recruited to serve as monitors. All of them were students at the School of Social Work at the Universidad de Granada taking a module on *Economic Analysis of Social Work*. None of them had any previous experience with economic experiments.

### 2.1. Stage 1: Monitors’ Training and Preparations

Monitors were trained for a total of 6 h. Training included a general description of the experimental methodology with special reference to the experimental protocols of the present study. Additional instructions regarding the experiment were also given in detail. Each monitor was asked to independently recruit ten subjects to participate in an economic experiment within 1 week’s time. The monitors had no information about the research focus of the study. By doing so it was ensured that subjects were not selected on the basis of any specific characteristic, thus avoiding any potential demand effect or sample bias.

The monitors were also told that they should aim for a balanced subject pool in terms of gender and employment status. This was done because we were interested in eliciting valuations

from individuals who were in a workplace environment. After the first week, the monitors were asked to submit a list with the codified names (in order to assure anonymity) of the ten subjects they had recruited.

### 2.2. Stage 2: Questionnaires and Implementation

In the second stage, every monitor answered a questionnaire ( $Q_m$ ) describing each one of the 10 subjects she had recruited. The questionnaire consisted of three parts; Part 1: appearance and personality questions of the subjects, Part 2: Sally-Ann task (Wimmer and Perner, 1983), which was simply used as a distraction from the research focus, Part 3: Monitors described the nature of the relationship between herself and each one of her subjects.

After completing and submitting  $Q_m$  to the researchers, the monitors received 10 new questionnaires  $Q_s$  and 10 envelopes for each one of her subjects. These envelopes were delivered by them to their subjects for enclosing their private answers. Note that the first two parts of  $Q_m$  and  $Q_s$  were identical. The only difference is that the questions in  $Q_m$  were answered by each of the interviewer (describing the 10 participants) while the questions in  $Q_s$  were self-reported by each one of the 10 subjects (See Supplementary Materials for an English translation of the main parts of the  $Q_s$  questionnaire).

Since Part 2 of the questionnaires was only used to distract participants (and monitors) from the main goal of the research, we will focus here only on Part 1. It consisted of four questions about their appearance, namely obesity, beauty, height and manner of dress, and five questions about their personality characteristics, namely ambition, self-esteem, sociality, creativeness and benevolence. All these questions were ranked on a 7-level Likert scale. *Obesity* is used as an explanatory variable while *beauty*, *ambition*, and *self-esteem* are used as control variables. The remaining questions were not related to the experiment but were used to distract subjects (just like the Sally-Ann task in Part 2).

At the end of the  $Q_s$  questionnaire, in Part 3, participants were also asked regarding how much money they would like to receive for the task. Specifically, subjects were asked the following question: *How much money you would like to request as a compensation for the effort you made to fill out the questionnaire and for the information you provided us.* An alternative elicitation mode would be to ask subjects to select between, for instance, 0€, 5€, 10€, 15€, and 20€. However, this would anchor our participants’ choices. In contrast, unrestricted question mode, avoid framing subjects elicitations. In fact, requesting for a very large amount was an option, which is of interest for the study.

It was also clarified that the money available for the research project was provided by the Spanish government and did not belong to either the monitors or the researchers. As the experiment took place in the field, subjects were also asked to give their names and home addresses for receiving the money that would be paid to them. Payments were realized 2 weeks later according to the following rule (unknown to the subjects and monitors ex-ante): Subjects who requested 10€ or more,

were paid exactly 10€. All the rest received the exact amount of their request. Finally, subjects were asked whether they would be willing to participate in any other similar study and how much money the would presumably request for doing so.

### 2.3. Ethical Concerns

All participants were assured that that their anonymity would always be preserved (in agreement with the Spanish Law 15/1999 for Personal Data Protection). Subjects were informed that no association will ever be made between subjects' real names, the corresponding codes and the final results. All experimental procedures were checked and approved by the Vice-Dean of Research of the School of Economics at the University of Granada, the institution coordinating the experiment.

## 3. RESULTS

### 3.1. Data Considerations and Descriptive Statistics

Subjects' "money requests" (henceforth *requests*) is the main dependent variable under consideration. The empirical distribution of this variable has been found (**Figure 1**) not linear including many zeros (93 subjects requested 0€ and 23 gave blank answers), discontinuities, focal points (10€, 20€, 30€, 50€, 100€) and extreme values (4 values  $\geq 18,000$ €, when standard deviation of *requests* is about 5830€). For this reason a 6-category variable (henceforth *6cat\_requests*, see Supplementary

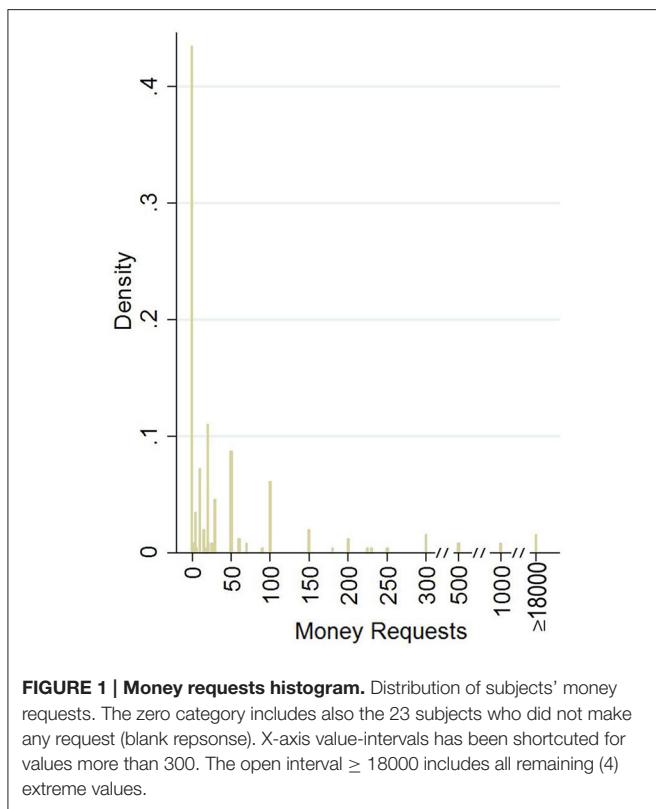


Table S1) with ordered values which are including at least one focal point of *requests* has been generated and will be analyzed in parallel with the original variable *requests*. Among others, such a transformation has the advantage of including all those extreme values which are eventually excluded as outliers from the regression analysis due to the distortion effect on the coefficients. These values are important for our analysis as they capture participants' intention to receive the highest possible stake.

Subjects were asked to fill in a questionnaire about their appearance and personality characteristics using 7-point Likert scales. Self-reported weight status (*self-weight*) was used as the main independent variable of this study. The original measure takes values from 1 (very thin) to 7 (very obese). However, for representation reasons which will become more obvious when describing the regression analysis, we separate "obese" ( $self-weight \geq 5$ , henceforth *self-obese*) from "thin" ( $self-weight \leq 3$ , henceforth *self-thin*) individuals. Self-reported beauty (1: very ugly to 7: very beautiful), self-esteem (1: no self-esteem at all, to 7: high self-esteem) and ambition (1: not ambitious at all to 7: very ambitious) are also included in the regressions analysis to control for possible confounding effects. The continuous variables *age* (and *age*<sup>2</sup>) and *wage* together with the dummy variable *female* were reported by the monitors and also used as controls variables. In the last part of the analysis, monitors' estimation of subjects' weight (*monitors' weight*) is also used for describing participants' weight status overstatement.

**Table 1** summarizes the descriptive statistics for all these variables in their original form. The subject pool was comprised of 55% females and 35% university students. About 37% of the subjects were unemployed and 18% worked in a low-paid job (i.e.,  $\leq 850$ €, corresponding to the lower quartile of our sample).

It is interesting to see that the mean, the median and the mode of the self-reported variables *beauty*, *ambition* and *self-esteem* are much higher than expected (i.e., 4, assuming a normal distribution). However, with regards to *obesity* the mean value approaches the expected one, while the mode and the median are exactly 4. This is probably due to the fact that weight status is an obvious appearance characteristic, leaving little space for subjective mis-estimations.

**TABLE 1 | Descriptive statistics.**

Variable	N	Mean	Median	Mode	Std. Dev	Min	Max
<i>self-weight</i>	269	4.18	4	4	1.05	1	7
<i>beauty</i>	269	4.79	5	5	0.97	1	7
<i>ambition</i>	269	4.52	5	5	1.34	1	7
<i>self-est.</i>	269	4.49	5	5	1.48	1	7
<i>monitors' weight</i>	270	3.65	4	4	1.43	1	7
<i>female</i>	270	0.55	1	1	0.50	0	1
<i>age</i>	270	29.33	25	24	9.47	18	65
<i>wage</i>	171	1316	1200	1500	848	100	7000

One subject did not answer the Q<sub>s</sub> questionnaire at all, reducing our sample to 269 self-reported observations. Monitors reported Q<sub>m</sub> questionnaires for all 270 subjects. The variable *wage* refers to the monthly salary of the 171 subjects who currently have a job.

### 3.2. Self-weight Bias

In this section, we will start our analysis with a graphical representation of the relation between the variables *requests* and *6cat\_requests* with *self-weight*. As a second step of analysis, we will conduct some preliminary non-parametric tests which will provide a first confirmation of our self-weight bias hypothesis. Finally, by performing linear (OLS) and non-linear (Tobit, Probit) regression analysis (Table 2), we will control for potential confounding factors and also account for some of the specific characteristics of our data (intra-group correlation, outliers, non-linearity of the dependent variable, etc.).

**Figure 2** shows the means of (**Figure 2A**) *requests* (including the 95% confidence intervals) and (**Figure 2B**) *6cat\_requests* by the 7 different levels of *self-weight*. The size of the bubble in (**Figure 2B**) is proportional to the number of people belonging to each level of *self-weight*. Note that the *self-weight* value 4 (horizontal axis) corresponds to those subjects who consider themselves neither *thin* nor *obese*.

At high values (5–7) of *self-weight*, a negative trend is already visible from the figure. This is also supported by Mann–Whitney (henceforth M-W) non-parametric tests; for the variable *requests*

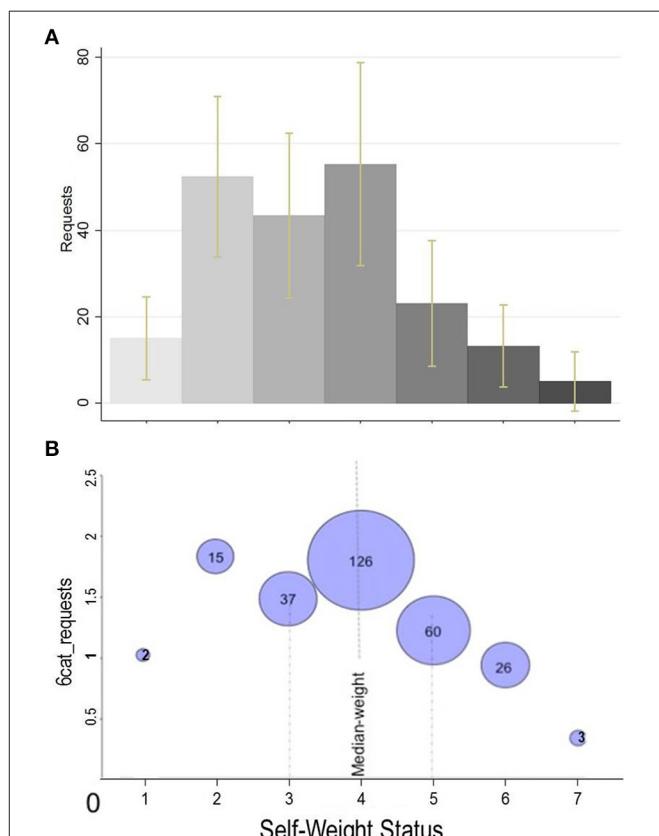
[after having dropped out the four extreme values (i.e.,  $\geq 3^*std$ ,  $n = 265$ )], we found that individuals with *self-weight* of levels 5 (*sw5*) or 6 (*sw6*) are requesting significantly less (at 5%) as compared to participants with *self-weight* level 4 (*sw4*). [M-W: (*sw4* vs. *sw5*;  $z = 2.49$ ,  $p = 0.013$ ), (*sw4* vs. *sw6*;  $z = 2.09$ ,  $p = 0.037$ )]. The same result is replicated when the variable *6cat\_requests* ( $n = 269$ ) is used [M-W: (*sw4* vs. *sw5*;  $z = 2.28$ ,  $p = 0.022$ ), (*sw4* vs. *sw6*;  $z = 2.24$ ,  $p = 0.025$ )]. We do not run any non-parametric test for *sw7* as too few observations ( $n = 3$ ) are included in this category.

On the other hand, there is no clear pattern for the average requests among self-identified thin (values 1–3) people. This is probably due to the fact that ranking of *self-weight* lower values are not really as straightforward as other variables. For instance, beauty is clearly monotonic in its self-ranking (e.g., a person of *beauty* = 6 is always considered better off than a person of *beauty* = 2). In contrast, really low values in the weight scale might be perceived equally bad as really high values (somebody can think of him/her self as too thin). It is therefore plausible to split the *self-weight* variable into two different dummies: *self-obese* taking the value of 1 if *self-weight*  $\geq 5$  and 0 otherwise,

**TABLE 2 |** Regressions on money requests.

	OLS (1)	Tobit (2)	Tobit (3)	o-Probit (4)	Probit (5)
	Requests	Requests	Requests	6cat_requests	2cat_requests
<i>self-obese</i>	−30.514*** (10.882)	−61.889*** (23.097)	−24.316*** (7.477)	−0.422*** (0.124)	−0.340** (0.150)
<i>self-thin</i>	−8.844 (15.767)	−29.150 (27.021)	−10.464 (11.595)	−0.230 (0.212)	−0.348 (0.253)
<i>female</i>	−18.545 (14.359)	−17.797 (22.687)	0.385 (10.081)	0.000 (0.170)	0.056 (0.200)
<i>beauty</i>	8.558 (11.740)	14.771 (15.863)	4.390 (5.043)	0.074 (0.078)	0.110 (0.082)
<i>age</i>	−13.776** (6.045)	−24.596** (9.825)	−7.774** (3.347)	−0.128** (0.057)	−0.133** (0.060)
<i>age</i> <sup>2</sup>	0.167** (0.080)	0.282** (0.130)	0.081* (0.044)	0.001* (0.001)	0.001* (0.001)
<i>wage</i>	0.018 (0.013)	0.012 (0.019)	0.001 (0.007)	−0.000 (0.000)	−0.000* (0.000)
<i>ambition</i>	9.297** (3.835)	15.226* (8.029)	6.123* (3.620)	0.098 (0.063)	0.060 (0.073)
<i>self-est</i>	−5.012 (7.095)	−2.859 (10.523)	1.317 (3.581)	0.027 (0.060)	0.042 (0.066)
<i>cons</i>	233.673** (109.691)	333.905** (153.168)	112.705** (53.569)		2.059* (1.107)
<i>N</i>	265	265	269	269	269
<i>R</i> <sup>2</sup> (pseudo)	0.092	0.017	0.023	0.047	0.109
Prob > <i>F/χ</i> <sup>2</sup>	0.0494	0.00958	0.00115	0.00000424	0.0000797

Standard errors (adjusted for 27 clusters in monitors) of parameters estimates in parentheses. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . (1) and (2): Four observations are excluded as outliers ( $>3^*s.d.$ ) (2) and (3): 154 left-censored observations at requests = 0. (3): 24 right-censored observations at requests  $>100$ . (4) *6cat\_requests*: six ordered values around the focal points (0, 10, 20, 30, 50, 100). Cut points are omitted. (5) *2cat\_requests*: dichotomous variable (=1 if requests  $> 0$ , 0 otherwise).



**FIGURE 2 |** Mean requests and *6cat\_requests* by self-reported obesity level. **(A)** Requests refer to the original variable (excluding outliers for requests  $\geq 18000$ ,  $n = 265$ ). Yellow bars: 95% confidence interval. **(B)** *6cat\_requests* is a 6-value ordinal transformation of the original variable ( $n = 269$ ). The size of the bubble is proportional to the number of individuals in that category.

and *self-thin* taking the value of 1 if *self-weight*  $\leq 3$  and 0 otherwise.

In the following regression analysis, money requests are regressed over these two dummy variables for facilitating presentation (main results are also replicated in Supplementary Table S1 when the original 7-point *self-weight* variable or *ob3* (*ob3* = *requests* if *requests*  $\geq 5$ , 0 otherwise) are used in the regression instead of the two dummies). The original 7-point measures of *beauty*, *ambition*, and *self-esteem*, the continuous measures *age* (and *age*<sup>2</sup>) and *wage* and the dummy variable *female* are also used in the regressions as control variables. Coefficients and standard errors (in parentheses) of all these regressors are presented in **Table 2**. We also account for the potential monitors' influence on their subjects decisions by allowing for intra-group correlation and relax the usual requirement that the observations be independent (i.e., 27 clusters for different monitors). Although monitors were specifically instructed not to influence subjects' answers, we cannot ignore that subjects may have been recruited from the monitor's proximate environment.

As robustness check, in **Table 2**, our dependent variable—money requests—is grouped and regressed in five different ways: In (1)–(3) the original variable *requests* is used. In (1) and (2) after the four extreme values ( $\geq 3^* \text{std}$ ) exclusion ( $n = 265$ ), OLS and Tobit (left-censored at *requests* = 0,  $N_{lc} = 154$ ) regressions are used respectively. In (3) all values are included ( $n = 269$ ) in the Tobit regression but eventually censored out ( $N_{rc} = 24$  at *requests*  $> 100$  and  $N_{lc} = 154$  at *requests* = 0). In (4) we use an ordered-Probit regression on the six-ordered variable *6cat\_requests* mentioned earlier. Finally, in (5) the dichotomous variable *2cat\_requests* (=1 if *requests*  $> 0$ , 0 otherwise) is regressed with a Probit model to answer the question who is more prone to request a positive amount of money.

Censoring from below in (2) and (3) seems quite plausible as zero appears as the natural lower bound, although some participants would be theoretically willing even to give money instead of receiving (alternatively, there were people willing to fill in even larger questionnaires without any compensation). This is probably the case of the 99 participants who requested 0€ not only in our main question but also when asked “*For which amount of money will you be willing to participate in a future study?*” (see Supplementary Material). Censoring from above in (3) has a post-experimental corrective scope. The open-ended question used for eliciting money-requests (Greig, 2008) has the advantage of excluding any anchoring effects (Tversky and Kahneman, 1974) but also the disadvantage of allowing really high requests which later complicate the analysis of our data. Assuming that the intention of those people requesting high stakes was simply to demonstrate that they want the highest possible payment (i.e., a person requesting 15,000€ or 500€ has the same intention with a person requesting 100€), we can also censor the data from the right. As mentioned earlier, this is also the logic behind the highest category in *6cat\_requests* which also includes the 4 outliers.

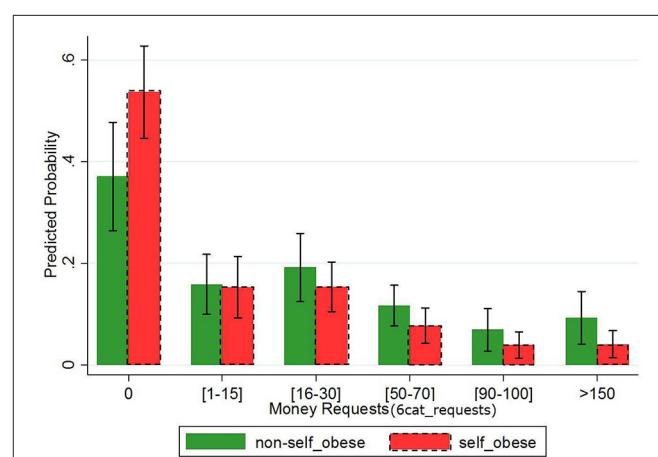
All regressions confirm the negative association between the dependent variable and *self-obese* at 1% significance level in (1)–(4) and at 5% in (5). In OLS regression (1) where coefficients have

a straightforward interpretation the result is striking: *self-obese* individuals' requests are at least 30€ less than the corresponding requests of the median group with *self-weight* = 4 (henceforth *non-obese*). Censoring out from below the 154 zero requests in (2), the linear effect of *self-obese* on the uncensored latent variable is doubled as *self-obese* individuals request almost 62€ less than the control group. When we additionally censor *requests* from above in (3) for high values ( $> 100$ ), the linear effect of *self-obese* on the uncensored latent variable is similar to the OLS result: *self-obese* individuals request 24€ less than their *non-obese* counterparts. Interestingly, this result remains significant (coefficient = 16.93,  $p_{\text{value}} = 0.002$ ) even when data is censored from above at a lower level (*requests*  $> 15$ €).

In (4) the effect of *self-obese* on *6cat\_requests* is negative and highly significant ( $p_{\text{value}} < 0.001$ ). In **Figure 3**, we present the predicted [after having performed (4)] probabilities to belong to each one of the *6cat\_requests* categories for *self-obese* and *non-obese* individuals separately, when all other predictors are fixed at their mean value. The probability of *self-obese* individual requesting 0 is 45% (i.e.,  $[pr(0|\text{self-obese})/pr(0|\text{non-obese})] - 1$ ) higher than the corresponding probability of a *non-obese* individual. At the same time, *non-obese* has 130%, 77%, and 51% more chance to fall in the category 5 (i.e., *requests*  $> 150$ ), 4 (i.e., *requests*  $\in [90 - 100]$ ) and 3 (i.e., *requests*  $\in [70 - 90]$ ), than their *self-obese* counterparts, respectively. *Self-obese* individuals' preference over zero requests and the one of *non-obese* for positive requests is exactly captured by the respective coefficient in the Probit model (5). These results are summarized as follows:

**Result 1:** *In comparison to non-obese, self-identified obese individuals request significantly less money and are more prone not to request any money at all.*

In other words, the self-weight bias hypothesis that obese people have internalized the negative attitudes toward themselves and behave in a different way than non-obese people by claiming less



**FIGURE 3 | Predicted probabilities for each category of 6cat\_requests by self-obese.** Ordered Probit predictions of *6cat\_requests* calculated for *self-obese* (red dash-framed bars) and *non-obese* (green bars) separately after having fixed all other predictors at their mean value.

or nothing is confirmed even after controlling statistically for a series of potential confounding factors.

Interestingly enough, no clear cut results are obtained when we study the self-reported measure of *self-thin* and *beauty* with any of our dependent variables and regressions. More specifically, the variable *beauty* does not capture any effect even in the absence of *self-obese* and *self-thin* variables from the models (not reported here). Moreover, the variable *self-esteem* was not found significant in any of the regressions, justifying the ambiguous role of self-esteem in Social Identity Theory. The fact that someone belongs to a “high-status” group (thin or normal-weight) may increase self-esteem but on the other hand the reason why someone is seeking to join in a group could be related either to low or to high self-esteem. Regarding the rest of the control variables, *age* is associated (negatively) with the dependent variables in a significance level lower than 5% in all regressions while *ambition* seems to have a positive effect only in OLS and Tobit regressions.

Now we turn our attention to gender effects. **Figure 4** illustrates the average *requests* or *6cat\_requests* by *self-weight* level and gender. In **Figure 4A** results are not really representative as the average requests in some obese categories

are influenced by some extreme values. This problem is eliminated with the *6cat\_requests* transformation illustrated in **Figure 4B**. Although we have not found the variable *female* significant in any of the earlier regressions, **Figure 4B** shows that the negative trend between *6cat\_requests* and *self-obese* (i.e., *self-weight*  $\geq 5$ ) is stronger in the female subsample.

However, the interaction between gender and *self-obese* (or *self-weight* or *self-ob3*, see Supplementary Table S2) is not significant in any of the OLS regressions. In absence of a direct calculation for standard errors for the interaction term (Ai and Norton, 2003) in Probit models, we repeat Probit analysis in Supplementary Table S3 for the female and male subsamples separately. Only *self-obese* females request significantly less money (both with *6cat\_requests* and *2cat\_requests*, at 1 and 5% significance level respectively) than non-obese females. In the male sample, although the negative sign holds, the variable is not significant. Result 2 is summarized as follows:

**Result 2:** *The evidence for gender difference on self-weight bias is weakly supported: The negative association between self-obese and the categorical variables 6cat\_requests or 2cat\_requests remains significant but only for the female subsample.*

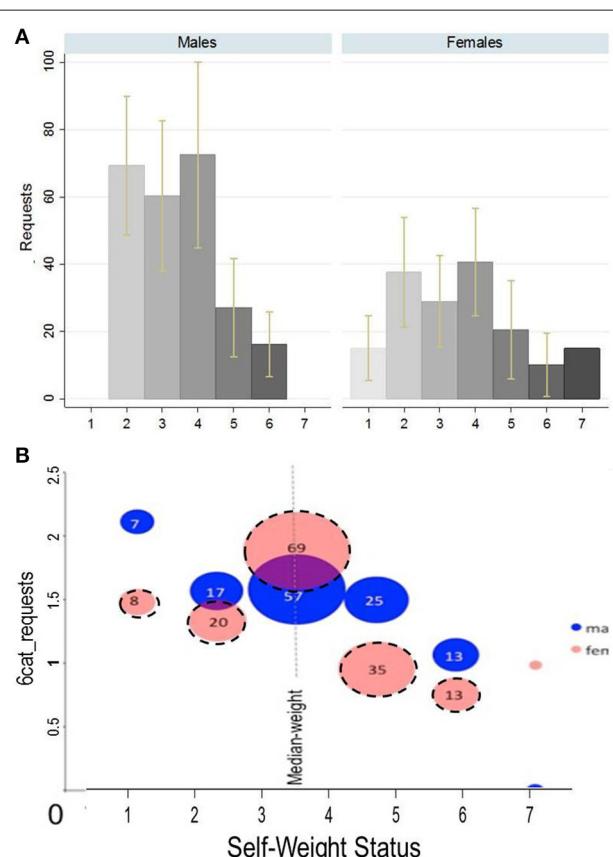
However, we do recognize that this result is partially affected by the loss of statistical power due to the restricted number of observations in the two subsamples.

### 3.3. Monitors' Evaluations and the Self-Weight Overstatement

As no traditional objective measure of obesity (actual weight, BMI, etc.) was included in our study, it is important to check the discrepancy between *self-weight*, and monitor's reports on subjects' weight status (*mon\_rep\_weight*). Interestingly, the percentage of individuals who overstate their weight status in the *self-obese* category (62%) is significantly higher than those who underestimate or accurately state it in both *self-thin* (42%) and *self-normal* (44%) categories (MW:  $p = 0.028$  and  $p = 0.010$  respectively, see also Supplementary Figure 2). We repeat OLS regressions using the monitor reported obesity variables and we find no significant effect [see **Table 3** for *mon\_rep\_obese* and also Supplementary Table S4 using *mon\_rep\_weight* and *mon\_rep\_ob3* (= *mon\_rep\_weight* if *mon\_rep\_weight*  $\geq 5$ , 0 otherwise) as main regressors]. This indicates that self-weight bias is only affected by subjective (self-reported) measures of obesity and not by others' evaluations. This result is summarized as follows:

**Result 3:** *The main determinant of the self-weight bias is the self-perceived own-weight status. Others' evaluations on subjects' weight status do not affect the self-weight bias.*

In regressions (7) and (8), we combine self-reported and monitor information in the same regressions by including the variable *weight\_overstate* and its interaction with *mon\_rep\_obese*, *mrobese* \* *overstate*. The variable



**FIGURE 4 | Mean requests and 6cat\_requests by obesity level and gender.** (A) Requests refer to the original variable (outliers are excluded). Yellow bars: 95% confidence interval. (B) 6cat\_requests is a 6-categorical ordinal variable. Red dash-framed bubbles correspond to females and blue ones refer to males. The size of the bubble is proportional to the number of individuals in that category.

**TABLE 3 | OLS on requests with monitors' reports**

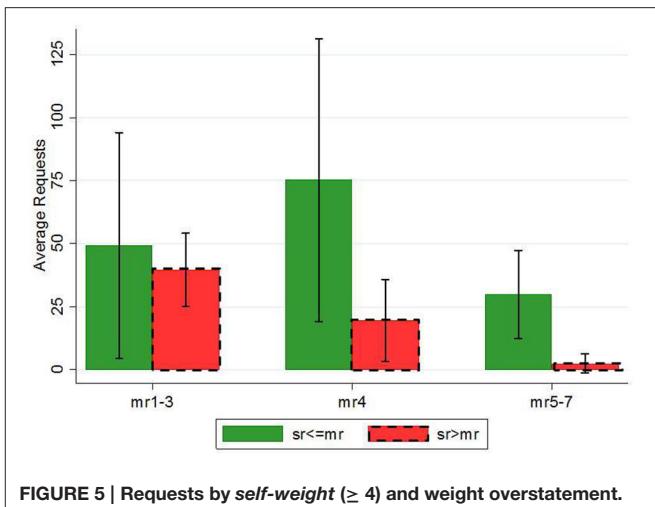
	(6) Requests	(7) Requests	(8) Requests
mon_rep_obese	−24.397 (16.748)	−31.806 (19.587)	−32.778 (22.428)
weight_overstate		−28.768** (12.080)	−29.939* (15.424)
mrobese*overstate			6.433 (23.616)
age	−14.185* (7.312)	−13.749* (6.988)	−13.715* (7.023)
age <sup>2</sup>	0.168* (0.096)	0.166* (0.092)	0.165* (0.093)
cons	297.032** (138.123)	292.520** (133.288)	292.693** (133.743)
N	264	264	264
R <sup>2</sup>	0.066	0.076	0.076
Prob > F	0.448	0.0933	0.0532

Standard errors (adjusted for 27 clusters in monitors) of parameters estimates in parentheses. \* $p < 0.10$ , \*\* $p < 0.05$ . Four observations are excluded as outliers ( $>3$ s.d.) and 1 as a missing value. All variables starting with mon\_rep\_ refer to monitors' reports. weight\_overstate takes the value 1 if self\_weight-mon\_rep\_weight  $> 0$ , and 0 otherwise. Controls based on monitors' reports mon\_rep\_thin, mon\_rep\_beauty, female, wage, mon\_rep\_ambition, mon\_rep\_self\_est are used but omitted as no significant.

weight\_overstate is a dummy variable which takes the value 1 if self\_weight – mon\_rep\_weight  $> 0$ , and 0 otherwise. In other words weight\_overstate captures all those subjects who perceive themselves more obese than their respective external evaluator. We see that weight\_overstate is significant in (7) while the mon\_rep\_obese remains insignificant. This means that self-weight bias (as approximated by money requests) is not associated with objective obesity (as evaluated by monitors) but only with the excessive weight (over monitors' estimation) which was self-reported by subjects. In particular, weight-status overstatement reduces requests by almost 29€, counterbalancing almost all the effect which was previously captured in (1) by the self-obese variable.

More importantly, the fact that in (8) the interaction term mrobese \* overstate is not significant shows that the negative effect of self-weight overstatement (henceforth overstate) applies to all weight-status levels and not only to mon\_rep\_obese individuals. **Figure 5** illustrates exactly this last result (as robustness test see also Supplementary Table S4 including the original variable mon\_rep\_weight or mon\_rep\_ob3). Although the effect is negative in all obesity levels (justifying the non-significance of overmrobese), the differences in requests between overstate and non-overstate individuals is significant (MW:  $z = 1.852$ ,  $p = 0.064$ ) only in mon\_rep\_obese (mon\_rep\_weight  $\geq 5$ ) category.

**Result 4:** The excessive weight felt by the “self” but not reported by the external evaluators determines the self-weight bias not only for obese but also for non-obese individuals.

**FIGURE 5 | Requests by self-weight ( $\geq 4$ ) and weight overstatement.**

Self-weight is considered overstated when self-report > monitor-report (red dash-framed bars). Error bars represent 95% confidence intervals.

Self-weight = 7 has been eliminated due to minor number of observations in the respective categories.

## 4. DISCUSSION

In this study, we have tested for the existence of internalized weight bias in people who self-report high weight status. Following experimental economics methodology, we have developed an implicit measure of self-weight bias by giving the same monetary incentives to both obese and non-obese persons. The experimental setting was actually simulating a salary negotiation environment in which participants were asked to state their money request for performing the same simple task. We found that self-identified obese individuals made significantly lower monetary requests as compared to non-obese. We therefore suggest that part of the obesity wage-gap is explained by obese individuals' lower reservation wages. We moreover have elicited monitors' estimations on subjects' weight status and used this information for comparison with subject's self-reports. We find that those individuals who overstate their weight status as compared to monitors' evaluation were those who were actually experiencing the self-weight bias presumably due to “false consciousness.” More importantly, we find that the self-weight bias is not only experienced by individuals who were characterized (by their monitors) as non-obese but also by those who were characterized as non-obese.

However, a different interpretation of this last result is possible, assuming that individuals' weight status is self-reported correctly but underestimated by monitors. Monitors' kindness or even sympathetic feelings especially toward obese individuals may give an alternative explanation to the self-weight bias. Monitors are more conservative to their evaluations in an attempt to be more gentle toward the sensitive (with the obesity issue) obese individuals. Regardless of the reference point and the consequent interpretation, the robust result remains the same: self-overstated or monitors' under-evaluated individuals are experiencing a self-weight bias.

To our great surprise self-esteem did not play any role in our study. Although socio-psychologists have highlighted the negative relationship between self esteem and obesity (French et al., 1995; Miller and Downey, 1999; Hesketh et al., 2004; Carr and Friedman, 2005; Biro et al., 2006), we find no association between our self-reported weight status and self-esteem variables. More importantly, self-esteem never appears significant in any of the regressions we have performed. One argument is that people who feel closely attached to an in-group are those with low self-esteem (see Baumeister and Leary, 1995) who expect to be benefited from the affiliation (Klaczynski et al., 2004). Particularly when a group has a high social standing (e.g., “thin” women), individuals with low self-esteem should seek membership benefits more often and should identify more closely with the in-group’s values than high self-esteem individuals (Bigler et al., 1997). On the other hand, the fact that someone belongs to a group may increase self-esteem due to solidarity feelings. The interaction between self-esteem and obesity becomes even more complicated when referred to low social standing groups (e.g., “obese” individuals) in which membership is not really desired.

Generally speaking, our findings are in accordance with the concept of *false consciousness*, extensively used by the System Justification Theory (Jost and Banaji, 1994). False anti-fat attitudes and stereotypes have been internalized by obese people leading to in-group devaluation and differential behavior. Along the same lines, Self-Fulfilling Prophecy Theory (Merton, 1948) would predict that obese people eventually shape their behavior in an expectancy-consistent manner, which justifies non-obese individuals’ false general beliefs and differential treatment toward obese people.

We claim that our standardized experimental setting creates the appropriate conditions for eliciting self-weight bias. The selection of a minor task to be performed minimized the opportunity cost discrepancies across individuals with different characteristics and skills (i.e., the task was equally difficult for all participants irrespectively of their weight status). At the same time the standardized monetary incentive given to participants have created equal opportunities for all of them. Thus, we have accurately measured participants’ reactions in our stimulus expressed in money requests. After controlling for other theoretically-based confounding factors, we have isolated the effect of obesity and estimated the self-weight bias.

Due to these controlled experimental conditions, we suggest that our findings can be extrapolated to other fields like in

the labor market. Without underestimating the importance of actual wage discrimination against obese people, we offer a complementary explanation to the wage gap across weight; the intrinsic tendency of obese people to claim less may result in lower salaries. We therefore conclude that discrimination in the working environment expressed by lower wages is exacerbated (rather than generated) by self-weight bias as obese people start their negotiation from an inauspicious initial position.

Such a generalization of course has its limitations. As with the vast body of experimental studies, standard criticisms of the representativeness of our subject pool apply. Furthermore, monitors’ influence on subject answers could only be controlled statistically. Another important caveat is that we model a one-shot interaction between subjects and monitors while in real life the salary negotiation process may last for longer, leaving time for both employers and candidates to readjust their strategies.

## AUTHOR CONTRIBUTIONS

The work is a product of the intellectual environment of both authors; and that both authors have contributed in various degrees to the analytical methods used, to the research concept, and to the experiment design.

## FUNDING

Financial support was received from grants by MCI (ECO2013-44879-R), the Regional Government of Andalusia (P12-SEJ-1436) and the European Commission.

## ACKNOWLEDGMENTS

We acknowledge and warmly appreciate the comments and suggestions from P. Kujal, T. García, H. Andersson, S. Neuman, G. Olcina, G. Attanasi, and A. Ebru, the participants in the FEDEA seminar (Madrid), IMEBE 2009, First Workshop in Gender Economics (Granada), 8th INRA-IDEI Conference (Toulouse).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01454>

## REFERENCES

- Ai, C., and Norton, E. (2003). Interaction terms in logit and probit models. *Econ. Lett.* 80, 123–129. doi: 10.1016/S0165-1765(03)00032-6
- Akerlof, G. A., and Kranton, R. E. (2000). Economics and identity. *Q. J. Econ.* 115, 715–753. doi: 10.1162/00355300554881
- Averett, S., and Korenman, S. (1996). The economic reality of the beauty myth. *J. Hum. Resour.* 31, 304–330. doi: 10.2307/146065
- Barron, L. A. (2003). Ask and you shall receive? Gender differences in negotiators’ beliefs about requests for higher salary. *Econ. Hum. Biol.* 56, 635–662. doi: 10.1177/00187267030566001
- Baum, C. L., and Ford, W. F. (2004). The wage effects of obesity: a longitudinal study. *Health Econ.* 13, 885–899. doi: 10.1002/hec.881
- Baumeister, R. F., and Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol. Bull.* 117:497.
- Bernhard, H., Fischbacher, U., and Fehr, E. (2006). Parochial altruism in humans. *Nature* 442, 912–915. doi: 10.1038/nature04981
- Bigler, R. S., Jones, L. C., and Lobliner, D. B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child Dev.* 68, 530–543. doi: 10.2307/1131676

- Biro, F. M., Striegel-Moore, R. H., Franko, D. L., Padgett, J., and Bean, J. A. (2006). Self-esteem in adolescent females. *J. Adolesc. Health* 39, 501–507. doi: 10.1016/j.jadohealth.2006.03.010
- Bosch-Domènech, A., Brañas-Garza P., and Espín, A. M. (2014). Can exposure to prenatal sex hormones (2D: 4D) predict cognitive reflection? *Psychoneuroendocrinology* 43, 1–10. doi: 10.1016/j.psyneuen.2014.01.023
- Brañas-Garza P., Espín, A. M., and Lenkei B. (2016). BMI is not related to altruism, fairness, trust or reciprocity: experimental evidence from the field and the lab. *Physiol. Behav.* 156, 79–93. doi: 10.1016/j.physbeh.2016.01.008
- Brunello, G., and D'Hombres, B. (2007). Does body weight affect wages? Evidence from Europe. *Econ. Hum. Biol.* 5, 1–19. doi: 10.1016/j.ehb.2006.11.002
- Burkhauser, R. C., and Cawley J. (2008). Beyond BMI: the value of more accurate measures of fatness and obesity in social science research. *J. Health Econ.* 27, 519–529. doi: 10.1016/j.jhealeco.2007.05.005
- Carr, D., and Friedman, M. A. (2005). Is obesity stigmatizing? Body weight, perceived discrimination, and psychological well-being in the united states. *J. Health Soc. Behav.* 46, 244–259. doi: 10.1177/002214650504600303
- Cawley, J. (2004). The impact of obesity on wages. *J. Hum. Resour.* 39, 451–474. doi: 10.2307/3559022
- Cawley, J., and Danziger, S. (2005). Morbid obesity and the transition from welfare to work. *J. Policy Anal. Manage.* 24, 727–743. doi: 10.1002/pam.20135
- Cawley, J. (2007). “The labour market impact of obesity,” in *Obesity, Business, and Public Policy*, eds Z. Acs and A. Lyles (Northampton: Edward Elgar Publishers), 76–89.
- Chen, Y., and Li, S. X. (2009). Group identity and social preferences. *Am. Econ. Rev.* 99, 431–457. doi: 10.1257/aer.99.1.431
- Crandall, C. S., and Reser, A. H. (2005). “Attributions and weight-based prejudice,” in *Weight Bias: Nature, Consequences, and Remedies*, eds K. D. Brownell, R. M. Puhl, M. B. Schwartz, and L. Rudd (New York, NY: The Guilford Press), 83–96.
- Cunningham, F. (1987). *False Consciousness. In Democratic Theory and Socialism*. Cambridge: Cambridge University Press.
- Durso, L. E., and Latner J. D. (2008). Understanding self-directed stigma: development of the weight bias internalization scale. *Obesity* 16, 80–86. doi: 10.1038/oby.2008.448
- Eagleton, T. (1991). *Ideology*. London: Verso.
- Elster, J. (1982). “Belief, bias, and ideology,” in *Rationality and Relativism*, eds M. Hollis and S. Lukes (Oxford: Blackwell), 123–148.
- French, S. A., Story, M., and Perry, C. L. (1995). Self-esteem and obesity in children and adolescents: a literature review. *Obes. Res.* 3, 479–490. doi: 10.1002/j.1550-8528.1995.tb00179.x
- Galinsky, A., and Mussweiler, T. (2001). First offers as anchors: the role of perspective taking and negotiator focus. *Pers. Soc. Psychol.* 81, 657–669. doi: 10.1037/0022-3514.81.4.657
- Garcia, J., and Quintana-Domeque C. (2006). “Obesity, employment and wages in Europe,” in *Advances in Health Economics and Health Services Research*, eds K. Bolin and J. Cawley (Amsterdam: Elsevier), 187–217.
- Gerhart, B. (1990). Gender differences in current starting salaries: the role of performance, college major and job title. *Indust. Labour Relat. Rev.* 43, 418–433. doi: 10.1177/001979399004300406
- Greig, F. (2008). Propensity to negotiate and career advancement: evidence from an investment bank that women are on a slow elevator. *Negot. J.* 24, 495–508. doi: 10.1111/j.1571-9979.2008.00200.x
- Han, E., Norton, E. C., and Stearns, S. C. (2009). Weight and wages: fat versus lean paychecks. *Health Econ.* 18, 535–548. doi: 10.1002/hec.1386
- Hesketh, K., Wake M., and Waters, E. (2004). Body mass index and parent-reported self-esteem in elementary school children: evidence for a causal relationship. *Int. J. Obes.* 28, 1233–1237. doi: 10.1038/sj.ijo.0802624
- Johansson, E., Bockeman, P., Kiiskinen U., and Heliovaara, M. (2009). Obesity and labour market success in Finland: the difference between having a high BMI and being fat. *Econ. Hum. Biol.* 7, 36–45. doi: 10.1016/j.ehb.2009.01.008
- Jost, J. T. (2011). “System justification theory as compliment, complement, and corrective to theories of social identification and social dominance” in *Social Motivation*, ed D. Dunning (New York, NY: Psychology Press), 223–263.
- Jost, J. T., and Banaji, M. R. (1994). The role of stereotyping in system justification and the production of false consciousness. *Brit. J. Soc. Psychol.* 33, 1–27. doi: 10.1111/j.2044-8309.1994.tb01008.x
- Klaczynski, P. A., Goold, K. W., and Mudry, J. J. (2004). Culture, obesity stereotypes, self-esteem, and the “thin ideal”: a social identity perspective. *J. Youth Adolesc.* 33, 307–317. doi: 10.1023/B:JOYO.0000032639.71472.19
- Lillis, J., Luoma, J. B., Levin, M. E., and Hayes, S. C. (2010). Measuring weight self-stigma: the weight self-stigma questionnaire. *Obesity* 18, 971–976. doi: 10.1038/oby.2009.353
- Merton, R. (1948). The self-fulfilling prophecy. *Antioch Rev.* 8, 193–210. doi: 10.2307/4609267
- Miller, C. T., and Downey, K. (1999). A meta-analysis of heavyweight and self-esteem. *Pers. Soc. Psychol. Rev.* 3, 68–84. doi: 10.1207/s15327957pspr0301\_4
- Pagan, J. A., and Davila, A. (1997). Obesity, occupational attainment, and earnings. *Soc. Sci. Q.* 78, 756–770.
- Puhl, R. M., and Heuer, C. A. (2009). The stigma of obesity: a review and update. *Obesity* 17, 941–964. doi: 10.1038/oby.2008.636
- Puhl, R. M., Moss-Racusin, C. A., and Schwartz, M. B. (2007). Internalization of weight bias: implications for binge eating and emotional well-being. *Obesity* 15, 19–23. doi: 10.1038/oby.2007.521
- Register, C. A., and Williams, D. R. (1990). Wage effects of obesity among young workers. *Soc. Sci. Q.* 71, 130–141.
- Ritov, I. (1996). Anchoring in simulated competitive market negotiation. *Organ. Behav. Hum. Decis. Process.* 67, 16–25. doi: 10.1006/obhd.1996.0062
- Roehling, M., Pilcher, S., Oswald, F., and Bruce, T. A. (2008). “The effects of weight bias on job-related outcomes: a meta-analysis of experimental studies,” in *Academy of Management Annual Meeting* (Anaheim, CA).
- Rosenbaum, J. (1984). *Career Mobility in a Corporate Hierarchy*. Orlando, FL: Academic Press.
- Rudman, L. A., Feinberg, J., and Fairchild, K. (2002). Minority members' implicit attitudes: automatic ingroup bias as a function of group status. *Soc. Cogn.* 20, 294–320. doi: 10.1521/soco.20.4.294.19908
- Schwartz, M. B., Vartanian, L. R., Nosek, B. A., and Brownell, K. D. (2006). The influence of one's own body weight on implicit and explicit anti-fat bias. *Obesity* 14, 440–447. doi: 10.1038/oby.2006.58
- Sidanius, J., and Pratto, F. (2004). *Social Dominance Theory: A New Synthesis*. New York, NY: Psychology Press.
- Tajfel, H., and Turner, J. C. (1979). An integrative theory of intergroup conflict. *Soc. Psychol. Intergroup Relat.* 74, 33–47.
- Tversky, A., and Kahneman D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124
- Wada, R., and Tekin E. (2010). Body composition and wages. *Econ. Hum. Biol.* 8, 242–254. doi: 10.1016/j.ehb.2010.02.001
- Wang, S. S., Brownell, K. D., and Wadden, T. A. (2004). The influence of the stigma of obesity on overweight individuals. *Int. J. Obes.* 28, 1333–1337. doi: 10.1038/sj.ijo.0802730
- Wimmer, H., and Perner J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 41–68. doi: 10.1016/0010-0277(83)90004-5

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Proestakis and Brañas-Garza. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Individual Characteristics vs. Experience: An Experimental Study on Cooperation in Prisoner's Dilemma

Iván Barreda-Tarazona<sup>1,2\*</sup>, Ainhoa Jaramillo-Gutiérrez<sup>1</sup>, Marina Pavan<sup>1</sup> and Gerardo Sabater-Grande<sup>1</sup>

<sup>1</sup> LEE and Economics Department, Universitat Jaume I, Castellón, Spain, <sup>2</sup> Departments of Management and Economics, Center for Experimental Research in Management and Economics (CERME), Università Ca' Foscari, Venezia, Italy

## OPEN ACCESS

### Edited by:

Mark Hallahan,  
College of the Holy Cross, USA

### Reviewed by:

Nobuyuki Takahashi,  
Hokkaido University, Japan  
Esther Kristina Diekhof,  
University of Hamburg, Germany

### \*Correspondence:

Iván Barreda-Tarazona  
ivan.barreda@eco.uji.es

### Specialty section:

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

Received: 17 June 2016

Accepted: 31 March 2017

Published: 20 April 2017

### Citation:

Barreda-Tarazona I, Jaramillo-Gutiérrez A, Pavan M and Sabater-Grande G (2017) Individual Characteristics vs. Experience: An Experimental Study on Cooperation in Prisoner's Dilemma. *Front. Psychol.* 8:596.

doi: 10.3389/fpsyg.2017.00596

Cooperative behavior is often assumed to depend on individuals' characteristics, such as altruism and reasoning ability. Evidence is mixed about what the precise impact of these characteristics is, as the subjects of study are generally randomly paired, generating a heterogeneous mix of the two characteristics. In this study we ex-ante create four different groups of subjects by factoring their higher or lower than the median scores in both altruism and reasoning ability. Then we use these groups in order to analyze the joint effect of the two characteristics on the individual choice of cooperating and on successful paired cooperation. Subjects belonging to each group play first 10 one-shot prisoner's dilemma (PD) games with ten random partners and then three consecutive 10-round repeated PD games with three random partners. In all games, we elicit players' beliefs regarding cooperation using an incentive compatible method. Individuals with high altruism are more optimistic about the cooperative behavior of the other player in the one-shot game. They also show higher individual cooperation and paired cooperation rates in the first repetitions of this game. Contrary to the one-shot PD games where high reasoning ability reduces the probability of playing cooperatively, the sign of the relationship is inverted in the first repeated PD game, showing that high reasoning ability individuals better adjust their behavior to the characteristics of the game they are playing. In this sense, the joint effect of reasoning ability and altruism is not linear, with reasoning ability counteracting the cooperative effect of altruism in the one-shot game and reinforcing it in the first repeated game. However, experience playing the repeated PD games takes over the two individual characteristics in explaining individual and paired cooperation. Thus, in a (PD) setting, altruism and reasoning ability significantly affect behavior in single encounters, while in repeated interactions individual and paired cooperation reach similarly high levels independently of these individual characteristics.

**Keywords:** altruism, cognitive ability, cooperation, prisoner's dilemma, experiment

## INTRODUCTION

Under the assumption of common knowledge of rationality and perfect information, the only Nash equilibrium of the finitely repeated Prisoner's Dilemma (PD) is mutual defection at each stage of the game. Reasoning by backward induction, a rational player's dominant strategy is to defect at the final stage, as is also the case in the one-shot game. Knowing this, each player should also defect at the second to last round, and so on, back to the first stage.

However, some cooperative play is observed, particularly at the earliest stages, in numerous experimental tests with this game (Andreoni and Miller, 1993; Cooper et al., 1996; Pothos et al., 2011, among others). One way to reconcile the theory with the experimental evidence is to assume some kind of incomplete information. If one player does not know the true payoffs of the opponent, for example, and assigns a positive probability that the other will not defect, mutual cooperation can be sustained as equilibrium (Kreps et al., 1982). One possible interpretation of the cooperation observed in experimental games, then, is that some players are "altruistic," in the sense that their true payoffs from cooperation are greater than the given monetary ones, and players' types are not common knowledge. Cooperation thus would be played by altruists. In a repeated game, an alternative explanation is that some players may try to "build a reputation" of cooperation in order to achieve a higher total payoff in the game. Both Andreoni and Miller (1993) and Cooper et al. (1996) use evidence of cooperation in the one-shot PD game as an indicator that a positive proportion of individuals are actually altruistic<sup>1</sup>. They also find that cooperation is higher when the PD is repeated for a finite number of times, consistent with reputation building. Players then turn to defection toward the end of the game, even if at a slower pace than predicted by Kreps et al. (1982). In their work, altruism is only a hypothesis to explain cooperative behavior, given that no independent measure is used to classify subjects as altruistic. However, even if altruists are expected to cooperate more, cooperation and altruism are not the same thing. Following Dreber et al. (2014) and Capraro and Marcelletti (2014), in our experiment we use as a treatment variable an ex ante measure of altruism: giving in a Dictator Game. That is, we define as altruism the willingness to sacrifice one's own payoff in order to increase the other's payoff. Furthermore, we elicit subjects' beliefs in the PD games in order to better understand the relationship between altruism and cooperative behavior.

Several laboratory experiments have been conducted to analyze whether the cooperators in a repeated Prisoner's Dilemma (RPD) can be identified by some kind of measurable characteristics. Dreber et al. (2014) find that altruism leads to more cooperation in a noisy version of the infinitely repeated PD game only if no cooperative equilibrium exists. However, altruism does not play any role in determining the outcome when cooperation can be sustained in equilibrium. Their results

support the view that social preferences are not important predictors of cooperation. Rather, individuals seem to cooperate mainly driven by payoff maximization motives. Using a dictator game to measure altruism and a standard PD game to measure cooperation, Capraro and Marcelletti (2014) find that being recipient of an altruistic act does not increase your probability of being cooperative with a third party.

We analyze the effect of altruism on cooperation, defined as the willingness to increase the joint payoffs of yourself and the other, which can be observed using one-shot and finitely repeated PD games. According to the Social Value Orientation (SVO) literature, prosocial individuals tend to maximize outcomes for both themselves and others (Van Lange et al., 1997). Eliciting beliefs about partner's cooperation allows us to tell apart whether participants classified as altruists in our study cooperate conditionally, i.e., based on the expectation that the other will also cooperate, or unconditionally, that is, even if thinking that the other will defect<sup>2</sup>.

Individuals' cognitive ability/intelligence has also been associated with cooperative play. One natural supposition is that more intelligent individuals should make more "rational" choices, exhibiting behavior consistent with game theoretic predictions, such as the sub-game perfect Nash equilibrium (Burks et al., 2009; Proto et al., 2015). Accordingly, these individuals should be observed to cooperate less in both one-shot and finitely repeated PD games. The empirical evidence, however, does not seem to support this conjecture. For instance, using a meta-study of repeated (PD) experiments run at numerous universities, Jones (2008, 2013) suggests that the average intelligence of game participants should be considered among the most robust factors driving individual cooperation. Specifically, this author finds that students at schools with higher average scores in the Scholastic Assessment Test (SAT) and the American College Test (ACT) tended to cooperate more often in a RPD<sup>3</sup>. Using a sample of 1,000 truck driving students in a one-shot sequential (PD), Burks et al. (2009) find that subjects with higher IQ more accurately forecast others' decisions and differentiate their actions more strongly given the first-mover's choice, exhibiting behavior that is far from the sub-game perfect equilibrium of the game.

Other experimental studies find mixed evidence regarding the conjecture about the negative link between cognitive ability and cooperative play. Yamagishi et al. (2014) find that decisions coherent with the maximization of self-interest are linked indeed to higher IQ. However, psychological assessment of the participants in their study leads to the conclusion that those classified as "Homo Economicus" might behave in a selfish

<sup>1</sup>In Cooper et al. (1996) a player is classified as an altruist (egoist) if he plays cooperatively more (less) than 50% of the time in 20 one-shot PD games. Andreoni and Miller (1993) consider any cooperation in the one-shot game as a sign of altruism.

<sup>2</sup>According to the SVO literature, prosocial individuals can be either altruists (unconditional cooperators) or cooperators (conditional cooperators). Differently from this literature, we do not identify altruists as unconditional cooperators.

<sup>3</sup>Jones (2014) increases the information processing necessary to implement strategies supporting cooperation of the RPD game through random switching between permutations of the payoff table. This additional strategic complexity attenuates the relationship between cognitive ability and cooperation observed in Jones (2008). In a related cognitive load experiment, Duffy and Smith (2014) find that a decrease in the cognitive load of subjects increases strategic defection near the end of the RPD game.

manner only in a situation in which no future consequence of their choice is expected. These subjects can better assess the future and adopt long-term strategies. In the same line, using an experimental design similar to ours but with just one factor (participants are allocated into two groups according to their level of intelligence), Proto et al. (2015) find that higher intelligence groups do not cooperate more in the initial rounds of an infinitely repeated PD game, but seem to learn better how to reciprocate their partner's behavior over time. However, there are no significant differences in the same design with lower continuation probability. Also recently and in contrast with Jones (2008, 2013), Al-Ubaydli et al. (2016) find that cognitive ability does not predict individual cooperation in a 10-round PD game but paired cooperation is positively correlated with the average cognitive ability of the two players. In their study, individuals with higher cognitive abilities reciprocate cooperation in the second round of the PD game significantly more than low cognitive ability subjects, like in Burks et al. (2009).

Given the previous findings, an alternative conjecture is then that more intelligent individuals better adapt to the circumstances in strategic situations<sup>4</sup>.

Our objective in this paper is to test the significance of the joint effect of cognitive ability and altruism on cooperative behavior in a series of one-shot and finitely repeated PD games. In order to do so, both characteristics are implemented as treatment variables, separating individuals in four distinct groups based on the interaction of their high/low level of cognitive ability (measured with the Differential Aptitude Test on Abstract Reasoning), and their high/low altruistic giving in a Dictator Game (DG). In the aforementioned literature, altruism or cognitive ability or both are treated as control variables rather than treatment variables, or not taken into account. Our 2 by 2 factorial design matches individuals with similar cognitive ability and level of altruism, allowing us to neatly observe the effect of these factors on cooperation. In other words, the effect of a high reasoning ability individual with high altruism might get diluted if she found for instance a low intelligence low altruism partner when playing a RPD. Our study tries to avoid this problem.

Subjects belonging to each group played 10 one-shot PD games and three 10-round repeated PD games where we elicited players' beliefs using an incentive compatible method. Our paper is the first introducing players' beliefs to analyze expectations and behavioral rules in the RPD game under different treatments of altruism and reasoning ability.

Based on the previous review, in our study we propose the following hypotheses:

**Hypothesis 1:** High altruism individuals should cooperate more in both one-shot and repeated PD.

Given our definition, an altruist should be willing to increase the other's payoff at the cost of decreasing her own expected payoff,

which is exactly what happens when an individual chooses the dominated cooperative strategy in our PD games.

**Hypothesis 2:** Individuals with higher cognitive ability should more accurately forecast their partner's actions in both types of games (one-shot and repeated), and thus be able to differentiate their behavior accordingly.

We assume that making better predictions is a necessary precondition to adapt successfully to a strategic situation. In line with Proto et al. (2015), we consider that more intelligent individuals should be capable of better assessing and adapting to the environment. Thus, they should better realize the scope for reputation building in the repeated game as opposed to the one-shot game.

**Hypothesis 3:** Reasoning ability should counteract the effect of altruism in the one-shot game, while it should reinforce it in the repeated PD game.

Our first two hypotheses propose that, while altruism should always increase cooperation, reasoning ability should lead to increased or decreased cooperation depending on the circumstances. This implies a non-linear interaction between the two factors.

Our results confirm the two first hypotheses using a clean experimental design. Reasoning ability is found to indeed counteract the effect of altruism in the one-shot games, but to reinforce it only in the first RPD. In general, the effect of the individual characteristics on the cooperation decision fades out with the repetition of the RPD game.

The article is organized as follows: Section Methods describes the experimental design and Section Results presents the results. Section Discussion discusses the results and concludes.

## METHODS

We turn to experimental economics methodology to create a controlled, saliently motivated and replicable environment in which to test our hypotheses. As a first step, we used an experimental setting to measure our subjects' reasoning ability and altruism. After creating four different groups according to the results of these measures, we invited again the same subjects to the lab for a different experiment. In this second step, subjects were randomly paired with other subjects of similar reasoning ability and altruism, without them knowing this information, and played four sets of (PD) games both one-shot and repeated. Thus, each subject whose data we present in this study has participated in two sessions in different days of two consecutive weeks in December 2014: all sessions of the second experiment were carried out during the week after the last session of the first experiment. As the participants did not receive any payment up to the end of the second session, the attrition rate was low: out of 178 subjects who participated in the first set of sessions, only 16 did not participate in the second set of sessions. Subjects were recruited among undergraduate students from different degrees at Universitat Jaume I (Spain), using ORSEE (Greiner, 2015). At the beginning of each session, subjects were given written

<sup>4</sup>Intelligence and adaptive behavior are found to be separate but related constructs exhibiting low to moderate correlations depending on the particular measures (Harrison, 1987; Keith et al., 1987; Platt et al., 1991). The underlying mechanism behind the relationship between intelligence and adaptive behavior is out of the scope of our paper.

instructions, which were also read aloud by the organizers. Any remaining questions were privately answered.

At the end of the second session, subjects found out their actual gains and were privately paid in cash the total amount obtained in both sessions. Average earnings were around 11€ for the first experiment and around 14€ for the second one, and the sessions lasted 1 and 1 h and a half, respectively. Experiments were computerized and carried out in a specialized computer lab (LEE at Universitat Jaume I), using software based on the Z-Tree toolbox by Fischbacher (2007).

Each of the two experimental designs is described in detail in the following subsections. Experimental instructions can be found in Section 1 of the Supplementary Material.

## Testing for Reasoning Ability and Altruism

In the first experimental setting, subjects were asked to complete two tasks. The first task consisted in completing the Abstract Reasoning part of the Differential Aptitude Test for Personnel and Career Assessment (DAT-AR for PCA, Bennett et al., 1974). The Abstract Reasoning (AR) scale of the DAT used in this experiment is included in the DAT-5 Spanish adaptation by the publisher TEA (Cordero and Corral, 2006). This test is usually used as a non-verbal measure of reasoning ability and involves the capacity to think logically and to perceive relationships in abstract figure patterns. It is considered as a marker of fluid intelligence (Colom et al., 2007), the component of intelligence most related to general intelligence or g factor (McGrew, 2009). The advantage of this test is that it is quite fast to implement: it is comprised of 40 multiple-choice items and has a 20 min time limit. Subjects were informed that they would receive 0.25€ for each right answer.

The second task included a Dictator Game where each subject played both as dictator (which we more neutrally called "sender") and recipient, and then was randomly assigned one of the two roles. An endowment of 10€ was provided to dictators, who could transfer any amount from 0 to 10€ to their respective anonymous recipient in increments of 0.1€. Subjects were informed that in this task the recipient would receive no payment other than the one they chose to give. In our analysis we use the amount given in the dictator game as a measure of subjects' altruism. The dictator game is positively correlated to altruistic acts in real-life situations (returning money to subjects in Franzen and Pointner (2013) using the misdirected letter technique), charitable giving (Benz and Meier, 2008) and willingness to help in a real-effort task (Peysakhovich et al., 2014). Additionally, Carpenter et al. (2008) find that the specific survey questions for altruism used in their study are positively correlated with DG giving. Using a related concept, Capraro et al. (2014) find benevolence to be correlated with cooperative behavior, but their definition of benevolence "to increase the benefit of someone else beyond one's own" has no cost to the "benevolent" player. We consider that a person acts altruistically if she unilaterally pays a cost  $c \geq 0$  to increase the benefit of someone else. More formally, Player 1 is altruist toward Player 2 if she prefers the allocation  $(x_1 - c, c)$  to the allocation  $(x_1, 0)$ , where  $c > 0$ . The larger the  $c$ , the more altruist we consider this subject to be.

After completing the aforementioned tasks, subjects were divided in four groups according to their reasoning ability and altruism and called again to the lab. Apart from 16 who decided not to continue with the second session and just came separately to the lab to get their gains in the first session, the rest continued. A subject was classified as "high altruism" if she chose to transfer more than the median transferred amount in the dictator game, and as "high reasoning" if her score was higher than the median score in the DAT-AR test. Following this classification, the final four treatment groups are named "Low Altruism and Low Reasoning" (LALR, 42 subjects), "Low Altruism and High Reasoning" (LAHR, 46 subjects), "High Altruism and Low Reasoning" (HALR, 42 subjects) and "High Altruism and High Reasoning" (HAHR, 32 subjects). Therefore, a total of 162 subjects (81 pairs of players) took part in the PD sessions. Subjects were not aware at any point of the existence of the four treatments. We could not control the gender composition of each treatment but it turned out quite balanced, always in the 60–40% of females range. In **Table 1** we summarize the treatments implemented.

## PD Games

We organized 8 PD sessions, 2 for each treatment group. Each PD session began with training questions on the PD to make sure that players fully understood the mechanism of the game. Then, subjects belonging to the same treatment group were faced with four consecutive PD tasks. Subjects were informed that they would be paid according to their decisions in only one of the four tasks, randomly selected at the end of their session.

### One-Shot PD Games

The first task consisted in a sequence of 10 one-shot PD games against potentially different anonymous opponents using a strangers-pairing mechanism. No player knew the identity of the player with whom she was currently paired or the history of decisions made by any of the other players.

**Table 2** shows the payoffs of the one-shot PD game. In each cell, the first (second) figure denotes the payoff in euros of player 1 (2). Clearly from the Table, "A" represents the decision to cooperate and "B" not to cooperate.

**TABLE 1 | Treatments summary.**

Treatment	Subjects	Female (%)	Altruism	Reasoning ability
LALR	42	43	Low	Low
LAHR	46	40	Low	High
HALR	42	59	High	Low
HAHR	32	47	High	High

**TABLE 2 | Payoffs of the one-shot game.**

Player 1	Player 2	
	A	B
A	(20, 20)	(0, 28)
B	(28, 0)	(10, 10)

In order to avoid endowment effects across the one-shot games in this task, we used the RLI (Random Lottery Incentive) system as payment mechanism. That is, if this task was selected for payment, only one randomly drawn PD game was remunerated. We didn't randomize task order and made all players play this task first, so that subjects could face a great number of opponents (up to 10 different ones) and in this way get some information about the population of players that they were facing.

### Finitely Repeated PD Games

In the last three tasks participants played a repeated PD game, in which each subject played 10 rounds of the same game with a given participant using a partners-pairing mechanism. Therefore, each subject played 10 consecutive rounds with the same opponent. Players were then anonymously re-matched with new opponents and played a new RPD lasting again 10 rounds. At the end of each period in a repetition, subjects were shown what their opponent had played. However, when players were re-matched, they were not told anything about the history of play of their new opponent.

The payoffs of each round for all three RPD tasks are shown in **Table 3**. It can be observed that they are just equal to those of a round of the one-shot game divided by ten.

### Beliefs

In order to gather more detailed information on players' strategic reasoning, subjects were asked the following questions before each round of each game:

- 1.- "Do you think your partner will choose A or B this period?"
- 2.- "What percentage of players will choose to play A this period?"

With the first question we elicit the "individual" belief and with the second one the "social" belief on individual cooperation.

Subjects could earn up to two additional euros for these questions, according to their answers<sup>5</sup>.

## RESULTS

Before reporting the detailed results related to cooperation behavior in the (PD) tasks, we first describe the outcomes of the reasoning ability test and of the Dictator Game, and subjects' beliefs in the PD tasks.

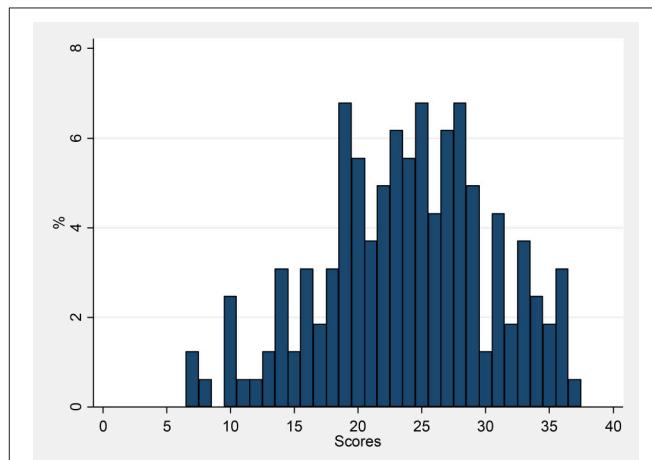
### Descriptive Statistics

**Figure 1** presents the distribution of the number of observed correct answers to the 40 multiple choice items in the DAT-AR test. The mean and the median number of right answers were 23.9 and 24 out of 40, respectively, and the standard deviation was 6.7. Mean and median number of correct answers are almost

<sup>5</sup>At each round of the one-shot PD, subjects received 1€ for answering the first question correctly and 1€ minus as many cents as the difference (in absolute value) between their answer to question 2 and the actual percentage of players choosing cooperation in that round. At each round of the repeated PD the stakes were one tenth of the one-shot PD, that is 0.1€ gain, and one tenth of the difference penalty.

**TABLE 3 | Payoffs of the RPD game.**

	Player 1		Player 2	
	A	B	A	B
A			(2, 2)	(0, 2.8)
B			(2.8, 0)	(1, 1)



**FIGURE 1 | Scores observed in the DAT-AR test.**

identical to the ones calculated for the Spanish population of a comparable age (Cordero and Corral, 2006).

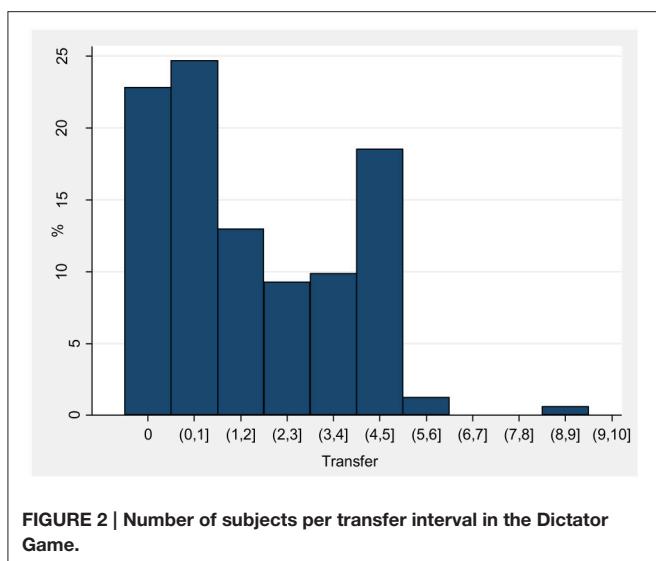
**Figure 2** shows the distribution of the transfers in the Dictator Game. About 80% of our subjects gave non-zero amounts. The mean and median transfer were of 2 and 1.4€ out of 10€, respectively, and the standard deviation was almost 2€. Comparing these results with the range of outcomes in the dictator game meta-analysis of Engel (2011), our values are within the range of what is typically observed (dictators on average give 28.35% of the pie).

**Table 4** shows descriptive statistics on reasoning ability and altruism for subjects included in the four treatment groups. On average, "high" altruism subjects transfer about 3€ more than "low" altruism ones, while subjects with "high" reasoning ability answered correctly to about 10 additional questions with respect to subjects with "low" reasoning ability. Comparing these results with the general ones for Spain from Cordero and Corral (2006), 19 correct answers correspond to about the 25% percentile of the DAT-AR scores distribution, and 29 correct answers to about the 75% percentile.

For the pooled data, there is a significantly negative correlation between altruism and reasoning ability, but it is quite low (Spearman's rho of  $-0.17$ ,  $p = 0.032$ ). Besides, the correlation between the two characteristics is not significant within each group. However, we test for collinearity in our regression analysis.

### Beliefs

**Figure 3** shows the percentage of participants whose belief is that their partner will cooperate in that particular period (the



**FIGURE 2 | Number of subjects per transfer interval in the Dictator Game.**

**TABLE 4 | Altruism (A) and Reasoning ability (R) descriptive statistics by treatment.**

	Mean		S.D		Min		Max	
	A	R	A	R	A	R	A	R
LALR	0.45	18.74	0.52	4.85	0.0	7	1.5	24
LAHR	0.43	29.67	0.53	3.33	0.0	25	1.5	37
HALR	4.15	18.62	1.24	3.99	2.0	8	8.2	24
HAHR	3.52	29.09	1.19	3.77	2.0	25	6.0	36

“individual belief,” that is, the answer to question 1 reported in Section Beliefs above) by task, period and treatment. In the one-shot game high altruism individuals with low reasoning ability (HALR) have a higher expectation of partner cooperation than the rest. This difference is significant for the first seven periods when we compare HALR vs. LALR (with the exception of period 6) and HALR vs. LAHR using a proportion test, and for the first period when we compare HALR vs. HAHR. The full test statistics are presented in Table SM2.1 in the Supplementary Material (all our tests  $p$ -values have been Bonferroni corrected to take into account the problem of false positives in multiple comparisons).

In the first period of each RPD task we observe that HALR individuals continue to have the most positive expectations about partner cooperation, while LAHR subjects are the most pessimistic, this difference being significant for tasks 2, 3, and 4 (see the proportions tests results in Tables SM2.2–SM2.4 in the Supplementary Material). However, these treatment differences level off over time within each RPD game.

On average over all periods in a task, high reasoning ability subjects have a lower expectation of partner's cooperation in the one-shot game (Mann-Whitney test  $z = -4.034$  and  $p = 0.0001$ ), while there are no significant differences in expectations in the repeated PDs. This shows that HR individuals' beliefs are more consistent with the Nash equilibrium of the game, but only in the one-shot.

The mean percentage of individuals expected to cooperate in each period (the “social belief,” that is, the answer to the second question reported in Section Beliefs), shows a similar pattern to that of the individual belief (see Figure SM2.1 in the Supplementary Material).

The elicitation of beliefs allows us to measure the number of individuals who have correctly guessed their partner's behavior in any given period, that is, they expected cooperation and the other has indeed cooperated, or they expected defection and the other has defected. Dividing this number by the total number of individuals in the treatment, we obtain the percentage of correct beliefs for each task, period and treatment (presented in Figure 4). According to Hypothesis 2 in the Introduction, we should observe that individuals with higher cognitive ability better forecast their partner's behavior. The percentage of correct individual beliefs is significantly higher for high reasoning ability subjects in the first four repetitions of the one-shot game (see Table SM2.5 in the Supplementary Material) and in the first period of task 2. In particular, LAHR participants reach 100% accuracy in almost half of the periods in all tasks, more often than the other treatments. However, there are no systematic differences in the remaining periods and tasks (Tables SM2.6–SM2.8 in the Supplementary Material). In the RPD tasks, the percentage of correct guesses is above 80% for most periods, for all treatments.

**Result 1:** *High cognitive ability subjects better forecast their partner's behavior in the first repetitions of the one-shot games and at the beginning of the first RPD. However, there are no systematic differences in the percentages of correct guesses in the remaining repetitions of the RPD.*

Notice that high altruism individuals with low reasoning ability less accurately forecast their partner's behavior in task 1. This is consistent with the fact that they have a too optimistic view of their partner's behavior in the one-shot game.

## Individual Cooperation in Period 1 of Each Task

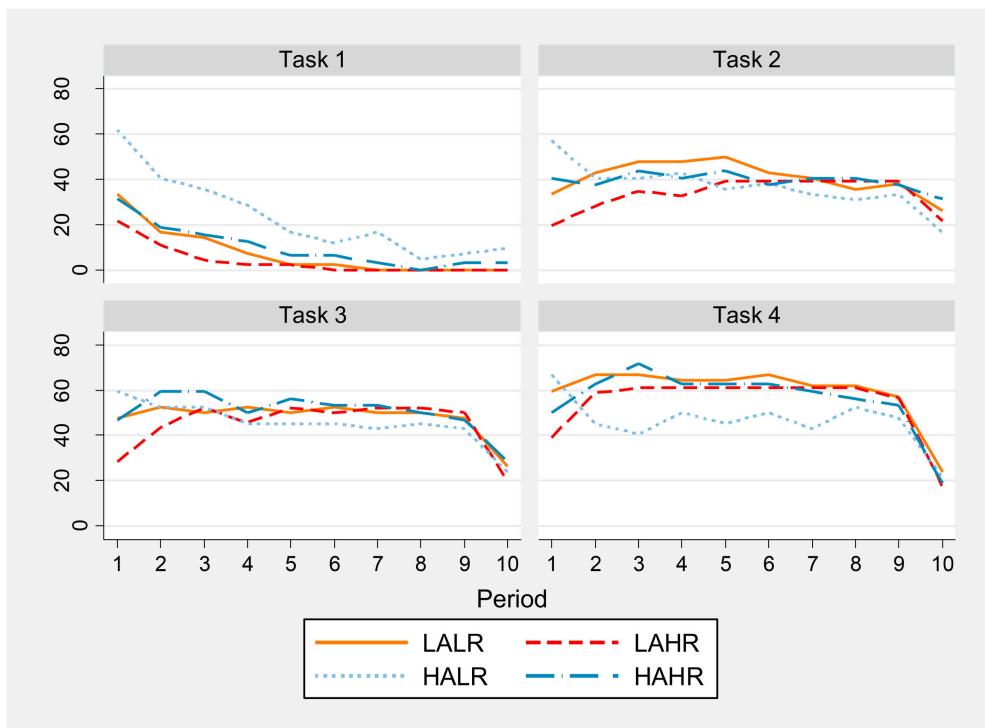
In Figure 5 we present the percentage of subjects choosing to cooperate in period 1 for each task and treatment.

The observed level of cooperation in the very first one-shot PD game depends on both altruism and reasoning ability.

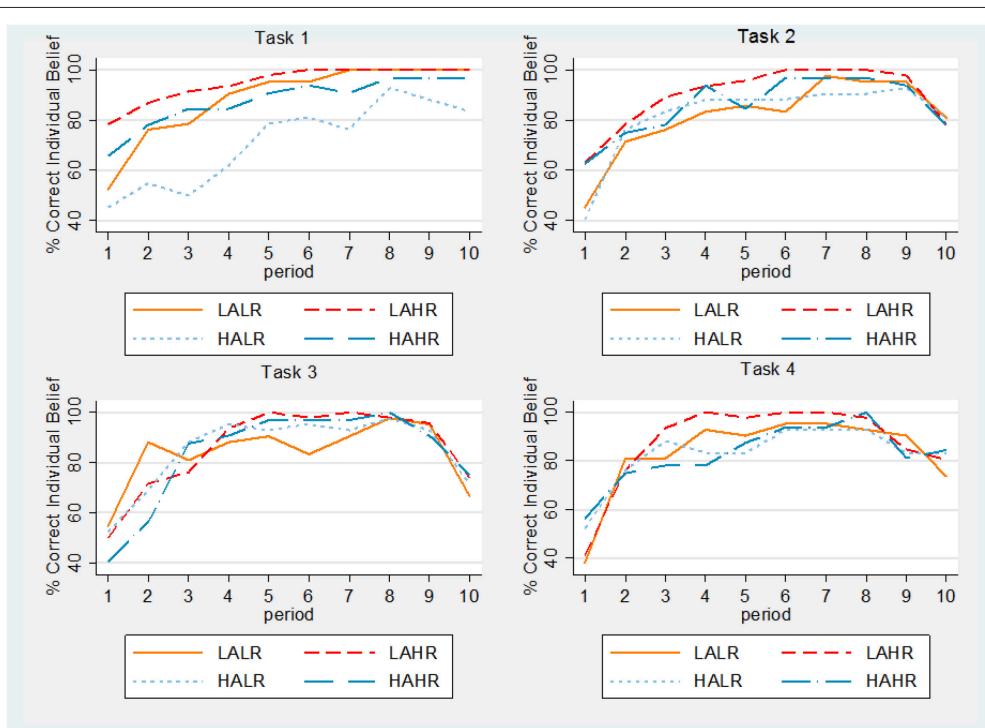
**Result 2:** *In the first PD game altruism tends to increase cooperation while reasoning ability tends to decrease it.*

Coherently with our Hypotheses 1 and 3, in the first one-shot PD game high altruism subjects cooperate more than low altruism subjects, and high reasoning ability subjects cooperate less than low reasoning ability ones. These differences are significant using a proportion test, as reported in SM2.12 (period 1).

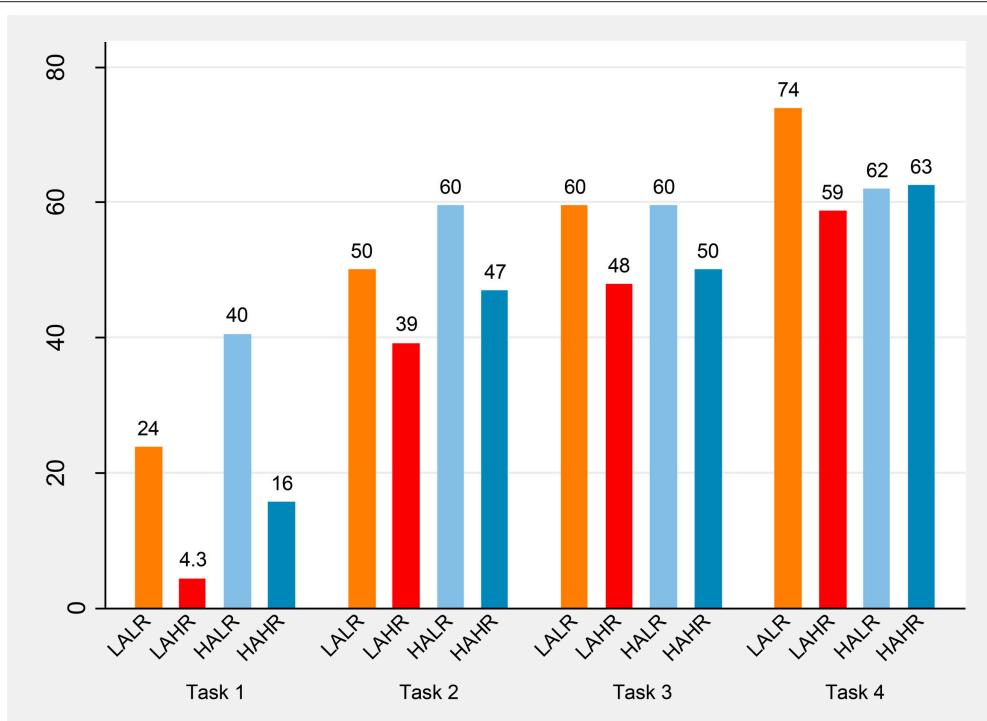
**Result 3:** *Individual cooperation rates are higher at the beginning of RPD games than at the beginning of the sequence of one-shot PD games, particularly for high reasoning ability subjects.*



**FIGURE 3 |** Percentage of individuals whose belief is partner cooperation in the current period by task, period and treatment.



**FIGURE 4 |** Percentage of correct individual beliefs by task, period and treatment.



**FIGURE 5 | Percentage of individuals cooperating in period 1 of each task.**

Using a proportion test we obtain that the percentage of individuals cooperating in period 1 is significantly higher in all repeated PD tasks than in task 1 for all treatments with the exception of the HALR treatment (see Table SM2.9 in the Supplementary Material). After a significant increase in first period cooperation from task 1 to task 2 especially for high reasoning ability subjects, the cooperation level remains stable at the beginning of the remaining tasks. Consistently with our Hypothesis 2, we observe a more marked difference in behavior between the one-shot and the repeated tasks for high reasoning ability individuals.

The observed differences in cooperation for the first one-shot PD game are no longer significant for the first period of each repeated game. The high reasoning ability subjects, who cooperated significantly less at the beginning of the one-shot games, show no significantly lower cooperation levels at the beginning of the subsequent tasks (tests results are available upon request). High reasoning ability individuals seem to better anticipate the lower cooperation rate that will be attained in a series of one-shot games with different partners as opposed to a sequence of repeated interactions with the same partner.

## Individual Cooperation Dynamics

**Figure 6** shows individual cooperation percentages by task, period and treatment.

The percentage of cooperation decreases for all treatments as the one-shot PD game is repeated (task 1). However, the group

with higher altruism and lower reasoning ability never reaches a 0% individual cooperation rate (the other treatment groups reach 0% individual cooperation in periods 5 to 9). Table SM2.10 in the Supplementary Material shows percentages of individual cooperation in the repetitions of the one-shot game, for all treatments.

Using a proportion test, in Table SM2.12 in the Supplementary Material we show that high reasoning ability participants (HR) cooperate significantly less in the one-shot PD game than low reasoning ability ones (LR) in the first two repetitions (column 1). Additionally, the percentage of cooperation is significantly higher for high altruism subjects (HA) than for low altruism ones (LA) for several periods, as can be seen in column 4.

As can be observed in **Figure 6**, in the RPD tasks individual cooperation not only is higher at the beginning but also sustained at around 40% to 60% until the very last period, when it falls abruptly (see details in Table SM2.11 in the Supplementary Material). However, last period individual cooperation rates are still positive, differently from task 1, for most treatments. No significant treatment effects appear in the RPD tasks, as we had already observed in our analysis of period one.

## Regression Analysis

In order to account for the effect of beliefs and of the stage game repetitions within each task together with the treatment, we run random-effects panel logit regressions. Results are reported in **Table 5**.

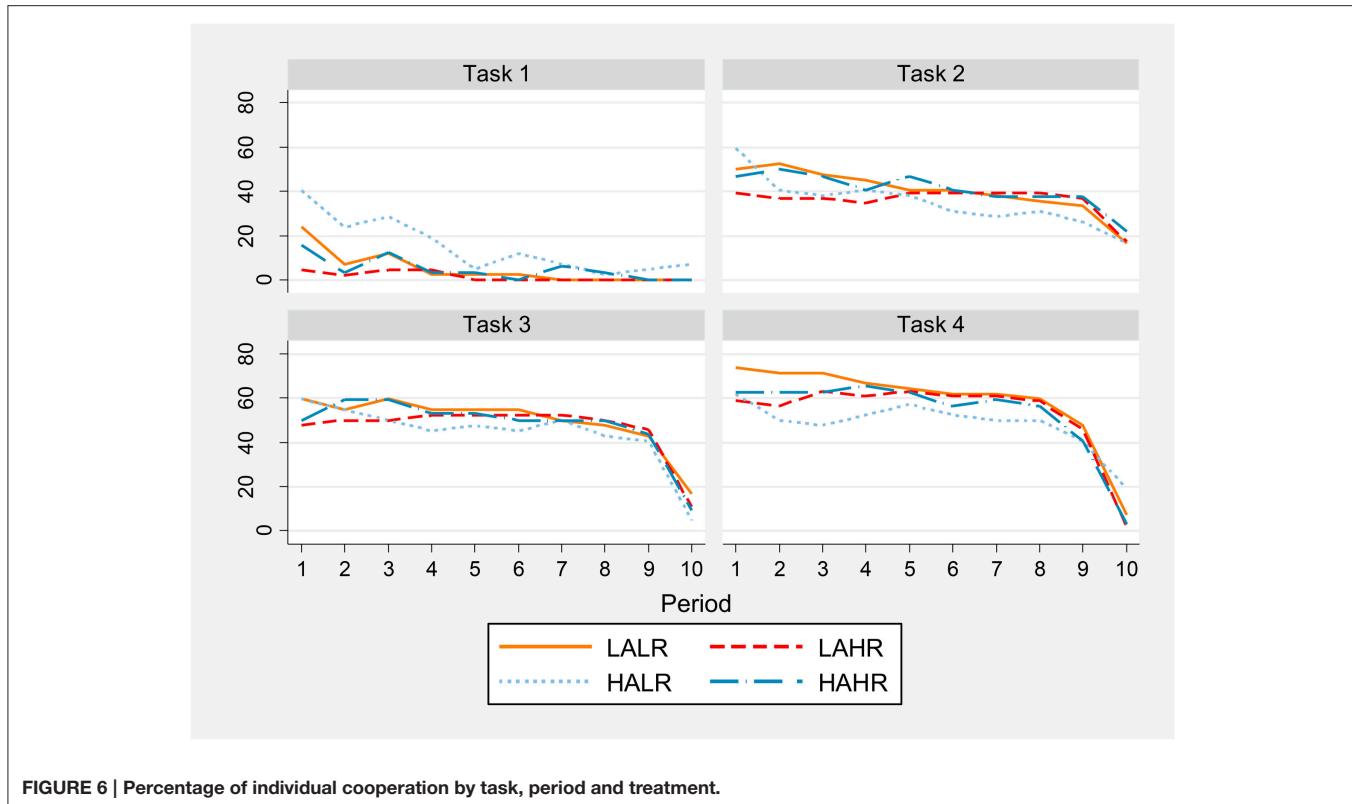


FIGURE 6 | Percentage of individual cooperation by task, period and treatment.

TABLE 5 | Random-effects panel logit regressions of individual cooperation on treatment, period and beliefs.

Individual cooperation	Task 1		Task 2		Task 3		Task 4	
	Coeff.	Sd.E.	Coeff.	Sd.E.	Coeff.	Sd.E.	Coeff.	Sd.E.
Social belief	0.02***	(0.01)						
Individual belief			5.31***	(0.31)	5.15***	(0.28)	5.19***	(0.29)
Female	0.23	(0.34)	0.18	(0.36)	0.01	(0.33)	-0.38	(0.33)
Period	-0.41***	(0.05)	-0.32***	(0.04)	-0.31***	(0.04)	-0.37***	(0.04)
LAHR	-1.52**	(0.60)	0.15	(0.49)	-0.23	(0.43)	-0.39	(0.44)
HALR	1.13***	(0.43)	-0.33	(0.50)	-0.52	(0.45)	-0.30	(0.46)
HAHR	-0.24	(0.52)	0.28	(0.53)	-0.64	(0.48)	-0.60	(0.49)
Constant	-2.62***	(0.46)	-1.64***	(0.42)	-0.84**	(0.39)	-0.17	(0.41)
N	1620		1620		1620		1620	
Wald Chi <sup>2</sup>	96.72***		295.07***		341.80***		320.63***	

\*\*\*Coefficient significant at 1%, \*\*Significant at 5%. Standard errors in parentheses.

The variables used are the following:

- Individual cooperation: dependent variable. Takes value 1 when the individual decides to cooperate in the current period, 0 otherwise.
- Social belief: individual expectation on the percentage of subjects cooperating in the current period and session. Ranges from 0 to 100.
- Individual belief: takes value 1 if the individual expects the partner to cooperate in the current period, 0 otherwise.
- Female: takes value 1 if the subject is female, 0 otherwise.
- Period: takes values 1 to 10 in each task.

- LAHR/HALR/HAHR: dummy variables that take value 1 for the corresponding treatment, 0 otherwise.

In the regression for task 1 (the one-shot PD game) we consider “social belief” more appropriate than “individual belief” as a regressor, given that the individual is not always playing with a same partner.

The baseline treatment is “Low Altruism and Low Reasoning” (LALR). Within the “Low Altruism” subjects, the treatment with “High Reasoning” (LAHR) shows significantly lower cooperation in the one-shot PD game. On the opposite, a high level of altruism significantly increases the probability of cooperating for

individuals characterized by “Low Reasoning” ability (HALR vs. the baseline LALR). The joint effect of high reasoning ability and high altruism appears to be null. In fact, there are no significant differences in cooperation between HAHR and LALR subjects, which could be due to the fact that the effects of a higher reasoning ability and a higher altruism go in opposite directions. This is coherent with the interaction effect we anticipated in Hypothesis 3.

We also observe that the higher the expectation on the percentage of players cooperating in that round, the higher individual cooperation. Moreover, each additional period significantly reduces the likelihood of cooperation. Gender has no significant effect.

Treatment effects disappear in the RPD tasks: none of the estimated coefficients for each of the three treatment dummies is significantly different from zero. In these tasks, thinking that the partner will cooperate significantly rises the probability of cooperation. There is a negative significant effect of period.

We can directly include reasoning ability and altruism measurements in these regressions rather than using a dummy for each group. Results are reported in **Table 6**. The variables used to measure reasoning ability and altruism are the following:

- Reasoning ability: number of correct answers in the DAT-RA test. Ranges from 7 to 37.
- Altruism: euros transferred to the recipient in the dictator game. Ranges from 0 to 8.2.

Although the correlation between reasoning ability and altruism was weak, we tested for collinearity in the estimated models. Results of these tests are reported in Table SM2.13 in the Supplementary Material. The Variance Inflation Factors are quite low (slightly above 1) for all regressors, indicating that there is no cause for concern.

For task 1 we obtain that reasoning ability has a significant negative effect while altruism increases the likelihood of cooperating, thus extending our Result 2 beyond the first period to all the one-shot PD games. The effect of the remaining variables is robust to the replacement of the treatment dummies by cognitive ability and altruism variables.

**Result 4:** *In the one-shot PD games, the effect of reasoning ability on the likelihood of cooperation is negative while that of altruism is positive. Additionally, individual beliefs and period also significantly affect the cooperation decision. Gender is not relevant.*

In task 2 reasoning ability continues to be significant for explaining cooperation. However, note that the direction of the effect is the opposite, that is, higher abstract reasoning leads to less cooperation in the one-shot PD and to more cooperation in RPD, thus confirming our Hypothesis 3. As we pointed out above, it seems that subjects with higher reasoning ability better recognize the different nature of the games played and the relatively lower opportunities of coordinating on cooperation that playing with a changing partner provides. Thus, these subjects seem to better adjust their behavior to the environment.

**Result 5:** *The effect of reasoning ability on cooperation is negative in the one-shot games but positive in the first RPD task.*

In tasks 3 and 4 neither reasoning ability nor altruism affect cooperation. Instead, the belief that the partner will cooperate significantly increases the likelihood of cooperating in all tasks. In fact, this belief turns out to be highly correlated with past partner cooperation (which we have not included in the regression for this reason: Spearman's rho of 0.76,  $p < 0.001$ ). Again, period has a significantly negative effect and gender plays no role.

**Result 6:** *Experience with the RPD game takes over individual characteristics of the subjects in explaining their decision.*

While reasoning ability significantly predicts cooperation behavior the first time the repeated game is played (task 2), individual characteristics do not seem to play a role when participants gain experience facing the RPD a second and a third time (tasks 3 and 4).

### Unconditional Cooperation

Using the information on beliefs, we computed the percentage of individuals who cooperate “unconditionally,” that is, even if expecting defection, for each period of each task. The result is that very few individuals choose to cooperate thinking that the partner will defect. In the one-shot, on average only 1.5% of low altruism and 2.8% of high altruism participants' decisions are A/B. In the repeated tasks, on average <6% of both high and low altruism subjects' decisions are unconditionally cooperative. We interpret this result as evidence of very low unconditional cooperation. In fact, taking into account the payoff table of the game, we can observe that even a high altruism subject would find it hard to cooperate unconditionally. On average high altruism subjects were willing to sacrifice 4€ out of 10€ in the dictator game, while in the one-shot PD they should give up 10€ and get nothing if they cooperate thinking that the partner is not going to cooperate. In fact no player gave up the whole 10€ endowment in the DG.

**Result 7:** *There is scarce evidence of unconditional cooperation, even for high altruism subjects.*

### Paired Cooperation

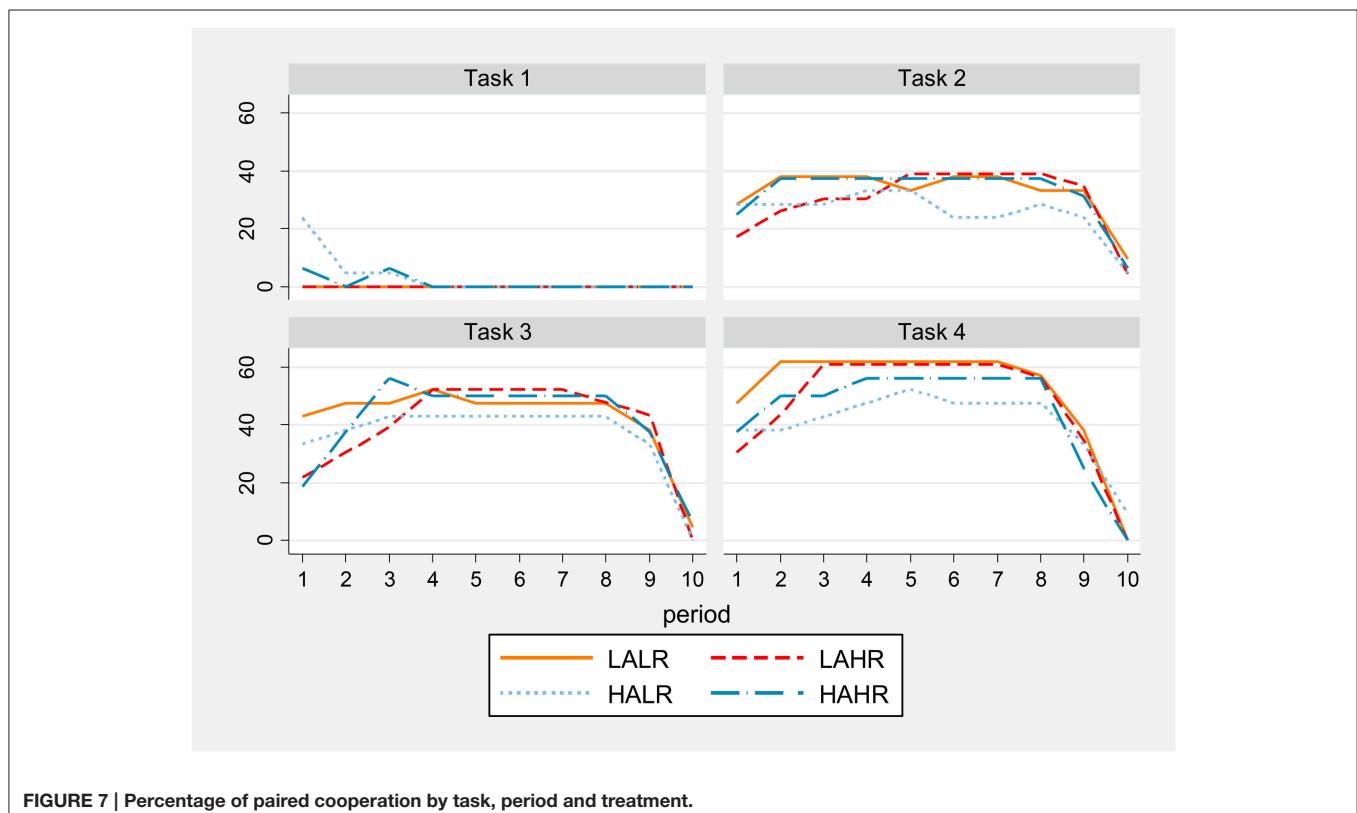
By paired cooperation we refer to the situation where both members of a pair simultaneously decide to cooperate in a given period, thus obtaining the cooperative payoff of the Prisoners' Dilemma.

As can be seen in **Figure 7**, successful paired cooperation is obviously much lower in the one-shot than in the repeated PD. Only altruists show some positive cooperation at the beginning of task 1. The difference in paired cooperation between low and high altruism pairs is significant for the first one-shot game ( $z = -2.78$  and  $p = 0.003$ ). All treatments increase paired cooperation at the beginning of the RPD games, particularly high reasoning ability subjects which show steep and significant increases in the first two periods. Specifically, we find significant differences comparing the level of paired cooperation in period 2 vs. period 1 for high reasoning ability pairs (at 5% in tasks 2 and 3, marginally in task 4; test details in Table SM2.14 in the Supplementary

**TABLE 6 | Random-effects panel logit regressions of individual cooperation on individual characteristics, period and beliefs.**

Individual cooperation	Task 1		Task 2		Task 3		Task 4	
	Coeff.	Sd.E.	Coeff.	Sd.E.	Coeff.	Sd.E.	Coeff.	Sd.E.
Reasoning ability	-0.09***	(0.03)	0.06**	(0.03)	0.04	(0.02)	-0.00	(0.02)
Altruism	0.20**	(0.09)	0.00	(0.09)	-0.01	(0.08)	-0.04	(0.09)
Social belief	0.02***	(0.01)						
Individual belief			5.31***	(0.31)	5.11***	(0.28)	5.17***	(0.29)
Female	0.31	(0.36)	0.20	(0.36)	0.02	(0.33)	-0.36	(0.33)
Period	-0.41***	(0.05)	-0.32***	(0.04)	-0.31***	(0.04)	-0.37***	(0.04)
Constant	-1.13	(0.80)	-3.03***	(0.76)	-2.00***	(0.70)	-0.36	(0.71)
N	1620		1620		1620		1620	
Wald Chi <sup>2</sup>	91.06***		297.67***		342.84***		319.96***	

\*\*\*Coefficient significant at 1%, \*\*Significant at 5%. Standard errors in parentheses.

**FIGURE 7 | Percentage of paired cooperation by task, period and treatment.**

Material). There are no other treatment differences in reaching and sustaining high cooperation. Tasks 2 and 3 present levels of paired cooperation close to 40%, and task 4 reaches 60%.

**Result 8:** *In the first one-shot game high altruism subjects exhibit higher levels of paired cooperation than low altruism ones.*

**Result 9:** *In the RPD game high reasoning ability subjects significantly increase paired cooperation in the first two periods, all treatments attaining and sustaining similarly high levels until one period before the last of each repetition, when cooperation crumbles.*

## DISCUSSION

We study cooperative behavior in (PD) games using a neat 2 by 2 factorial design, considering high vs. low altruism and high vs. low reasoning ability. As in all the previous experiments with these games, we find evidence of cooperation in both one-shot and finitely repeated (PD). In particular, we confirm the result by Andreoni and Miller (1993) and Cooper et al. (1996) that a certain amount of cooperative play appears to be due to the altruistic nature of subjects. In fact, by using an external measure of altruism (giving in a Dictator's Game), we show that altruism positively affects the likelihood of cooperation in

the one-shot PD games. Moreover, high altruism players seem to be more optimistic about their partners' behavior and they cooperate mainly thinking that their partner will also cooperate. Successful paired cooperation is very low in the one-shot games, with high altruism pairs being the only ones to reach positive levels.

As in the aforementioned studies and coherent with the "reputation building" hypothesis, we find that both individual and paired cooperation rates are much higher (40–60%) in the repeated PD games, and sustained for almost all periods, only to fall sharply in the last period of each task. Thanks to the elicitation of players' beliefs, we show that in our experiment cooperation is almost never unconditional: even altruistic subjects hardly cooperate if they think that their partner is going to defect. Altruism does not significantly increase neither individual nor paired cooperation in RPDs.

Interestingly, the effect of reasoning ability on individual cooperation changes sign depending on the type of PD game. Reconciling part of the previous literature and consistently with Burks et al. (2009)'s result for sequential PD, higher cognitive ability subjects appear to better adapt to the particular game played. In particular, they more accurately forecast their partner's behavior in the first repetitions of the one-shot games and at the beginning of the first RPD. Coherently, they tend to cooperate significantly less in the one-shot PD, as hinted in the lower continuation probability treatments of Proto et al. (2015). Also, they are more likely to cooperate in the first RPD, in line with what Jones (2008) found in his analysis using average intelligence scores. Differently to Al-Ubaydli et al. (2016), where paired cooperation is predicted by cognitive ability whereas individual cooperation is not, we do not find fundamental differences between individual and paired cooperation.

Reasoning ability is found to counteract the effect of altruism in the one-shot game. In fact, the joint effect of high reasoning ability and high altruism on the likelihood of cooperation appears to be no different from that of low reasoning ability and low altruism. However, while low reasoning ability individuals display similar behavior in both one-shot and RPD games, high reasoning ability subjects appear to better understand the nature of the one-shot (PD), changing then their decisions in the repeated version of the game.

Individual characteristics, however, fast reduce their weight in affecting subjects' decisions. While both reasoning ability and altruism explain individual cooperation in the one-shot PD and reasoning ability continues to be significant in the first RPD game, both characteristics become irrelevant as explicative variables when subjects gain experience in the RPD game. Instead, the variables affecting individual cooperation are period and subject beliefs. The latter could still be mediated by subject type, but in a more dynamic and adaptive way, as beliefs in the RPD are highly correlated with past partner cooperation. With experience in the RPD, reached and sustained

cooperation end up being similar among all groups. Thus, in a (PD) setting, altruism and reasoning ability significantly affect behavior in a situation in which no future consequence of choices is expected. This effect appears to be diluted when building a reputation can be used to reach higher payoffs. Indeed, transforming a social relationship into repeated interactions appears to be key to achieve mutual cooperation (Axelrod, 1984).

As future research, personality traits could also be added as determinants of cooperation, such as agreeableness or extraversion, as in Pothos et al. (2011), Proto et al. (2015), or Kagel and McGee (2014). They could be added as controls rather than as treatment variables, because the latter option would much complicate the treatment structure and impose high demands on the number of participants. An efficient alternative would be to program algorithmic players with a selection of frequently studied strategies and make them interact with human players, as in Hilbe et al. (2014). Also, having an increased age and culture variability could add insights on the determinants of cooperation.

## ETHICS STATEMENTS

This study was carried out in accordance with the recommendations of the ethical committee from the Universitat Jaume I. Participants gave informed consent in accordance with the Declaration of Helsinki. All participants in the subject database from the LEE at Universitat Jaume I in Castellón have voluntarily signed to participate in economic experiments and can freely decide whether they want to take part or not in each proposed experiment. No deception takes place in any experiment run at the LEE. No vulnerable populations were involved in the study.

## AUTHOR CONTRIBUTIONS

All authors collaborated in the development of the idea, the design of the project and the running of the sessions. IB programmed the software. AJ and IB developed the database and carried out most of the analyses. MP, IB, and GS wrote the article. All authors revised and accepted the written version.

## FUNDING

Financial support by Universitat Jaume I (project P1.1B2015-48) and the Spanish Ministry of Economics and Competitiveness (projects ECO2013-44409-P and ECO2015-68469-R) is gratefully acknowledged.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.00596/full#supplementary-material>

## REFERENCES

- Al-Ubaydli, O., Jones, G., and Weel, J. (2016). Average player traits as predictors of cooperation in a repeated prisoner's dilemma. *J. Behav. Exp. Econ.* 64, 50–60. doi: 10.1016/j.socec.2015.10.005
- Andreoni, J., and Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: experimental evidence. *Econ. J.* 103, 570–585. doi: 10.2307/2234532
- Axelrod, R. (1984). *The Evolution of Cooperation*. New York, NY: Basic Books.
- Bennett, G. K., Seashore, H. G., and Wesman, A. G. (1974). *DAT: Differential Aptitude Test*. New York, NY: The Psychological Corporation.
- Benz, M., and Meier, S. (2008). Do people behave in experiments as in the field? Evidence from donations. *Exp. Econ.* 11, 268–281. doi: 10.1007/s10683-007-9192-y
- Burks, S. V., Carpenter, J. P., Goette, L., and Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7745–7750. doi: 10.1073/pnas.0812360106
- Capraro, V., and Marcelletti, A. (2014). Do good actions inspire good actions in others? *Sci. Rep.* 4:7470. doi: 10.1038/srep07470
- Capraro, V., Smyth, C., Mylona, K., and Niblo, G. A. (2014). Benevolent characteristics promote cooperative behavior among humans. *PLoS ONE* 9:e102881. doi: 10.1371/journal.pone.0102881
- Carpenter, J., Connolly, C., and Myers, C. K. (2008). Altruistic behavior in a representative dictator experiment. *Exp. Econ.* 11, 282–298. doi: 10.1007/s10683-007-9193-x
- Colom, R., Escorial, S., Shih, P. C., and Privado, J. (2007). Fluid intelligence, memory span, and temperament difficulties predict academic performance of young adolescents. *Pers. Individ. Dif.* 42, 1503–1514. doi: 10.1016/j.paid.2006.10.023
- Cooper, R., DeJong, D. V., Forsythe, R., and Ross, T. W. (1996). Cooperation without reputation: experimental evidence from prisoner's dilemma games. *Games Econ. Behav.* 12, 187–218. doi: 10.1006/game.1996.0013
- Cordero, A., and Corral, S. (2006). *DAT – 5 Tests de Aptitudes Diferenciales*. Madrid: TEA Ediciones.
- Dreber, A., Fudenberg, D., and Rand, D. G. (2014). Who cooperates in repeated games: the role of altruism, inequity aversion, and demographics. *J. Econ. Behav. Organ.* 98, 41–55. doi: 10.1016/j.jebo.2013.12.007
- Duffy, S., and Smith, J. (2014). Cognitive load in the multi-player prisoner's dilemma game: are there brains in games? *J. Behav. Exp. Econ.* 51, 47–56. doi: 10.1016/j.socec.2014.01.006
- Engel, C. (2011). Dictator games: a meta study. *Exp. Econ.* 14, 583–610. doi: 10.1007/s10683-011-9283-7
- Fischbacher, U. (2007). Z-Tree: zurich toolbox for ready-made economic experiments. *Exp. Econ.* 10, 171–178. doi: 10.1007/s10683-006-9159-4
- Franzen, A., and Pointner, S. (2013). The external validity of giving in the dictator game. *Exp. Econ.* 16, 155–169. doi: 10.1007/s10683-012-9337-5
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *J. Econ. Sci. Assoc.* 1, 114–125. doi: 10.1007/s40881-015-0004-4
- Harrison, P. L. (1987). Research with adaptive behavior scales. *J. Spec. Educ.* 21, 37–68. doi: 10.1177/002246698702100108
- Hilbe, C., Röhl, T., and Milinski, M. (2014). Extortion subdues human players but is finally punished in the prisoner's dilemma. *Nat. Commun.* 5:3976. doi: 10.1038/ncomms4976
- Jones, G. (2008). Are smarter groups more cooperative? Evidence from prisoner's dilemma experiments, 1959–2003. *J. Econ. Behav. Organ.* 68, 489–497. doi: 10.1016/j.jebo.2008.06.010
- Jones, G. (2013). *Are Smarter Groups More Cooperative? Results for Corrected and Extended Datasets*. Working Paper George Mason University.
- Jones, M. T. (2014). Strategic complexity and cooperation: an experimental study. *J. Econ. Behav. Organ.* 106, 352–366. doi: 10.1016/j.jebo.2014.07.005
- Kagel, J., and McGee, P. (2014). Personality and cooperation in finitely repeated prisoner's dilemma games. *Econ. Lett.* 124, 274–277. doi: 10.1016/j.econlet.2014.05.034
- Keith, T. Z., Fherman, P. G., Harrison, P. L., and Pottebaum, S. M. (1987). The relation between adaptive behavior and intelligence: testing alternative explanations. *J. Sch. Psychol.* 25, 31–43. doi: 10.1016/0022-4405(87)90058-6
- Kreps, D. M., Milgrom, P., Roberts, J., and Wilson, R. (1982). Rational cooperation in the finitely repeated prisoner's dilemma. *J. Econ. Theory* 27, 245–252. doi: 10.1016/0022-0531(82)90029-1
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: standing on the shoulders of the giants of psychometric intelligence research. *Intelligence* 37, 1–10. doi: 10.1016/j.intell.2008.08.004
- Peysakhovich, A., Nowak, M. A., and Rand, D. G. (2014). Humans display a "cooperative phenotype" that is domain general and temporally stable. *Nat. Commun.* 5:4939. doi: 10.1038/ncomms5939
- Platt, L. O., Kamphaus, R. W., Cole, R. W., and Smith, C. L. (1991). Relationship between adaptive behavior and intelligence: additional evidence. *Psychol. Rep.* 68, 139–145. doi: 10.2466/pr.0.1991.68.1.139
- Pothos, E. M., Perry, G., Corr, P. J., Matthew, M. R., and Busemeyer, J. R. (2011). Understanding cooperation in the Prisoner's Dilemma game. *Pers. Individ. Dif.* 51, 210–215. doi: 10.1016/j.paid.2010.05.002
- Proto, E., Rustichini, A., and Sofianos, A. (2015). *Higher Intelligence Groups Have Higher Cooperation Rates in the Repeated Prisoner's Dilemma*. CAGE Working Paper 255.
- Van Lange, P. A. M., De Bruin, E. M. N., Otten, W., and Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: theory and preliminary evidence. *J. Pers. Soc. Psychol.* 73, 733–746. doi: 10.1037/0022-3514.73.4.733
- Yamagishi, T., Li, Y., Takagishi, H., Matsumoto, Y., and Kiyonari, T. (2014). In search of Homo economicus. *Psychol. Sci.* 25, 1699–1711. doi: 10.1177/0956797614538065

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Barreda-Tarazona, Jaramillo-Gutiérrez, Pavan and Sabater-Grande. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# At Least I Tried: The Relationship between Regulatory Focus and Regret Following Action vs. Inaction

Adi Itzkin, Dina Van Dijk and Ofer H. Azar\*

Guilford Glazer Faculty of Business and Management, Ben-Gurion University of the Negev, Beer Sheva, Israel

## OPEN ACCESS

**Edited by:**

Nikolaos Georgantzis,  
University of Reading, UK

**Reviewed by:**

Pablo Brañas-Garza,  
Middlesex University, UK  
Giuseppe Attanasi,  
University of Strasbourg, France

**\*Correspondence:**

Ofer H. Azar  
azar@som.bgu.ac.il

**Specialty section:**

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

**Received:** 18 June 2016

**Accepted:** 13 October 2016

**Published:** 27 October 2016

**Citation:**

Itzkin A, Van Dijk D and Azar OH  
(2016) At Least I Tried: The  
Relationship between Regulatory  
Focus and Regret Following Action vs.  
Inaction. *Front. Psychol.* 7:1684.  
doi: 10.3389/fpsyg.2016.01684

Regret is an unpleasant feeling that may arise following decisions that ended poorly, and may affect the decision-maker's well-being and future decision making. Some studies show that a decision to act leads to greater regret than a decision not to act when both resulted in failure, because the latter is usually the norm. In some cases, when the norm is to act, this pattern is reversed. We suggest that the decision maker's regulatory focus, affects regret after action or inaction. Specifically, promotion-focused individuals, who tend to be more proactive, view action as more normal than prevention-focused individuals, and therefore experience regulatory fit when an action decision is made. Hence, we hypothesized that promotion-focused individuals will feel less regret than prevention-focused individuals when a decision to act ended poorly. In addition, we hypothesized that a trigger for change implied in the situation, decreases the level of regret following action. We tested our hypotheses on a sample of 330 participants enrolled in an online survey. The participants received six decision scenarios, in which they were asked to evaluate the level of regret following action and inaction. Individual regulatory focus was measured by two different scales. Promotion-focused individuals attributed less regret than prevention-focused individuals to action decisions. Regret following inaction was not affected by regulatory focus. In addition, a trigger for change decreases regret following action. Orthodox people tend to attribute more regret than non-orthodox to a person who made an action decision. The results contribute to the literature by showing that not only the situation but also the decision maker's orientation affects the regret after action vs. inaction.

**Keywords:** regulatory focus, regulatory fit, promotion focus, prevention focus, regret, action, inaction

## INTRODUCTION

Every decision that we make in our life carries the risk that we might regret it. But what type of decisions will be regretted more: decisions of doing something or decisions of not doing anything? In the current paper we suggest that individual differences in regulatory focus would affect individuals' tendency to regret more what they did or what they did not do.

Regret is an unpleasant feeling that is aroused after retrospection that involves awareness to the negative aspects of a decision. The regret process involves running a mental re-creation of what actually happened vs. what could have happened, comparing these two options and deciding that the decision process and the outcome were suboptimal (Zeelenberg et al., 2002; Roes et al., 2009; Das and Joffe, 2012). The level of regret depends on an individual's perception of the mental gap

between what happened as opposed to what could have happened. The greater the gap, the stronger the regret (Das and Kerr, 2010). Regret could lead to regret aversion, which further encourages people to learn from past decisions in order to avoid similar experiences in the future (Zeelenberg and Beattie, 1997; Zeelenberg et al., 2002; Roese et al., 2009; Das and Kerr, 2010; Das and Joffe, 2012).

There are contradictory findings regarding what induces more regret: action or inaction. Action is considered as doing something that changes the current situation, such as deciding to go out for dinner or changing a strategy when trying to solve a problem. Inaction, on the other hand, is considered as doing nothing or keeping the status quo, such as staying home or keeping the same strategy already used. Early research found that because inaction is usually the norm, action, which violates the norm, leads to greater regret (Kahneman and Tversky, 1982; Kahneman and Miller, 1986). This finding has been later replicated in numerous studies (e.g., Landman, 1987; Gleicher et al., 1990; Baron and Ritov, 1994; Gilovich and Medvec, 1995; Miller and Taylor, 1995; Ritov and Baron, 1995; N'gbala and Branscombe, 1997; Van der Pligt et al., 1998; Ordóñez and Connolly, 2000). This effect has been termed the *action effect*, namely, an action that leads to a failure will cause greater regret than inaction that leads to similar failure (Zeelenberg et al., 2002).

Yet, other studies demonstrated that under certain conditions inaction can produce more regret than action (Gilovich and Medvec, 1995; Zeelenberg et al., 2002). For example, Gilovich and Medvec (1995) indicated that regret perception depends on the time horizon. Specifically, action is more regrettable in the short term, while inaction is more regrettable in the long term. Following this view, Zeelenberg et al. (2002) broadened Kahneman and Tversky (1982) theory and have added the *inaction effect*, which becomes relevant when action is perceived as desirable and needed, whereas inaction is perceived as less desired. Zeelenberg et al. suggested that what is considered normal can be influenced by a relevant past decision. In particular, when prior outcomes are positive or absent, inaction is considered more normal and people will attribute more regret to action than to inaction. However, when prior outcomes are negative, action becomes the more normal decision and more regret will be attributed to inaction.

Bar-Eli et al. (2007) showed that sometimes action can be the norm even without prior outcomes. They found that soccer goalkeepers in penalty kicks perceive action (jumping to one of the sides) as more normal than inaction (staying at the center of the goal) and consequently failed inaction produces more regret than failed action. Consequently, goalkeepers almost always choose action even though this actually reduces their chances to stop the ball. Azar (2013) examined the impact of previous outcomes not on regret but on decisions in a business strategy context. He found that whether a strategy was previously successful or not did not affect the likelihood that it will be continued or changed, in scenarios where the previous outcome was not informative about the future, but could trigger emotional reaction, such as regret.

Other factors that affect the relationship between failure associated with action vs. inaction and level of regret are

desirability and consistency (Seta et al., 2001). Seta et al. suggested that errors associated with action or inaction are less desirable and produce more regret when they are inconsistent with the decision maker's orientation (action or inaction) than when they are consistent. The effect of consistency and desirability was further demonstrated by McElroy and Dowd (2007). Building upon Seta et al. (2001) and McElroy and Dowd (2007) we suggest in the current paper that the level of regret following action or inaction is determined by the individual's regulatory focus, through the mechanism of regulatory fit. We rely on a well-established comprehensive motivational theory of regulatory focus (Higgins, 1997, 1998) to explain mixed evidences regarding regret following decisions of action vs. inaction.

Regulatory Focus Theory (Higgins, 1997, 1998), proposes that human motivation consists of two regulatory foci: promotion and prevention focus. People with promotion focus are motivated to achieve accomplishments, aspirations, and ideals, they are sensitive to gain - non-gain situations and to the presence or absence of positive outcomes. In contrast, individuals with prevention focus are motivated to attain security, responsibility, and duties. They are sensitive to loss - non-loss situations and to the presence or absence of negative outcomes (Higgins and Tykocinski, 1992; Friedman and Förster, 2001; Cesario et al., 2004; Förster et al., 2004). Regulatory focus can emerge as a chronic characteristic (personal disposition) as well as a situational (context-induced) variable (Higgins, 1997, 1998).

The two motivational foci are related to different types of strategies that are used to achieve individuals' goals. As was shown in numerous studies (Crowe and Higgins, 1997; Higgins, 1997; Shah et al., 1998; Liberman et al., 1999; Freitas and Higgins, 2002; Chernev, 2004; Avnet and Higgins, 2006), promotion focus individuals use approach and eagerness means to pursue their goals, tend to make changes and to take risks, try to achieve gains, and are prone to action; whereas prevention focus individuals use avoidance and vigilance means to pursue their goals, tend to maintain stability and to keep the status quo, try to avoid losses, and tend to caution and inaction. The inaction preference can be so profound that prevention-focused individuals might choose the status quo even if it is not the profitable one (Chernev, 2004).

According to regulatory focus theory, when the individuals' regulatory focus matches their goal pursuit means, they experience regulatory fit, which subsequently enhances their belief in what they are doing and the significance of their decisions (Higgins, 2000, 2005, 2006). Promotion focus fits goal pursuit means, such as eagerness and approach strategies (e.g., taking risks), whereas prevention focus fits goal pursuit means, such as vigilance and avoidance strategies (e.g., avoid risks). Under regulatory fit, people will judge a decision they made as more "right," value it more, and feel more engaged to their decision, than under non-fit condition (Camacho et al., 2003; Higgins, 2005). For example, Camacho et al. (2003) asked subjects to imagine themselves having a conflict with another person and to evaluate the other person's strategy of resolving the conflict. Promotion-focused subjects evaluated eager strategies (e.g., encouraging you to succeed) as more right than prevention-focused subjects, whereas prevention-focused subjects evaluated vigilant strategies (e.g., removing anything that might cause

trouble) as more right than promotion-focused subjects. Based on the regulatory fit principle, it can be argued that in situations of regulatory fit, because people feel more right about what they are doing, there will be less regret.

Although regret is a central emotion, the influence of regulatory fit on regret is still in the first stages of investigation. Kwak and Park (2008) showed that fear from anticipated regret increases the “sunk cost” effect (continue investing in a hopeless situation) under regulatory fit condition. However, our research focuses on the influence of regulatory fit on experienced regret in post-choice evaluation. Although conceptual models have suggested including regulatory fit as an integral part of the regret process, little has been conducted in this research area (Roese et al., 2007; Das and Kerr, 2010). We have found only one study that examined the effect of regulatory focus on regret (Church and Iyer, 2012), however, this study was not designed to test the different level of regret following action vs. inaction. The authors found (in contrast to their prediction) that people with higher promotion focus tended to regret their actions more than people higher on prevention focus. However, this was not compared to the level of regret following inaction. Thus, we cannot infer from these findings about our research questions.

In the present study we hypothesize that regulatory fit leads to less regret than regulatory non-fit. Specifically, a decision of action will fit individuals in promotion focus, and a decision of inaction will fit individuals in prevention focus. In other words, we argue that inaction will be considered as more normal behavior under prevention focus than under promotion focus, whereas action decisions will be viewed as more normal under promotion focus. Consequently, in line with norm theory (Kahneman and Miller, 1986), we hypothesize that the phenomenon of attaching more regret to action that fails than to inaction that fails, will be lower under promotion focus and higher under prevention focus. Thus, our hypotheses are:

**Hypothesis 1:** When two decisions resulted in failure, one is an action decision and another is an inaction decision, individuals in promotion focus will be more likely to attribute the lower regret to the action decision than individuals in prevention focus.

**Hypothesis 2:** Individuals in promotion focus will attribute less regret than individuals in prevention focus to an **action** decision that resulted in failure.

**Hypothesis 3:** Individuals in promotion focus will attribute more regret than individuals in prevention focus to an **inaction** decision that resulted in failure.

Hypothesis 1 relates to a binary question of regret (i.e., “Who feels more regret, the person who chose action or the person who chose inaction?”). Due to the binary nature of the question, this hypothesis is essentially identical to the symmetric hypothesis (“When two decisions resulted in failure, one is an action decision and another is an inaction decision, individuals in prevention focus will be more likely to attribute the lower regret to the inaction decision than individuals in promotion focus”) and therefore the results, which are presented according to Hypothesis 1, can also be interpreted as addressing this symmetric hypothesis.

Hypotheses 2 and 3 relate to two continuous variables of regret (i.e., level of regret following action and level of regret following inaction). Additionally, since regulatory focus is an individual tendency but could also be temporarily induced, all hypotheses will be tested by both individual and induced regulatory focus.

Along the lines of Zeelenberg et al. (2002), we hypothesize that a negative prior outcome makes action (a change in the status quo) more normal than absent such a negative prior outcome. In other words, a negative prior outcome creates a trigger for a change, and such a trigger increases the normality of choosing action and reduces the normality of choosing inaction, leading to reduced regret from failed action and increased regret from failed inaction. Moreover, we believe that not only a prior negative outcome but also a change in the environment may create a trigger for change, increasing the normality of action and reducing the normality of inaction, and therefore reducing the regret from action decisions compared to inaction decisions. For example, a trigger for change could stem from changing the targeted production level in one’s work environment, which implies that a change in one’s work strategy might be needed. To sum, we hypothesize:

**Hypothesis 4:** A trigger for change (caused by prior negative outcome or by a change in the environment) will reduce regret after failed **action** and increase regret after failed **inaction**.

Several demographic variables might affect the tendency to regret action or inaction. People’s level of religiosity may affect their tendency to favor action vs. inaction in different life situations. For example the Orthodox in Israel are known for their aversion of changes and by their clinging to the status quo, holding the motto of “to any proposal for change say ‘no’” (Lehmann and Siebzehner, 2009). Therefore, they might show more regret following action. Other demographic variables, such as gender, age, and income level could also affect the level of regret and therefore were taken into account in our analysis.

## METHODS

To test our hypotheses, we conducted an online experiment that included six scenarios in which regulatory focus (prevention vs. promotion) was induced, and regret level was measured after scenarios of failed action decisions vs. failed inaction decisions. In addition, chronic individual regulatory focus as well as demographic and personal variables were measured. The study received an approval from the Human Subjects Research Committee of the University.

## Sample

A total of 330 Israeli subjects were recruited voluntarily through a polling service company and were paid in exchange for their participation. One hundred and seventy three (52.4%) were female, age range was between 25 to 60 years old, and the mean age was  $M = 39$  ( $SD = 10.14$ ). Income level ranged between much below average ( $n = 41$ ; 12.4%), below average ( $n = 69$ ; 20.9%), average ( $n = 128$ ; 38.7%), above average ( $n = 66$ ; 20%), and much above average ( $n = 21$ ; 6.3%), with 5 missing values. In terms of religiosity level, 171 (51.8%) were secular,

76 (23%) traditionalists, 45 (13.6%) orthodox, and 38 (11.5%) ultra-orthodox.

## Procedure

The study consisted of three parts. In the first part participants filled a consent form and then filled the chronic regulatory focus measure. In the second part the subjects were divided randomly into three treatment groups: Induced promotion focus ( $n = 94$ ), induced prevention focus ( $n = 116$ ), and a control group ( $n = 120$ ). The manipulation included a word-completion task, in which subjects were asked to complete missing words in a text, using specific words that were provided in a list. The induced promotion focus manipulation used a list of promotion words (e.g., gain, aspirations, success), whereas the induced prevention focus manipulation used a list of prevention words (e.g., loss, obligations, failure). In the control condition no task was given. In order to check the manipulation the participants were asked to rate 8 behavior tendencies related to either a promotion focus (e.g., eager) or a prevention focus (e.g., vigilant), on a 10-point scale.

The third part of the experiment involved six scenarios. Each scenario presents an uncertain situation with two possible decisions: to retain the status quo (inaction) or to make a change (action). One decision maker in the scenario chooses action and the other chooses inaction, and both fail. The first two scenarios replicated those used in previous studies (Kahneman and Tversky, 1982; Zeelenberg et al., 2002), and the sixth scenario is somewhat similar to that in Gilovich and Medvec (1995). The additional new scenarios were developed to examine the robustness of the results to different contexts and situations, keeping the same structure of failed action vs. inaction. Three of the scenarios (2, 5, 6) contained a trigger for change, either a prior negative outcome or a change in the environment, while the other three scenarios did not contain any signal for the need of change (1, 3, 4). The scenarios were always presented in the same order, from 1 to 6.

After each scenario the subjects were asked to indicate who feels more regret (the one who acted or the one who did not act). This was the question used in Kahneman and Tversky (1982) and in Zeelenberg et al. (2002). Thus, in order to precisely replicate the original studies, we did not add any other questions that may affect the answers to the original question. In the four additional scenarios, however, we added two questions that asked the subjects to estimate the regret level of each decision maker on a 0–100 scale. We assumed that when participants are asked to estimate the regret level of a person in a hypothetical scenario they will use their own experience and personality to make their estimation. The six scenarios appear in the Appendix.

## Measures

*Individual Regulatory Focus* was measured with two measures: one is the scale of Lockwood et al. (2002), which is the most common scale for measuring regulatory focus, and the other is the Outcome-Based Measure (OBM; Schödl et al., 2013), a recently developed scale for regulatory focus.

1. *The Lockwood's Regulatory Focus Scale* consists of 18 items with a 9-point Likert scale. An example of a prevention item is "In general, I am focused on preventing negative events in my life," and an example of a promotion item is "In general, I am focused on achieving positive outcomes in my life." To calculate the total regulatory focus measure, an average score for each regulatory focus was calculated, then the gap between the promotion focus score to the prevention focus score was calculated (deducting the prevention score from the promotion score). The higher the score, the higher is the tendency toward promotion focus.
2. *The OBM Scale of Regulatory Focus* consists of 11-paired items describing cognitive, emotional, and strategic outcomes of regulatory focus based on Higgins (1997, 1998) theory. Each pair has two endings, one of promotion and one of prevention. A sample pair of items for a cognitive outcome is: "In general, I pay attention to: (a) negative information, (b) positive information." A sample pair of items for an emotional outcome is: "When I complete a task successfully: (a) I feel relief, (b) I feel joy." A sample pair of items for a strategic outcome is: "In general, I am: (a) enthusiastic, (b) cautious." Both endings of each item are rated on a 0–10 scale (0 = not at all true about me, 10 = very true about me). Reliabilities were  $\alpha = 0.83$  for promotion and  $\alpha = 0.82$  for prevention. A score for each regulatory focus was calculated by the sum of the answers to the relevant endings (prevention or promotion). The total regulatory focus measure is then obtained by deducting the prevention focus score from the promotion focus score. The higher the score, the higher is the tendency toward promotion focus.

Demographic variables age, gender, income, and religiosity level, were provided by the polling service company. Income was measured on an ordinal five-point scale (much below average, below average, average, above average, much above average). Religiosity level contained four categories that represent four main Israeli sectors: secular, traditionalist, orthodox, and ultra-orthodox. We recoded religiosity into a dichotomous variable with secular and traditionalist coded as "0" (non-orthodox), and orthodox and ultra-orthodox coded as "1."

## RESULTS

### Induced Regulatory Focus Manipulation

We first conducted a manipulation check for the regulatory focus manipulation. A set of eight Independent-Sample  $t$ -tests showed no differences between promotion and prevention conditions in terms of the behavior tendencies evoked by the word-completion task ( $t$ -tests ranged between  $-0.11 < t < 1.61$ ; and significance levels  $.12 < p < 0.91$ ). None of the eight behavior tendencies revealed significant difference between the two regulatory focus manipulations. As a result, the induced regulatory focus was not used in further analyses, but in order to control the potential effect of the manipulation, we added to the regressions two dummy variables for the promotion and prevention treatments (denoted by *Promotion\_Tr* and *Prevention\_Tr* in the regressions), where the control treatment

with no word-completion task is the benchmark. Thus, further analyses tested hypotheses 1–3 only with regards to the individual measures of regulatory focus and not regarding the induced regulatory focus.

## Who Feels More Regret?

Next, we considered the first question in each scenario, asking which of the two decision makers feel more regret, the person who acted or the person who did not act (the dichotomous measure of regret). For each subject we only have a binary response about who felt more regret, but aggregating over all the subjects we can get the proportion of subjects that attributed more regret to action or to inaction. **Table 1** presents these proportions and the test of whether the underlying probability is different from 0.5 (using the binomial distribution).

The results show that in scenario 5 the proportions are exactly 50–50% and in scenario 2 more people attribute greater regret to inaction (55.2 vs. 44.8%), but the difference is not statistically significant. In the other four scenarios a higher regret was attributed more often to the person who acted than to the person who did not act, and the difference is statistically significant at the 5%-level (using a 2-tailed binomial test). **Table 1** also shows that in three scenarios (2, 5, and 6) there were less than 57% who attributed more regret to action (than inaction). In the other three scenarios (1, 3, 4) more than 68% attributed more regret to action. The difference between these two groups of scenarios will be discussed later.

To test how regulatory focus is related to regret following action vs. inaction, we conducted two sets of six logistic regressions (one for each scenario) on the dichotomous measure of regret, namely, which person regret more, the one who acted (coded 1) or the one who did not act (coded 0). The first set of regressions included the following predictors: Demographics (age, gender, religiosity, and income), two dummy variables of the manipulation treatment of regulatory focus, and the individual regulatory focus measure of Lockwood. The second set of regressions was similar but used the OBM scale instead of Lockwood as the measure of individual regulatory focus. The results of the 12 logistic regressions, summarized in **Table 2**, show some support for Hypothesis 1. Specifically, **Table 2** demonstrates that in scenarios 2 and 5 (with both measures of regulatory focus), and scenario 6 (only with the

OBM scale), individual regulatory focus had a significant effect in the predicted direction, namely, the higher the promotion focus the lesser the probability of attributing more regret to the person who acted (compared with the one who did not act).

To be able to analyze the six scenarios together and derive more general conclusions, we created a database that aggregates the scenarios but records the unique subject ID in each observation. We then ran regressions on the combined data of scenarios 1–6 (clustered by subject ID), which are reported at the bottom of **Table 2**. These two regressions revealed that individual regulatory focus was statistically significant in the predicted direction ( $p = 0.017$  using Lockwood's scale,  $p = 0.001$  using the OBM scale). That is, the higher the promotion level, the less likely is the subject to attribute greater regret to the action decision (vs. inaction).

To sum, the effect of regulatory focus on the likelihood of attributing lower regret to the action decision (compared with the inaction decision) was obtained in three out of six scenarios when they are considered separately, and in the total measure of regret across all six scenarios. In addition, except for scenario 6, these effects were consistent across two different measures of regulatory focus. Thus, our results partially support hypothesis 1.

In addition to the effect of regulatory focus, subjects' religiosity level also had a significant effect on attributed regret. Specifically, in scenarios 2 and 5 and in the aggregated scenarios 1–6 (see **Table 2**) religiosity was positively and significantly related to the probability of attributing more regret to the person who acted. In scenarios 1 and 3 the effect of religiosity was positive and marginally significant ( $p$ -levels ranged between 0.06 and 0.07). All these mentioned effects of religiosity were consistently found across the two sets of regressions with both scales of regulatory focus. The positive effect of religiosity indicates that orthodox people are more likely than non-orthodox people to attribute more regret to action (compared to inaction).

## Regret Levels Following Action vs. Inaction

In order to test Hypothesis 2 we conducted two sets of linear regressions on the continuous measure of regret following action, which was measured in scenarios 3–6. Two sets of four linear regressions (for each of the four scenarios 3–6) were conducted on the level of regret attributed to the person who acted in the

**TABLE 1 | Who feels more regret—the person choosing action or inaction?**

Scenario	Regret following action is higher			Regret following inaction is higher		
	N	Frequency	Percent (%)	Frequency	Percent (%)	p-value (2-tailed)
1	330	246	74.5	84	25.5	0.000
2	330	148	44.8	182	55.2	0.069
3	307	211	68.7	96	31.3	0.000
4	297	212	71.3	85	28.7	0.000
5	314	157	50	157	50	1.000
6	299	169	56.5	130	43.5	0.028

The right column presents the 2-tailed p-value of the test (using the binomial distribution) of whether the probability of a subject attributing more regret to action (or inaction) is different from 0.5.

**TABLE 2 | Logistic regressions: does action produce more regret than inaction?**

Scenario	Variables	Lockwood's scale			OBM Scale		
		Coef.	Std. Err.	P	Coef.	Std. Err.	P
1	Age	-0.010	0.0131	0.405	-0.013	0.013	0.321
	Female	-0.003	0.263	0.990	-0.009	0.265	0.972
	Religiosity	0.614	0.339	0.070	0.643	0.341	0.060
	Income	0.050	0.119	0.673	0.068	0.119	0.568
	Promotion_Tr	-0.333	0.310	0.282	-0.340	0.310	0.273
	Prevention_Tr	0.087	0.316	0.783	0.100	0.317	0.752
	Lockwood's promotion	0.075	0.072	0.301			
	OBM promotion				-0.012	0.089	0.890
2	Age	-0.029	0.012	0.018	-0.025	0.012	0.043
	Female	-0.021	0.232	0.928	-0.071	0.231	0.759
	<b>Religiosity</b>	<b>0.707</b>	<b>0.278</b>	<b>0.011</b>	<b>0.672</b>	<b>0.270</b>	<b>0.013</b>
	Income	-0.086	0.110	0.430	-0.113	0.110	0.306
	Promotion_Tr	-0.355	0.294	0.228	-0.299	0.294	0.310
	Prevention_Tr	-0.054	0.267	0.839	0.002	0.269	0.992
	<b>Lockwood's promotion</b>	<b>-0.178</b>	<b>0.062</b>	<b>0.004</b>			
	<b>OBM promotion</b>				<b>-0.191</b>	<b>0.080</b>	<b>0.016</b>
3	Age	0.016	0.012	0.204	0.017	0.012	0.168
	Female	0.197	0.253	0.435	0.152	0.257	0.554
	Religiosity	0.564	0.308	0.067	0.581	0.312	0.062
	Income	-0.060	0.108	0.577	-0.062	0.105	0.553
	Promotion_Tr	0.550	0.319	0.085	0.579	0.321	0.072
	Prevention_Tr	0.331	0.293	0.259	0.383	0.295	0.194
	Lockwood's promotion	-0.067	0.074	0.362			
	OBM promotion				-0.134	0.086	0.122
4	Age	0.018	0.014	0.196	0.018	0.014	0.196
	Female	-0.212	0.266	0.427	-0.216	0.269	0.422
	Religiosity	0.340	0.306	0.267	0.346	0.309	0.262
	Income	-0.042	0.117	0.720	-0.037	0.114	0.745
	Promotion_Tr	0.128	0.332	0.698	0.129	0.330	0.697
	Prevention_Tr	-0.116	0.304	0.702	-0.111	0.305	0.715
	Lockwood's promotion	0.010	0.073	0.894			
	OBM promotion				-0.007	0.081	0.927
5	Age	-0.009	0.012	0.467	-0.004	0.012	0.714
	Female	0.017	0.241	0.944	-0.029	0.243	0.903
	<b>Religiosity</b>	<b>0.787</b>	<b>0.290</b>	<b>0.007</b>	<b>0.747</b>	<b>0.286</b>	<b>0.009</b>
	Income	-0.101	0.111	0.365	-0.132	0.106	0.213
	Promotion_Tr	-0.552	0.298	0.065	-0.499	0.296	0.092
	Prevention_Tr	-0.141	0.279	0.612	-0.095	0.277	0.731
	<b>Lockwood's promotion</b>	<b>-0.184</b>	<b>0.063</b>	<b>0.004</b>			
	<b>OBM promotion</b>				<b>-0.174</b>	<b>0.080</b>	<b>0.030</b>
6	Age	-0.004	0.012	0.728	-0.003	0.012	0.796
	Female	0.373	0.240	0.120	0.294	0.243	0.226
	Religiosity	0.132	0.284	0.641	0.171	0.295	0.562
	Income	0.188	0.099	0.058	0.203	0.101	0.044
	Promotion_Tr	-0.247	0.299	0.409	-0.198	0.302	0.511
	Prevention_Tr	-0.270	0.282	0.339	-0.166	0.289	0.566
	Lockwood's promotion	-0.088	0.065	0.178			
	<b>OBM promotion</b>				<b>-0.252</b>	<b>0.084</b>	<b>0.003</b>

(Continued)

**TABLE 2 | Continued**

Scenario	Variables	Lockwood's scale			OBM Scale		
		Coef.	Std. Err.	P	Coef.	Std. Err.	P
1–6 total	Age	−0.004	0.005	0.480	−0.002	0.005	0.668
	Female	0.059	0.110	0.592	0.018	0.111	0.868
	<b>Religiosity</b>	<b>0.494</b>	<b>0.138</b>	<b>0.000</b>	<b>0.497</b>	<b>0.140</b>	<b>0.000</b>
	Income	−0.013	0.0474	0.778	−0.019	0.048	0.682
	Promotion_Tr	−0.147	0.138	0.288	−0.118	0.138	0.391
	Prevention_Tr	−0.039	0.128	0.763	0.006	0.127	0.964
	<b>Lockwood's promotion</b>	<b>−0.074</b>	<b>0.031</b>	<b>0.017</b>			
	<b>OBM promotion</b>				<b>−0.124</b>	<b>0.037</b>	<b>0.001</b>

The dependent variable is *ActMoreRegret*, a dummy variable that equals one if the subject thinks that the person who acted feels more regret than the one who did not act. The table reports the robust standard errors. The last regressions, on the combined data of scenarios 1–6, are clustered by subject ID. Significant effects are bold.

scenario. The predictors in the logistic regressions were used also here. In addition, the level of regret attributed to the person who did not act (by the same subject in the same scenario) was also included as an independent variable, in order to control for regret following inaction when predicting regret following action.

**Table 3** presents the regression results that show what affects the regret attributed to the action decision. Because the question about the level of regret after the action decision was introduced only in scenarios 3–6, the results do not include scenarios 1–2. In scenarios 4, 5 (with both measures of regulatory focus), and 6 (only with the OBM scale), individual regulatory focus was significant at the 5% level in the predicted direction, namely, the higher the promotion focus, the lower the regret level attributed to the action decision. This effect was obtained beyond the positive effect of the level of regret attributed to the inaction decision. In other words, despite the fact that the level of regret attributed to inaction was positively and significantly related to the level of regret attributed to the action decision, the unique effect of regulatory focus on regret attributed to action was significant, supporting Hypothesis 2. To get an overview of the general findings across all scenarios, we also ran two regressions on the combined data of scenarios 3–6 (clustered by subject ID), reported at the bottom of **Table 3**. In line with Hypothesis 2, individual regulatory focus measured by both Lockwood's scale and the OBM scale was significant, such that the higher the promotion focus, the lower the regret attributed to the action decision.

To sum, the effect of regulatory focus on the attribution of regret to an action decision was obtained in three out of four scenarios and in the total measure of regret across all four scenarios. In addition, these effects were consistent across two different measures of regulatory focus. Thus, our results support hypothesis 2.

In order to test Hypothesis 3, we ran two additional sets of regressions on the level of regret following *inaction*. The same predictors were used as in the previous regressions, but this time we controlled for the regret level following *action*, since we predicted the level of regret following inaction. The results of these linear regressions are shown in **Table 4**.

and surprisingly do not support Hypothesis 3. Specifically, no effect of regulatory focus on the level of regret following inaction was revealed (except for one effect of the Lockwood's scale in scenario 6). As can be seen in **Table 4** the regret level following action positively predicts the level of regret following inaction, but regulatory focus has no unique effect on regret following inaction. Thus, the results did not confirm Hypothesis 3.

## The Effect of a Trigger for Change

We now turn to examine Hypothesis 4, according to which a trigger for change lowers the level of regret attributed to action. Scenarios 2, 5, and 6, included a trigger for change, whereas scenarios 1, 3, and 4 did not include any trigger for change. Scenario 2, which replicates a study of Zeelenberg et al. (2002), includes a negative prior outcome (losing the prior game), after which the coach has to decide whether to change the team. The prior loss creates a trigger to do something different, i.e., a trigger for change. Similarly, in Scenario 6, which deals with a decision of students to change or not their university, it is mentioned that the students are unhappy with their university, again creating a trigger for change. In scenario 5, which deals with two employees who have weekly manufacturing targets, it is mentioned that this week the target was higher than usual. This is not a prior negative outcome but it is an important change in the environment, which can be a trigger for change in the decision (which machine parameters to adopt). In contrast to those three scenarios, scenarios 1, 3, and 4, describe a decision of two people to change or not to change, without any additional information that could be a trigger for change. For example, scenario 1 (a replication of a scenario from Kahneman and Tversky, 1982) describes two people who decide to change/not change a stock, but no reason or additional information regarding the necessity of a change is given. Similarly, scenarios 3 and 4 present two decisions to change/not change a project (scenario 3), or a supplier (scenario 4), but no additional information is given for a prior negative outcome of the current project or supplier, or a significant change in the environment. Therefore, no apparent trigger for change is created in scenarios 1, 3, and 4.

**TABLE 3 | Linear regressions explaining regret following action.**

Scenario	Variables	Lockwood's scale			OBM scale		
		Coef.	Std. Err.	P	Coef.	Std. Err.	P
3	Regret for inaction	<b>0.191</b>	<b>0.067</b>	<b>0.005</b>	<b>0.194</b>	<b>0.068</b>	<b>0.005</b>
	Age	0.016	0.153	0.918	0.046	0.154	0.763
	Female	0.660	2.762	0.811	0.149	2.787	0.957
	Religiosity	1.592	3.120	0.610	1.448	3.070	0.638
	Income	-1.049	1.285	0.415	-1.215	1.339	0.365
	Promotion_Tr	-0.204	3.129	0.948	0.278	3.119	0.929
	Prevention_Tr	3.304	3.190	0.301	3.930	3.200	0.220
	Lockwood's promotion	-1.375	0.764	0.073			
	OBM promotion				-1.758	0.939	0.062
4	Regret for inaction	<b>0.181</b>	<b>0.065</b>	<b>0.006</b>	<b>0.184</b>	<b>0.065</b>	<b>0.005</b>
	Age	0.039	0.159	0.803	0.077	0.161	0.631
	Female	1.568	2.831	0.580	0.697	2.867	0.808
	Religiosity	0.437	3.291	0.894	0.355	3.274	0.914
	Income	0.946	1.309	0.470	0.785	1.401	0.576
	Promotion_Tr	-5.223	3.319	0.117	-4.501	3.281	0.171
	Prevention_Tr	-2.666	3.119	0.393	-1.620	3.111	0.603
	<b>Lockwood's promotion</b>	<b>-1.868</b>	<b>0.739</b>	<b>0.012</b>			
	<b>OBM promotion</b>				<b>-2.894</b>	<b>0.875</b>	<b>0.001</b>
5	Regret for inaction	<b>0.141</b>	<b>0.069</b>	<b>0.041</b>	<b>0.148</b>	<b>0.068</b>	<b>0.031</b>
	Age	0.170	0.137	0.215	0.220	0.138	0.112
	Female	1.027	2.819	0.716	0.119	2.863	0.967
	Religiosity	3.900	3.518	0.268	3.730	3.467	0.283
	Income	-0.953	1.330	0.474	-1.209	1.370	0.378
	Promotion_Tr	-4.840	3.339	0.148	-4.026	3.287	0.222
	Prevention_Tr	-4.649	3.324	0.163	-3.589	3.356	0.286
	<b>Lockwood's promotion</b>	<b>-2.260</b>	<b>0.796</b>	<b>0.005</b>			
	<b>OBM promotion</b>				<b>-3.062</b>	<b>0.988</b>	<b>0.002</b>
6	Regret for inaction	<b>0.227</b>	<b>0.072</b>	<b>0.002</b>	<b>0.235</b>	<b>0.070</b>	<b>0.001</b>
	Age	0.035	0.156	0.823	0.055	0.159	0.730
	Female	2.393	3.034	0.431	1.69	3.073	0.582
	Religiosity	2.227	3.563	0.532	2.370	3.491	0.498
	Income	0.359	1.333	0.788	0.318	1.381	0.818
	Promotion_Tr	-2.040	3.788	0.590	-1.480	3.761	0.694
	Prevention_Tr	-3.369	3.517	0.339	-2.471	3.494	0.480
	Lockwood's promotion	-1.133	0.836	0.176			
	<b>OBM promotion</b>				<b>-2.340</b>	<b>1.053</b>	<b>0.027</b>
3–6 total	Regret for inaction	-.053	0.054	0.320	-.045	0.054	0.407
	Age	0.050	0.130	0.703	0.090	0.131	0.492
	Female	2.085	2.344	0.374	1.335	2.362	0.572
	Religiosity	1.418	2.864	0.620	1.276	2.828	0.652
	Income	0.183	1.345	0.892	-0.028	1.414	0.984
	Promotion_Tr	-3.499	2.588	0.176	-2.827	2.530	0.264
	Prevention_Tr	-1.944	2.735	0.477	-1.061	2.759	0.701
	<b>Lockwood's promotion</b>	<b>-1.845</b>	<b>0.652</b>	<b>0.005</b>			
	<b>OBM promotion</b>				<b>-2.501</b>	<b>0.809</b>	<b>0.002</b>

The dependent variable is the regret (on a 0–100 scale) following a failed action decision. The table reports the robust standard errors. The last regressions, on the combined data of scenarios 3–6, are random-effects GLS regressions clustered by subject ID.  
Significant effects are bold.

**TABLE 4 | Linear regressions explaining regret following inaction.**

Scenario	Variables	Lockwood's scale			OBM scale		
		Coef.	Std. Err.	P	Coef.	Std. Err.	P
3	Regret for action	<b>0.209</b>	<b>0.070</b>	<b>0.003</b>	<b>0.21244</b>	<b>0.0711</b>	<b>0.003</b>
	Age	-0.147	0.149	0.325	-0.1552	0.150	0.302
	Female	2.368	2.821	0.402	2.692	2.830	0.342
	Religiosity	-3.087	3.154	0.328	-3.155	3.138	0.315
	Income	1.665	1.247	0.183	1.670	1.231	0.176
	Promotion_Tr	-2.799	3.410	0.412	-3.037	3.409	0.374
	Prevention_Tr	-5.534	3.248	0.089	-5.956	3.212	0.065
	Lockwood's promotion	0.474	0.779	0.543			
	OBM promotion				1.073	1.035	0.301
4	Regret for action	<b>0.210</b>	<b>0.075</b>	<b>0.005</b>	<b>0.216</b>	<b>0.076</b>	<b>0.005</b>
	Age	0.132	0.161	0.413	0.1372	0.161	0.394
	Female	<b>6.566</b>	<b>2.978</b>	<b>0.028</b>	<b>6.698</b>	<b>2.982</b>	<b>0.025</b>
	Religiosity	1.489	3.389	0.661	1.333	3.381	0.694
	Income	0.259	1.299	0.842	0.177	1.297	0.891
	Promotion_Tr	2.055	3.569	0.565	2.026	3.551	0.569
	Prevention_Tr	2.105	3.429	0.540	1.916	3.443	0.578
	Lockwood's promotion	-0.102	0.833	0.902			
	OBM Promotion				0.438	1.094	0.689
5	Regret for action	<b>0.151</b>	<b>0.072</b>	<b>0.037</b>	<b>0.159</b>	<b>0.071</b>	<b>0.027</b>
	Age	-0.138	0.158	0.381	-0.134	0.160	0.405
	Female	3.417	2.918	0.242	3.616	2.943	0.220
	Religiosity	-5.634	3.507	0.109	-5.854	3.467	0.092
	Income	1.845	1.308	0.159	1.755	1.309	0.181
	Promotion_Tr	-0.523	3.653	0.886	-0.575	3.636	0.875
	Prevention_Tr	3.353	3.356	0.318	3.103	3.370	0.358
	Lockwood's promotion	-0.098	0.824	0.905			
	OBM Promotion				0.627	1.005	0.533
6	Regret for action	<b>0.216</b>	<b>0.068</b>	<b>0.002</b>	<b>0.230</b>	<b>0.068</b>	<b>0.001</b>
	Age	-0.132	0.154	0.391	-0.081	0.156	0.605
	Female	-1.683	2.983	0.573	-1.561	3.037	0.608
	Religiosity	-4.416	3.586	0.219	-5.222	3.617	0.150
	Income	1.985	1.298	0.127	1.497	1.303	0.251
	Promotion_Tr	-2.768	3.706	0.456	-2.527	3.779	0.504
	Prevention_Tr	0.493	3.479	0.887	0.265	3.530	0.940
	<b>Lockwood's promotion</b>	<b>-1.724</b>	<b>0.781</b>	<b>0.028</b>			
	OBM Promotion				0.171	0.936	0.855
3–6 total	Regret for action	-0.022	0.059	0.702	-0.014	0.059	0.805
	Age	-0.061	0.121	0.611	-0.040	0.121	0.744
	Female	3.095	2.247	0.169	3.132	2.244	0.163
	Religiosity	-2.594	2.723	0.341	-2.937	2.706	0.278
	Income	1.475	1.254	0.240	1.272	1.261	0.313
	Promotion_Tr	-1.733	2.606	0.506	-1.612	2.598	0.535
	Prevention_Tr	-0.288	2.562	0.910	-0.375	2.587	0.885
	Lockwood's promotion	-0.751	0.601	0.212			
	OBM Promotion				0.035	0.792	0.964

The dependent variable is the regret (on a 0–100 scale) following a failed inaction decision. The table reports the robust standard errors. The last regressions, on the combined data of scenarios 3–6, are random-effects GLS regressions clustered by subject ID.

Significant effects are bold.

**Table 1** already demonstrates that something is different between the trigger for change (TFC) scenarios 2, 5, and 6, and the no-TFC scenarios 1, 3, and 4. In particular, the proportion of subjects who attribute greater regret to action than to inaction ranged between 68.7 and 74.5% in the no-TFC scenarios, but only 44.8–56.5% in the TFC scenarios. Considering the continuous variables of regret levels following action and inaction, we see again a remarkable difference between the TFC scenarios (now only scenarios 5 and 6 because no continuous regret levels were elicited for scenarios 1 and 2) and the no-TFC scenarios 3 and 4. More specifically, in the no-TFC scenarios (3 and 4), the regret from action was higher than regret from inaction and the difference was statistically significant (68.31 vs. 56.29,  $p = 0.0000$  in Scenario 3; 69.33 vs. 55.94,  $p = 0.0000$  in Scenario 4). However, in the TFC scenarios (5 and 6) the regret levels from action and inaction were very close and not statistically significant (61.38 vs. 58.35,  $p = 0.1027$  in Scenario 5; 59.27 vs. 57.79,  $p = 0.4274$  in Scenario 6). Overall, the level of regret after action was significantly higher when no trigger for change exists compared to the TFC scenarios (68.82 vs. 60.33,  $p = 0.0000$ ). However, the regret from inaction was similar regardless of a trigger for change (56.11 vs. 58.07,  $p = 0.1747$ ).

Hypothesis 4 was tested on the combined data of scenarios 1–6 (clustered by subject ID). We ran logistic regressions on the dichotomous measure of regret with the same independent variables as in the previous logistic regressions, but also adding a dummy variable for the trigger for change (coded “0” for no-TFC scenarios, and “1” for TFC scenarios). In addition, in order to test whether the effect of regulatory focus differs between TFC and no-TFC scenarios, we added the interaction between the trigger for change and the individual regulatory focus (TFC X promotion focus). **Table 5** summarizes the results of the two regressions (one with Lockwood’s promotion focus and one with the OBM promotion focus).

As can be seen in **Table 5**, according to our prediction, the trigger for change had a significant negative effect on the

probability of attributing more regret to action, meaning that when there is a trigger for change, less regret is attributed to action (compared to no trigger for change). This finding was consistent across the two measures of individual regulatory focus and further confirmed Hypothesis 4. In addition, while the main effect of regulatory focus was non-significant, the interaction between regulatory focus and the trigger for change was significant and negative. This significant interaction together with the lack of significant effect of the promotion focus variable itself, suggests that when asking subjects the binary question of who feels more regret, there is no significant effect of promotion focus in scenarios without a trigger for change, but there is a significant effect of promotion focus once a trigger for change is introduced. In particular, a trigger for change makes it less likely that the greater regret will be attributed to the person who chose action. These findings were consistent across the two measures of individual regulatory focus.

In addition, subjects’ religiosity level also had a significant effect on attributed regret, indicating that orthodox people are more likely than non-orthodox people to attribute more regret to action ( $p = 0.000$  for both measures of regulatory focus). This effect was consistent with the effects of religiosity that were found in the previous logistic regressions (see **Table 2**).

We also tested Hypothesis 4 on the two continuous measures of regret: regret following action (see **Table 6**) and regret following inaction (see **Table 7**). Two sets of linear regression models were conducted on the combined data of scenarios 3–6 (clustered by subject ID). The independent variables were the same as in the previous regression, except that we controlled for regret following inaction when predicting regret following action; and we controlled regret following action when predicting regret following inaction.

As can be seen in **Table 6**, the trigger for change had a significant negative effect on regret following action, meaning that when there is a trigger for change, less regret is attributed to action. This finding further confirms Hypothesis 4. In addition,

**TABLE 5 | Logistic regressions explaining regret following action vs. inaction: adding the trigger for change.**

Variables	Lockwood’s scale			OBM scale		
	Coef.	Std. Err.	P	Coef.	Std. Err.	P
Age	−0.004	0.006	0.489	−0.002	0.006	0.667
Female	0.062	0.116	0.592	0.020	0.117	0.864
<b>Religiosity</b>	<b>0.519</b>	<b>0.145</b>	<b>0.000</b>	<b>0.521</b>	<b>0.147</b>	<b>0.000</b>
Income	−0.011	0.051	0.832	−0.017	0.051	0.734
Promotion_Tr	−0.156	0.146	0.287	−0.125	0.145	0.390
Prevention_Tr	−0.043	0.135	0.751	0.005	0.134	0.971
<b>TFC</b>	<b>−0.710</b>	<b>0.135</b>	<b>0.000</b>	<b>−0.867</b>	<b>0.105</b>	<b>0.000</b>
Lockwood’s promotion	−0.008	0.042	0.838			
<b>TFC X Lockwood’s promotion</b>	<b>−0.125</b>	<b>0.051</b>	<b>0.014</b>			
OBM Promotion				−0.052	0.048	0.277
<b>TFC X OBM promotion</b>				<b>−0.148</b>	<b>0.065</b>	<b>0.022</b>

The dependent variable is *ActMoreRegret*, a dummy variable that equals one if the subject thinks that the person who acted feels more regret than the one who did not act. TFC, Trigger for Change. The table reports the robust standard errors. The regressions are clustered by subject ID. Significant effects are bold.

**TABLE 6 | Linear regressions explaining regret following action: adding the trigger for change.**

Variables	Lockwood's scale			OBM scale		
	Coef.	Std. Err.	P	Coef.	Std. Err.	P
Regret for Inaction	-0.055	0.053	0.302	-0.045	0.053	0.392
Age	0.050	0.131	0.704	0.090	0.131	0.493
Female	2.088	2.346	0.373	1.338	2.363	0.571
Religiosity	1.415	2.868	0.622	1.274	2.831	0.653
Income	0.184	1.347	0.891	-0.027	1.415	0.985
Promotion_Tr	-3.501	2.591	0.177	-2.828	2.534	0.264
Prevention_Tr	-1.944	2.739	0.478	-1.061	2.763	0.701
<b>TFC</b>	<b>-7.794</b>	<b>1.434</b>	<b>0.000</b>	<b>-8.263</b>	<b>1.254</b>	<b>0.000</b>
<b>Lockwood's promotion</b>	<b>-1.645</b>	<b>0.693</b>	<b>0.018</b>			
TFC X Lockwood's promotion	-0.401	0.636	0.529			
<b>OBM promotion</b>				<b>-2.210</b>	<b>0.820</b>	<b>0.007</b>
TFC X OBM promotion				-0.581	0.751	0.440

The dependent variable is the regret (on a 0–100 scale) following a failed action decision. TFC, Trigger for Change. The table reports the robust standard errors. The regressions are random-effects GLS regressions clustered by subject ID. Significant effects are bold.

**TABLE 7 | Linear regressions explaining regret following inaction: adding the trigger for change.**

Variables	Lockwood's scale			OBM scale		
	Coef.	Std. Err.	P	Coef.	Std. Err.	P
Regret for action	-0.013	0.060	0.822	-0.004	0.061	0.947
Age	-0.062	0.120	0.607	-0.041	0.121	0.736
Female	3.077	2.237	0.169	3.119	2.231	0.162
Religiosity	-2.608	2.710	0.336	-2.952	2.690	0.272
Income	1.474	1.245	0.237	1.273	1.249	0.308
Promotion_Tr	-1.702	2.595	0.512	-1.582	2.585	0.540
Prevention_Tr	-0.271	2.550	0.915	-0.364	2.572	0.887
<b>TFC</b>	<b>3.557</b>	<b>1.736</b>	<b>0.040</b>	2.085	1.440	0.148
Lockwood's promotion	-0.264	0.725	0.715			
TFC X Lockwood's promotion	-0.941	0.758	0.214			
OBM promotion				0.128	0.969	0.895
TFC X OBM promotion				-0.131	0.879	0.881

The dependent variable is the regret (on a 0–100 scale) following a failed inaction decision. TFC, Trigger for Change. The table reports the robust standard errors. The regressions are random-effects GLS regressions clustered by subject ID. Significant effects are bold.

the effect of regulatory focus was significant such that the higher the promotion focus, the lower the regret following action (supporting Hypothesis 2 as in our earlier findings). The interaction between trigger for change and regulatory focus was non-significant. This pattern of results was consistent in both measures of individual regulatory focus.

Finally, as can be seen in **Table 7** and in line with Hypothesis 4, the trigger for change had a positive effect on regret following inaction, meaning that when there is a signal that a change might be needed, there is more regret following inaction. However, this effect was weaker than the effect of TFC on regret from action (the coefficients of TFC on regret from action are -7.8 and -8.3 in the two regressions, compared to coefficients of

+3.6 and +2.1 on regret from inaction). In addition, this effect was statistically significant for the regret from inaction only when the Lockwood's scale was used. When using the OBM scale this effect was not statistically significant, although it had a positive coefficient as predicted. The individual regulatory focus had no effect on regret following inaction, similar to the results in **Table 4**, and once again not consistent with Hypothesis 3. The interaction between individual regulatory focus and the trigger for change also had no effect on regret from inaction.

In sum, the data strongly support our prediction that the existence of a trigger for change decreases the level of regret following action, but only partially support our prediction that

it increases the level of regret following inaction. In addition, the effect of regulatory focus was similar to our earlier findings, namely, promotion focus decreases regret following action (supporting Hypothesis 2), but does not increase regret following inaction (not supporting Hypothesis 3).

## Summary of Results

Our results provide partial support for hypotheses 1, 2, and 4, but did not support hypothesis 3. When testing whether more regret is attributed to action decision or to inaction decision, we found that regulatory focus was significantly related to regret in three out of six scenarios (2, 5, and 6) and when the effect is calculated across all six scenarios. The direction of the effect indicates that the higher the promotion focus, the lower the probability of attributing more regret to action. Similarly, when testing the regret following action (where it was measured on a 0–100 scale, i.e., in scenarios 3–6), the same effect of regulatory focus was found. Specifically, regulatory focus was related to regret in three out of four scenarios (4, 5, and 6) and when the effect is calculated across all four scenarios, such that the higher the promotion focus, the lower the attributed regret following action. However, when testing regret following inaction, there was no effect of regulatory focus in any of the scenarios (except for scenario 6 in Lockwood's scale), and also not when calculating the total effect across all four scenarios. In addition, according to our prediction, we found that when the situation contains a trigger for change, less regret is attributed to action and more regret is attributed to inaction (although the effect of TFC on inaction was not always statistically significant and it was weaker than its effect on action). Finally, relatively high consistency was found in the results pattern between the two scales of regulatory focus. This consistency further strengthens the robustness of our findings.

## DISCUSSION AND CONCLUSIONS

The present study examines the effect of regulatory focus on regret feelings following action vs. inaction decisions. The results indicate that individual differences in regulatory focus are related to the level of regret that emerges after making a decision that results in failure, and in particular, after making an action decision. The mechanism that explains the effect of regulatory focus on regret stems from the principle of regulatory fit. According to this principle, when the individual regulatory focus of decision makers fits their goal pursuit means or strategies, they feel more right about what they are doing (Higgins, 2000, 2005, 2006). Because an action decision fits more promotion focus, whereas an inaction decision fits more prevention focus (e.g., Chernev, 2004), we predicted that action will be less regrettable for promotion-focused individuals, whereas inaction will be less regrettable for prevention-focused individuals. Our results indeed show that promotion-focused individuals attribute less regret to action decisions than prevention-focused individuals. However, no difference was found between individuals with promotion and prevention foci with regard to regret from inaction decisions.

## Regulatory Focus and Regret

Our findings contribute to the regulatory focus research arena by expanding the role of individual regulatory focus to the domain of regret. So far numerous studies have found that regulatory focus affects people's decisions and choices (e.g., Aaker and Lee, 2001; Chernev, 2004; Lee and Aaker, 2004; Avnet and Higgins, 2006), strategies (e.g., Crowe and Higgins, 1997; Lockwood et al., 2002) and emotions (e.g., Higgins et al., 1997). However, as far as we know no study has investigated the effect of regulatory focus on post-choice regret. While previous research showed that people valued more decisions that were made under conditions of regulatory fit, than under non-fit (e.g., Higgins et al., 2003), the current research extends previous research by showing that under regulatory fit condition, people are also less likely to regret their decisions. Specifically, since action decision fits promotion focus orientation, an action decision is regretted less by promotion-focused individuals than by prevention-focused individuals. Understanding the impact of regulatory focus on regret from action vs. inaction could have implications for individuals' well-being and emotional regulation. For example, we can predict that prevention-focused individuals will be more sensitive to the negative effects of regret emerged by action decisions that failed; such negative effects could be reduced well-being, guilt or other negative feelings. On the other hand, our results do not suggest that the opposite effect is true for promotion-focused individuals, namely, inaction decisions that failed do not seem to harm promotion-focused individuals (compared to prevention ones). Thus, we suggest that prevention-focused individuals will be more sensitive to the harmful effect of regret following action decision, while promotion-focused individuals will be more resilient to such harmful effect. This notion is consistent with previous research suggesting that prevention-focused individuals might be more vulnerable to reduced well-being, whereas promotion focus is related to more resiliency (Van Dijk et al., 2013). Future studies are encouraged to further investigate the effect of regulatory focus on regret and regret consequences, such as reduced well-being, negative feelings and regret aversion.

## Action and Inaction Asymmetry

Our findings show asymmetry in the effect of regulatory focus on regret following action vs. inaction. This asymmetry has not been revealed by previous studies. When using a binary measure of regret (i.e., who regrets more: a person who acted or a person who did not act), we found that promotion focus decreased the probability of attributing more regret to action than to inaction. However, using the binary question we still do not know whether this effect results from promotion-focused individuals attributing less regret to action, more regret to inaction, or both. The use of additional two continuous measures of regret (i.e., regret following action and regret following inaction) revealed an asymmetric pattern between action and inaction. Specifically, promotion-focused individuals attribute less regret to action than prevention-focused individuals, but the two groups attribute similar regret levels to inaction decisions. This asymmetry between action and inaction implies that a decision not to act is the default or the norm, as suggested by

the norm theory (Kahneman and Miller, 1986). This means that inaction or leaving things as they are, without making any change, is the first and basic option in a situation of choice. Taking action or changing the status quo, on the other hand, is a less trivial choice and it requires more intent and deliberate plan. Therefore, inaction decisions are similarly perceived by different individuals, and even among individuals who tend to use action strategies or a promotion focus—inaction is still an acceptable and normal option. Action decisions, on the other hand, are perceived as a desirable option only by individuals who are predisposed to action; since an action decision is beyond the default and it takes more effort and intent to choose it, such a decision will not fit all individuals. This extends the norm theory of Kahneman and Miller (1986), suggesting that individual differences (at least with respect to the regulatory focus) are more notable in regret following action than in regret following inaction. We encourage future studies to further test the asymmetric effect of individual regulatory focus as well as other individual tendencies on action vs. inaction decisions.

### Prior Negative Outcome and Trigger for Change

Zeelenberg et al. (2002) suggested that when there is information regarding a prior negative outcome, action decision becomes more normal and acceptable, and therefore is less regrettable, than when no such information exists. Our results support this idea that a prior negative outcome makes action more normal than otherwise and consequently reduces regret following action. However, we go beyond this and find that not only prior negative outcomes but also other situational cues that signal the need for change, such as changing a weekly target at work (scenario 5), reduce regret following action. We suggest that a trigger for change makes action more normal than without such a trigger, and therefore it reduces regret following action, in line with the norm theory of Kahneman and Miller (1986), which suggests that regret is greater when it follows less normal decisions. Our findings add to other studies that show particular situations in which action is the norm and therefore produces less regret, such as the decision of goalkeepers in penalty kicks to jump (Bar-Eli et al., 2007). However, although our results show that a trigger for change reduces the level of regret attributed to action, it is not reversing the regret attribution pattern. In the three scenarios that contained a trigger for change (2, 5 and 6), only in scenario 2 a reversed pattern was evident (i.e., inaction was perceived as more regrettable than action). However, even in scenario 2, where the percent of attributing more regret to action is only 44.8% (the lowest among the scenarios), it is not statistically significant different from 50%.

### Practical Implications

An implementation of our results to decision making situations in both individual and organizational contexts would be to select promotion-focused individuals for decision making assignments in which actions must be made. Since there is less regret following action among promotion-focused individuals, it is more likely that such individuals will have less regret aversion and will be more willing to take action when it is needed.

Examples of contexts in which action decisions are mostly preferred would be Hi-Tech industries, or organizations who operate in a dynamic and turbulent environments that require frequent changes in technology, products, human resources, and so on. Another context that requires action decisions would be an entrepreneurial environment, in which individuals must be creative and innovative, discover opportunities, and develop new products. We are not suggesting that only promotion-focused individuals are required to make decisions in such environments and contexts, but in comparison to stable environments, high doses of promotion-focused individuals would be desirable. In contrast, in stable and less dynamic environments, changes and action decisions are required less frequently, and therefore the advantage of promotion-focused individuals is less significant. Yet, as our results show, inaction decisions are generally more preferred and less regretted by all individuals, regardless of their regulatory focus. Therefore, in steady environments, we suggest that both prevention—and promotion-focused individuals will tend to prefer inaction decisions. However, this idea needs further examination in both lab and field studies.

Another practical implication for effective decision making in organizations stems from our findings regarding the effect of a “trigger for change.” In order to encourage action decisions (in contexts that require changes), a useful suggestion would be to provide such triggers for change. For example, a manager who emphasizes to the employees the differences between the current situation and the previous one creates more triggers for change than a manager who emphasizes the similarities between the situations or who does not emphasize anything. As another example, consider two universities in which the Dean asks the faculty to update their courses and propose beneficial changes to the program. In the first university the Dean emphasizes that due to increased competition from colleges there is a reduced demand for the program. In the second university, although the situation is similar, the Dean just asks to try to improve the program as much as possible, or may be even emphasizes the similarities (e.g., that after the proposed changes, courses should still be semester-based, and the BA should still take 4 years). The first Dean, who emphasizes the changes in the environment, creates a trigger for change, and therefore is likely to encourage a more proactive and innovative mindset, more changes, and more needed action than the second Dean who did not create a trigger for change. According to our findings, triggers for change reduce the level of regret from action decisions, and thus increase the tendency to adopt action decisions.

### Research Limitations and Future Research

One limitation of our study is that the regulatory focus manipulation did not produce the expected effect. The current manipulation was chosen because we observed in other studies that Israeli subjects do not react as expected to the more common manipulations for regulatory focus (Higgins et al., 2001; Freitas and Higgins, 2002), i.e., these manipulations did not create promotion and prevention foci in Israeli samples. One possible reason is different interpretation of Israelis (compared to American subjects) of the terms used in Higgins' manipulations, namely oughts, duties, and obligations vs. ideals, dreams and

aspirations. The manipulation that we have used is based on similar technique used by Lockwood et al. (2002) and it was recently tested by Schödl and Van Dijk (2014). Although the manipulation was independent of the individual measure of regulatory focus (because it was randomly allocated to subjects and because it was carried out after measuring individual regulatory focus), as another precaution we controlled for the potential effect of the manipulation by adding it as an independent variable in the regressions.

Another limitation of our study is that we tested the effect of regulatory focus on regret with hypothetical scenarios rather than creating true regret in the individual. However, creating real regret in the individual is very difficult. One needs to have the subject make a decision, then to make sure the decision results in failure so that a potential regret may arise. Even then, if the individuals do not attribute the failure to a significant mistake they made, they might not feel regret. For example, if one guesses the numbers in a lottery and then does not win, he probably does not feel strong regret, because there was no way in which he could know the winning numbers. So running an experiment in which subjects make decisions and the experimenter informs them that they made a mistake and they lose, will not necessarily create regret. Furthermore, even if one can design an experiment that creates real regret in the lab, it is likely to be regret about losing small amounts of money in an artificial setting. On the other hand, with the scenarios we were able to describe situations that involve more significant regret than losing a few dollars, and with a greater diversity of situations. By using six different hypothetical scenarios in different contexts, three different measures for regret, and two different measures for individual regulatory focus, we further increase the robustness, validity and the richness of the results. Although the above arguments explain our choice of hypothetical scenarios, it is a worthwhile direction for future research, albeit not an easy one, to think about lab experiments with real consequences that induce regret and use them to analyze how personality differences in general and regulatory focus in particular affect regret. Such studies may be interesting complements to our results.

## REFERENCES

- Aaker, J. L., and Lee, A. Y. (2001). "I" seek pleasures and "we" avoid pains: The role of self-regulatory goals in information processing and persuasion. *J. Consum. Res.* 28, 33–49. doi: 10.1086/321946
- Avnet, T., and Higgins, E. T. (2006). How regulatory fit affects value in consumer choices and opinions. *J. Mark. Res.* 43, 1–10. doi: 10.1509/jmkr.43.1.1
- Azar, O. H. (2013). *The Impact of Risk and the Past Performance of a Strategy on Strategic Decision Making*. Working Paper, Ben-Gurion University of the Negev, Beer Sheva.
- Bar-Eli, M., Azar, O. H., Ritov, I., Keidar-Levin, Y., and Schein, G. (2007). Action bias among elite soccer goalkeepers: The case of penalty kicks. *J. Econ. Psychol.* 28, 606–621. doi: 10.1016/j.jeop.2006.12.001
- Baron, J., and Ritov, I. (1994). Reference points and omission bias. *Organ. Behav. Hum. Decis. Process.* 59, 475–498. doi: 10.1006/obhd.1994.1070
- Camacho, C. J., Higgins, E. T., and Luger, L. (2003). Moral value transfer from regulatory fit: what feels right is right and what feels wrong is wrong. *J. Pers. Soc. Psychol.* 84, 498–510. doi: 10.1037/0022-3514.84.3.498
- Cesario, J., Grant, H., and Higgins, E. T. (2004). Regulatory fit and persuasion: Transfer from "feeling right." *J. Pers. Soc. Psychol.* 86, 388–404. doi: 10.1037/0022-3514.86.3.388
- Chernev, A. (2004). Goal orientation and consumer preference for the status quo. *J. Consum. Res.* 31, 557–565. doi: 10.1086/425090
- Church, E. M., and Iyer, L. S. (2012). An experimental investigation of regulatory orientation and post-choice regret in online product selection. *J. Electron. Commer. Res.* 13, 289–304. Retrieved from: <http://search.proquest.com/docview/1266226039/accountid=14484>
- Crowe, E., and Higgins, E. T. (1997). Regulatory focus and strategic inclinations: Promotion and prevention in decision-making. *Organ. Behav. Hum. Decis. Process.* 69, 117–132. doi: 10.1006/obhd.1996.2675
- Das, N., and Joffe, B. (2012). The interactive effects of decision-making and expertise on the experience of regret. *J. Appl. Bus. Res.* 28, 333–346. doi: 10.19030/jabrv28i3.6953
- Das, N., and Kerr, A. H. (2010). "Woulda, coulda, shoulda:" a conceptual examination of the sources of post purchase regret. *J. Mark. Theory Pract.* 18, 171–180. doi: 10.2753/MTP1069-6679180205

Future studies should explore the impact of one's religiosity level on regret following action vs. inaction. Our findings show that orthodox people tend to attribute more regret than non-orthodox to a person who made an action decision. One explanation could be that orthodox people are more conservative and oriented to keep the status quo and avoid changes and risks. However, this finding is found only when using the dichotomous measure of regret and was not replicated with other measures of regret. Therefore, more research is needed in order to verify this effect.

Further research can be useful in order to verify our findings about the influence of regulatory focus on regret and confirm it in diverse situations, with different samples of subjects. We suggest to further explore the asymmetric effect of regulatory focus on action vs. inaction. An interesting direction would be to examine whether inaction is a type of decision that is perceived as the norm by most individuals, regardless of their personality, whereas an action decision is perceived differently according to the individual tendency, because it is considered as a less normal strategy. Another direction could be to present to the subjects various scenarios in different orders and analyze whether the order makes a difference. Additionally, the trigger for change should be tested in future studies in order to clarify and identify what types of information are perceived as a trigger for change, and consequently weaken the general tendency to regret more action than inaction decisions. Finally, the interaction effect that was found between the trigger for change and regulatory focus calls for future research to explore whether (and in what conditions) a trigger for change, which signals deviation from the norm, increases the impact of individual differences on regret feeling.

## AUTHOR CONTRIBUTIONS

AI: Experimental and study design, Analysis of the results, Writing the article. DV: Experimental and study design, Analysis of the results, Writing the article. OA: Experimental and study design, Analysis of the results, Writing the article.

- Förster, J., Higgins, E. T., and andWerth, L. (2004). How threat from stereotype disconfirmation triggers self-defense. *Soc. Cogn.* 22, 54–74. doi: 10.1521/soco.22.1.54.30982
- Freitas, A. L., and Higgins, E. T. (2002). Enjoying goal-directed action: the role of regulatory fit. *Psychol. Sci.* 13, 1–6. doi: 10.1111/1467-9280.00401
- Friedman, R. S., and Förster, J. (2001). The effects of promotion and prevention cues on creativity. *J. Pers. Soc. Psychol.* 81, 1001–1013. doi: 10.1037/0022-3514.81.6.1001
- Gilovich, T., and Medvec, V. H. (1995). The experience of regret: What, when, and why. *Psychol. Rev.* 102, 379–395. doi: 10.1037/0033-295X.102.2.379
- Gleicher, F., Kost, K. A., Baker, S. M., Strathman, A. J., Richman, S. A., and Sherman, S. J. (1990). The role of counterfactual thinking in judgments of affect. *Pers. Soc. Psychol. Bull.* 16, 284–295. doi: 10.1177/0146167290162009
- Higgins, E. T. (1997). Beyond pleasure and pain. *Am. Psychol.* 52, 1280–1300. doi: 10.1037/0003-066X.52.12.1280
- Higgins, E. T. (1998). Promotion and prevention: regulatory focus as a motivational principle. *Adv. Exp. Soc. Psychol.* 30, 1–46. doi: 10.1016/S0065-2601(08)60381-0
- Higgins, E. T. (2000). Making a good decision: value from fit. *Am. Psychol.* 55, 1217. doi: 10.1037/0003-066x.55.11.1217
- Higgins, E. T. (2005). Value from regulatory fit. *Am. Psychol. Soc.* 14, 209–213. doi: 10.1111/j.0963-7214.2005.00366.x
- Higgins, E. T. (2006). Value from hedonic experience and engagement. *Psychol. Rev.* 113:439. doi: 10.1037/0033-295X.113.3.439
- Higgins, E. T., Chen Idson, L., Freitas, A. L., Spiegel, S., and Molden, D. C. (2003). Transfer of value from fit. *J. Pers. Soc. Psychol.* 84, 1140–1153. doi: 10.1037/0022-3514.84.6.1140
- Higgins, E. T., Friedman, R. S., Harlow, R. E., Idson, L. C., Ayduk, O. N., and Taylor, A. (2001). Achievement orientations from subjective histories of success: promotion pride versus prevention pride. *Eur. J. Soc. Psychol.* 31, 3–23. doi: 10.1002/ejsp.27
- Higgins, E. T., Shah, J., and Friedman, R. (1997). Emotional responses to goal attainment: strength of regulatory focus as moderator. *J. Pers. Soc. Psychol.* 72, 515–525. doi: 10.1037/0022-3514.72.3.515
- Higgins, E. T., and Tykocinski, O. (1992). Self-discrepancies and biographical memory: personality and cognition at the level of psychological situation. *Soc. Pers. Soc. Psychol.* 18, 527–535. doi: 10.1177/0146167292185002
- Kahneman, D., and Miller, D. T. (1986). Norm theory: comparing reality to its alternatives. *Psychol. Rev.* 93, 136–153. doi: 10.1037/0033-295X.93.2.136
- Kahneman, D., and Tversky, A. (1982). The psychology of preferences. *Sci. Am.* 246, 160–173. doi: 10.1038/scientificamerican0182-160
- Kwak, J., and Park, J. (2008). "Sunk cost effects in consumer decisions: role of anticipated regret," in *Proceedings of the Advances in Consumer Research - North American Conference* (Duluth, MN), 719.
- Landman, J. (1987). Regret and elation following action and inaction: affective reactions to positive versus negative outcomes. *Pers. Soc. Psychol. Bull.* 13, 524–536. doi: 10.1177/0146167287134009
- Lee, A. Y., and Aaker, J. L. (2004). Bringing the frame into focus: the influence of regulatory fit on processing fluency and persuasion. *J. Pers. Soc. Psychol.* 86:205. doi: 10.1037/0022-3514.86.2.205
- Lehmann, D., and Siebzehner, B. (2009). Power, boundaries and institutions: marriage in ultra-orthodox judaism. *Eur. J. Soc.* 50, 273–308. doi: 10.1017/S0003975609990142
- Lberman, N., Idson, L. C., Camacho, C. J., and Higgins, E. T. (1999). Promotion and prevention choices between stability and change. *J. Pers. Soc. Psychol.* 77, 1135–1145. doi: 10.1037/0022-3514.77.6.1135
- Lockwood, P., Jordan, C. H., and Kunda, Z. (2002). Motivation by positive or negative role models: regulatory focus determines who will best inspire us. *J. Pers. Soc. Psychol.* 83, 854. doi: 10.1037/0022-3514.83.4.854
- McElroy, T., and Dowd, K. (2007). Action orientation, consistency and feelings of regret. *Judgm. Decis. Mak.* 2, 333–341. Retrieved from: <http://search.proquest.com/docview/1010989381?accountid=14484>
- Miller, D. T., and Taylor, B. R. (1995). "Counterfactual thinking, regret, and superstition: how to avoid kicking yourself," in *What Might Have Been: The Social Psychology of Counterfactual Thinking*, eds N. J. Roese and J. M. Olson (Hillsdale, NJ: Erlbaum), 305–331.
- N'gble, A., and Branscombe, N. R. (1997). When does action elicit more regret than inaction and is counterfactual thinking the mediator of this effect? *J. Exp. Soc. Psychol.* 33, 324–343.
- Ordóñez, L. D., and Connolly, T. (2000). Regret and responsibility: a reply to Zeelenberg et al. (1998). *Organ. Behav. Hum. Decis. Process.* 81, 132–142. doi: 10.1006/obhd.1999.2834
- Ritov, I., and Baron, J. (1995). Outcome knowledge, regret and omission bias. *Organ. Behav. Hum. Decis. Process.* 64, 119–127. doi: 10.1006/obhd.1995.1094
- Roese, N. J., Epstude, K., Fessel, F., Morrison, M., Smallman, R., Summerville, A., et al. (2009). Repetitive regret, depression, and anxiety: Findings from a nationally representative survey. *J. Soc. Clin. Psychol.* 28, 671–688. doi: 10.1521/jscp.2009.28.6.671
- Roese, N. J., Summerville, A., and Fessel, F. (2007). Regret and behavior: comment on zeelenberg and pieters. *J. Consum. Psychol.* 17, 25–28. doi: 10.1207/s15327663jcp1701\_5
- Schödl, M. M., Kluger, A. N., and Van Dijk, D. (2013). "Regulatory focus measurement confusion: prevention/promotion or approach/avoidance?" in *A Paper Presented at the European Association of Work and Organizational Psychology* (Münster).
- Schödl, M. M., and Van Dijk, D. (2014). "I have a dream, but I ought to do something else: time allocation to prevention and promotion goals," in *A paper presented at the 74th Annual Meeting of the Academy of Management* (Philadelphia, PA).
- Seta, J. J., McElroy, T., and Seta, C. E. (2001). To do or not to do: Desirability and consistency mediate judgments of regret. *J. Pers. Soc. Psychol.* 80, 861–870. doi: 10.1037/0022-3514.80.6.861
- Shah, J., Higgins, E. T., and Friedman, R. S. (1998). Performance incentives and means: how regulatory focus influences goal attainment. *J. Pers. Soc. Psychol.* 74, 285–293. doi: 10.1037/0022-3514.74.2.285
- Van Dijk, D., Seger-Guttmann, T., and Heller, D. (2013). Life-Threatening event reduces subjective well-being through activating avoidance motivation: a longitudinal study. *Emotion* 13, 216–225. doi: 10.1037/a0029973
- Van der Pligt, J., Zeelenberg, M., and Manstead, A. S. (1998). Undoing regret on dutch television: apologizing for interpersonal regrets involving actions or inactions. *Pers. Soc. Psychol. Bull.* 24, 1113–1119. doi: 10.1177/01461672982410008
- Zeelenberg, M., and Beattie, J. (1997). Consequences of regret aversion 2: additional evidence for effects of feedback on decision making. *Organ. Behav. Hum. Decis. Process.* 72, 63–78. doi: 10.1006/obhd.1997.2730
- Zeelenberg, M., Van den Bos, K., Van Dijk, E., and Pieters, R. G. M. (2002). The inaction effect in the psychology of regret. *J. Pers. Soc. Psychol.* 82, 327–334. doi: 10.1037/0022-3514.82.3.314

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Itzkin, Van Dijk and Azar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## APPENDIX: THE SCENARIOS USED IN THE EXPERIMENT

### Scenario 1: Stock Investment (Kahneman and Tversky, 1982)

Paul owns shares in Company A. During the past year he considered switching to stock in company B, but he decided against it. He now finds out that he would have been better off by \$1200 if he had switched to the stock of Company B. George owned shares in Company B. During the past year he switched to stock in Company A. He now finds out that he would have been better off by \$1200 if he had kept his stock in Company B. Who feels more regret?

### Scenario 2: Soccer Teams (Zeelenberg et al., 2002)

Jacob and Noah are both coaches of a soccer team. Jacob is the coach of team A, and Noah is the coach of team B. Both coaches lost the prior game with a score of 4–0. This Sunday Jacob decides to do something: He fields three new players. Noah decides not to change his team. This time both teams lose with 3–0. Who feels more regret, coach Jacob or coach Noah?

### Scenario 3: Project Management

Shirley and Rene are both project managers in a global company. As part of their jobs they decide with which projects to continue and which to terminate every quarter based on performance. At the beginning of the year, both of them were required to make a decision regarding projects that started earlier. Shirley decided to terminate project A and switch it with project B. Rene on the other hand decided to continue with project C that she started earlier. At the end of the year it turned out that both projects B and C failed, produced losses, and it was decided to terminate them.

1. Who feels more regret, Shirley or Rene?
2. a. What is the level of regret that Shirley feels on a scale of 0–100 (0 - no regret at all, 100 - very high level of regret)?
- b. What is the level of regret that Rene feels on a scale of 0–100 (0 - no regret at all, 100 - very high level of regret)?

### Scenario 4: Supplier Choice

Emma and Mia both work as purchasing managers in a big pharmaceutical company. As part of their jobs they decide with which raw materials suppliers to work. The company has been purchasing a variety of raw materials for the past five years from supplier A. Emma needed raw material X and received for it offers from both supplier A and supplier B, who is a supplier that has not yet been working with the company. Mia needed raw material

Y and received for it offers from both supplier A and supplier C, who is a supplier that has not yet been working with the company. Emma decided to purchase the raw material X from the new supplier B. Mia decided to purchase the raw material Y from the old supplier A. After some time it was discovered that both new raw materials X and Y, from both suppliers B and A respectively, were of low quality and caused the company losses.

1. Who feels more regret, Emma or Mia?
2. a. What is the level of regret that Emma feels on a scale of 0–100 (0 - no regret at all, 100 - very high level of regret)?
- b. What is the level of regret that Mia feels on a scale of 0–100 (0 - no regret at all, 100 - very high level of regret)?

### Scenario 5: Machine Parameters

Michael and Daniel are both machine operators in a company that manufactures plastic products. Every week each of them receives his weekly target and has to make sure that the machine under his responsibility will produce this target. This week the target was higher than usual for both of them and therefore Michael and Daniel pondered what to do. Michael decided to change the machine parameters. Daniel decided to stay with the regular parameters. At the end of the week both Michael and Daniel did not succeed to reach the weekly target.

1. Who feels more regret, Michael or Daniel?
2. a. What is the level of regret that Michael feels on a scale of 0–100 (0 - no regret at all, 100 - very high level of regret)?
- b. What is the level of regret that Daniel feels on a scale of 0–100 (0 - no regret at all, 100 - very high level of regret)?

### Scenario 6: Academic Studies (Based in Part on Gilovich and Medvec, 1995)

Roy and Alex both studied for a Bachelor's degree in management and decided to continue to a Master's degree in business administration in the same university. After a short period in the degree both Roy and Alex felt that the degree is not contributing to them and the general feeling was that the attitude towards them is unpleasant and they do not enjoy the degree. Roy and Alex each considered whether to quit the university for a similar track in another university. Roy decided to stay and Alex decided to move to a different university. After half a year, they met and updated each other. They found that both of them are still unsatisfied with the degree they study.

1. Who feels more regret, Roy or Alex?
2. a. What is the level of regret that Roy feels on a scale of 0–100 (0 - no regret at all, 100 - very high level of regret)?
- b. What is the level of regret that Alex feels on a scale of 0–100 (0 - no regret at all, 100 - very high level of regret)?



# Don't Always Prefer My Chosen Objects: Low Level of Trait Autonomy and Autonomy Deprivation Decreases Mere Choice Effect

Zhe Shang, Tuoxin Tao and Lei Wang\*

Department of Psychology and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China

Choice effect is a robust phenomenon in which even “mere choice” that does not include actual choosing actions could result in more preference for the self-chosen objects over other-chosen objects. In the current research, we proposed that autonomy would impact the mere choice effect. We conducted two studies to examine the hypothesis. The results showed that the mere choice effect measured by Implicit Association Test (IAT) significantly decreased for participants with lower levels of trait autonomy (Study 1) and when participants were primed to experience autonomy deprivation (Study 2). The theoretical and practical implications are discussed.

## OPEN ACCESS

### Edited by:

Aurora García-Gallego,  
Universitat Jaume I, Spain

### Reviewed by:

Wing-Yee Cheung,  
University of Southampton, UK

Eric Mayor,  
University of Neuchâtel, Switzerland

### \*Correspondence:

Lei Wang  
leiwang@pku.edu.cn

### Specialty section:

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

**Received:** 31 December 2015

**Accepted:** 29 March 2016

**Published:** 19 April 2016

### Citation:

Shang Z, Tao T and Wang L (2016)  
Don't Always Prefer My Chosen Objects: Low Level of Trait Autonomy and Autonomy Deprivation Decreases Mere Choice Effect.  
*Front. Psychol.* 7:524.  
doi: 10.3389/fpsyg.2016.00524

## INTRODUCTION

People make choices according to their preferences, indicating the important role that preferences play in choices. In addition, choice also has an impact on post-choice preferences. After the choice has been made, people's liking of the chosen objects tends to increase while that of the rejected objects tends to decrease, known as the post-decisional spreading of alternatives (Brehm, 1956; Hammock and Brehm, 1966). In other words, people would prefer one of two similar objects simply because they chose one rather than the other, which is also known as the choice effect (Huang et al., 2009). Since the initial work of Brehm (1956), this phenomenon has got widespread attention (Ariely and Norton, 2008). The choice-induced preference has been found to exist strongly in several forms, such as in real choice actions (Pattal et al., 2008) and illusory choices (Huang et al., 2009).

Over the past decades, the cognitive dissonance theory and self-concept related theories have been widely used in explaining the mechanism underlying the choice-induced preferences. The main classical explanation is based on the cognitive dissonance theory (Festinger, 1957), which argues that people are motivated to maintain internal consistency between cognitive inputs and behavioral outputs to reduce the uncomfortable feeling of dissonance. An individual is likely to experience cognitive dissonance if he/she holds negative attitudes toward an object, given that the object has been chosen by oneself, because thus a conflict would occur between the cognitive input (“I don't like this thing”) and the behavioral output (“I chose this thing”) (Van Overwalle and Jordens, 2002). In order to reduce this uncomfortable feeling of cognitive dissonance, individuals would increase their liking of the chosen objects when the actual choice action has been taken (Olson and Stone, 2005). This theory helps to explain why people prefer a chosen object to an unchosen one simply because they took an explicit action to choose the object. However, when there is no explicit choosing action, the cognitive dissonance theory loses its power in explaining

the mechanism of the choice-induced preference. Recent empirical research has indicated that the choice effect happens even when individuals lack the awareness of their choosing behaviors (Lieberman et al., 2001; Coppin et al., 2010), suggesting that cognitive dissonance may not be a necessary prerequisite of choice-induced preference. Indeed, choice-induced preferences are found even when the choices were seemingly trivial (Langer and Rodin, 1976) or wholly illusory (Langer, 1975). That is, choice-induced preferences existed even when there was no explicit choice action and thus the awareness of cognitive dissonance may not be present. The phenomenon where choice itself is powerful enough to induce liking, even in the condition that choosing is illusory and does not actually occur, was termed by Huang et al. (2009) as the mere choice effect.

The theory referring to the positive valence of self node helps to explain the choice effect as well (Greenwald et al., 2002). "Self node" means that self was treated as a node in the self-related concept tree. "Self" is the sum of all that one can call his/her own (James, 1890). "My choice" is also a part of the self-concept. Theories and phenomena associated with self-serving or self-protecting biases (Sedikides and Strube, 1997), such as self-enhancement (Kurman, 2001), self-affirmation (Brown and Dutton, 1995), and self-verification (Chen et al., 2006) imply that people are prone to evaluate "my choice" as better than "others' choice" to maintain a positive self-image, and thus would display a positive evaluation on self-chosen objects. "Self node" affects the choice-preference link by increasing preference on self-chosen objects in an implicitly way, which leads to mere choice effect. On the other hand, to choose is to express a preference and to assert the self (Leotti et al., 2010). Consequently, attaching a high evaluation to "my choice" implies acceptance of the self and thus in turn brings higher self-satisfaction and self-esteem. The scope of self-concept is broader than just one's possessions or decisions (choices). As the Ryan and Deci (2000) theorized self-determination theory (SDT), the need for competence, autonomy, and psychological relatedness are three psychological needs that motivate the self to initiate behavior (Deci and Ryan, 1985, 2000). The act of self-regulation, such as autonomy, is also related to self-concept. Experiencing autonomy promotes the sense that an individual's behavior is self-motivated and self-determined and thus maintain a positive self-image. Applied in the objects evaluation, another possible theoretical explanation refers to the role that autonomy plays in the choice effect.

The sense of autonomy refers to the extent to which people feel free to make their own decisions and experience a sense of volition in their actions (van Prooijen, 2009). Choosing behavior increases the experience of autonomy by allowing people to exert their right to make a decision. Previous research has demonstrated that people evaluate the chosen alternative as more desirable than the rejected alternative, in order to reassert their autonomy (Hammock and Brehm, 1966). Experiments have suggested that manipulations designed to enhance one's experience of autonomy can boost intrinsic motivation and energize behavior (Swann and Pittman, 1977; Zuckerman et al., 1978; Simon and McCarthy, 1982, Unpublished). Offering people an optimal amount of choice enhanced their intrinsic

motivation and energy to persist (e.g., deCharms, 1968; Deci and Ryan, 1985). As demonstrated by plenty of research, autonomy is associated with intrinsic motivation (Deci et al., 1999), persistence (Moller et al., 2006), goal attainment (Sheldon and Elliot, 1998), and creativity (Sheldon, 1995), indicating that autonomy elicits positive outcomes. Additionally, perceived autonomy has an effect on enhancing happiness (Chekola, 2007; Demir et al., 2011), job satisfaction, and a general increase in subjective well-being (Sheldon et al., 2004), all of which conclude that autonomy elicits positive personal feelings. Preference for an object represents the positive objective valence that one attaches to the object in the process of evaluation. As has been validated by previous research, personal positive state and feelings influence evaluation, in terms of increasing personal preference/sensitivity on surroundings and targeted objects (Gu et al., 2010; Yang et al., 2014). It is possible that the sense of autonomy elicited by choosing enhances an individual's evaluation toward an object, because experiencing autonomy induces positive feelings, which in turn have a positive effect on the evaluation of the object.

We speculate that the experience of autonomy may enhance the preference on self-chosen objects. Thus, we measured the relationship of trait autonomy and choice effect in Study 1 and propose the hypothesis: (1) trait autonomy is positively correlated with choice effect. We infer that autonomy may moderate the choice-preference link. When an individual takes a choosing action or is simply acknowledged that something has been chosen by himself/herself, the sense of autonomy is generated, which brings him/her positive feelings. These positive feelings in turn may enhance his/her positive evaluation to the surroundings. On the contrary, the lack of autonomy may reduce the preference for self-chosen objects. To our best knowledge, however, no study has provided empirical evidence for the role of autonomy in the choice effect. In Study 2, we investigated the influence of different levels of autonomy experience on the choice effect by using a priming task to set three conditions: the autonomy fulfillment condition, the autonomy deprivation condition, and the control condition. Here we propose the hypothesis: (2) the choice effect would occur in the autonomy fulfillment condition and in the control condition, but not in the autonomy deprivation; (3) autonomy deprivation would decrease or even eliminate the choice effect when compared with the autonomy fulfillment condition; (4) autonomy deprivation would decrease or even eliminate the choice effect when compared with the control condition; (5) autonomy fulfillment would increase the choice effect when compared with the control condition.

## Overview of Two Studies

In order to study the influence of autonomy on the choice effect while excluding the impact of cognitive dissonance, we employed a modified illusory choice paradigm, adapted from Huang et al. (2009) to measure the presence of the mere choice effect. We adopted the Implicit Association Test (IAT, Greenwald et al., 1998) that records the response time when participants respond to settled categories of objects framed by positive or negative adjectives.

In measuring autonomy, we treated it as an individual difference variable (Deci and Ryan, 2000). It can be either dispositional or situational. Thus, we tested our hypothesis through two studies. In Study 1, we recorded participants' self-report trait autonomy and divided participants into high and low autonomy groups accordingly. Study 2 adopted a priming paradigm to manipulate the situational autonomy in three levels.

## **STUDY 1: THE RELATIONSHIP BETWEEN TRAIT AUTONOMY (BETWEEN-SUBJECT VARIABLE) AND THE MERE CHOICE EFFECT**

In Study 1, we used a scale to measure trait autonomy as an individual differential variable. Subsequently, we tested the mere choice effect using an IAT paradigm. We then calculated the relationship of trait autonomy and the mere choice effect.

### **Materials and Methods**

#### **Participants**

A total of 91 graduate and undergraduate students (50 female, 41 male, average age = 22.2 years,  $SD = 2.39$ , ranging from 19 to 25 years old) participated in the experiment for a cash reward (US\$2). We asked all the participants to conduct an object chosen task in which they would see some texts in a computer screen and react by push some buttons on the keyboard. Each of them wrote informed written consent before the test. They were told there would be no any dangers while they were doing the experiment. They were told their rights and they can decide to or not to participate in this experiment, and they had the right to quit the experiment at any time of the experiment. This study was in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Department of Psychology, Peking University.

#### **Procedure and Materials**

Participants were told that there were two unrelated tasks. After each experiment session, we asked the participant whether he/she thought the two parts were related. None of them replied yes. In the first task, they were required to complete questionnaires including trait autonomy and demographic survey. The second task was a mere choice task presented on computer, adapted from Huang et al. (2009), which created a mere choice situation to the participants. As previous research has demonstrated (Huang et al., 2009), when participants are asked to choose something for a third party, they would implicitly prefer the self-chosen object to the other-chosen objects (i.e., the choice effect).

#### **Trait Autonomy**

The five-item Choicefullness Subscale of the Self Determination Scale (Sheldon, 1995; Sheldon et al., 1996) was used to measure trait autonomy. Each item presented participants with two opposing statements. Participants were asked to indicate which of the two statements was more appropriate for describing themselves. An example item is showed as follows: "I always feel

like I choose the things I do" (Statement A) versus "I sometimes feel that it's not really me choosing the things I do" (Statement B) (5-likert scale: 1 = *only A feels true*; 5 = *only B feels true*). The answers were coded such that lower scores indicated lower level of autonomy. Our data showed good internal consistency (Cronbach's  $\alpha = 0.697$ )

#### **Mere Choice Effect**

To evaluate the mere choice effect (an affect reflected the degree of preference on self-chosen objects over other-chosen objects), both of the two studies used a modified illusory choice paradigm developed by Huang et al. (2009), in which participants were asked to imagine a scenario about choice instead of taking an actual choice action.

This part was completed on computers using the Inquisit laboratory software.

In Step 1, participants read the following two-page scenario on the computer screen:

*Please visualize the following scenario. You and your friend (marked as **the other** in the experiment) bought six products in a supermarket for another friend: a mug, a small figurine, a piece of chocolate, a piece of candy, a pen, and a ruler. Please visualize and remember these products. They will be used in the following experiment (Page 1).*

*Among these six products, three of them were chosen by you, and the other three were chosen by the friend (the other) shopping with you. You chose the mug, the chocolate, and the pen. Your friend chose the small figurine, the candy, and the ruler. Please spend 2 min to visualize and separately remember your choices and your friend's choices. They will be used in the following experiment (Page 2).*

Half of the participants were shown the aforementioned scenario. The other half read similar instructions except that we swapped the products assigned to the self and the other.

Then the participants began the modified illusory choice IAT (Huang et al., 2009). This IAT followed the procedure designed by Greenwald et al. (1998), involving two target categories (objects chosen by the self vs. objects chosen by the other) and two attribute categories (positive vs. negative). Target categories followed the scenario described previously. The attribute categories were previously used in many studies (Maison et al., 2004; Huang et al., 2009). The positive stimuli included the Sun, luck, love, fun, happiness, pleasure, holiday, and friendship. The negative stimuli included disease, death, murder, accident, poison, war, tragedy, and vomit. Our study was consistent with the classical IAT paradigm (Lane et al., 2007), target words and attribute words were presented together in the IAT paradigm. In the two main tasks of IAT, there were two situations: in one situation, the words "self-chosen objects/positive attributes" appeared in the top left-hand corner while the words "other-chosen objects/negative attributes" appeared in the top right-hand corner; in the other situation, the words "self-chosen objects/negative attributes" appeared in the top left-hand corner while the words "other-chosen objects/positive attributes" appeared in the top right-hand corner. As can be seen, in both situations, the self-chosen and other-chosen objects appeared together as the

target words, making it impossible to analyze their effects separately.

The IAT consisted of five classification tasks (see **Table 1**): attribute discrimination task (Block 1, 24 trials), initial target-category discrimination task (Block 2, 24 trials), initial combined task (Block 3, 24 trials for practice, and Block 4, 48 trials for data collection), reversed target-category discrimination task (Block 5, 48 trials), reversed combined task (Block 6, 24 trials for practice and Block 7, 48 trials for data collection).

In the attribute discrimination task (Block 1, 24 trials), participants were asked to press a left key (F) when a positive word appeared on the screen and a right key (J) for a negative word. Similarly, in the initial target-category discrimination task (Block 2, 24 trials), objects chosen by the self (responding by pressing the left key) and objects chosen by the other (responding by pressing the right key) were discriminated. In the initial combined task (Block 3, 24 trials for practice and Block 4, 48 trials for data collection), attribute and target discrimination trials were combined and participants had to press the left key when either a positive word or an object chosen by the self was presented and the right key when a negative word or an object chosen by the other was presented (the compatible condition, we replicated the IAT paradigm in accordance with a previous study (Huang et al., 2009), in which the participants' responses showed that self-chosen objects were implicitly linked with positive words (e.g., happiness, sunshine), as opposed to negative words (e.g., death, war), and in which other-chosen objects were implicitly linked to negative words, as opposed to positive words. Thus, we argue that the compatible condition was composed of self-chosen objects with positive descriptions and other-chosen objects with negative descriptions, just as Huang et al., 2009 showed). In the reversed target-category discrimination task (Block 5, 48 trials), Block 2 was repeated with a switch of the categorization keys by pressing left key when an object chosen by the other appeared on the screen and a right key when an object chosen by the self appeared. The reversed combined task (Block 6, 24 trials for practice and Block 7, 48 trials for data collection) again combined two individual tasks. Participants were instructed to press the left key when either a positive word or an object chosen by the other was presented and press the right key when a negative word or an object chosen by the self was presented (incompatible condition). Each block started with a brief instruction for the following task and a request to respond

as fast as possible while trying to minimize mistakes. Participants were also reminded that their error rate and response times would be recorded.

Different random orders of trials were used for different participants. Half of the participants went through the seven blocks in the order presented previously; to remove any order effect, Blocks 2, 3, and 4 were swapped with Blocks 5, 6, and 7 for the other half of the participants. Only data from Blocks 4 and 7 were used for analysis. Each block started with a brief instruction.

After each experiment session, the participant was fully debriefed, thanked, and paid for his/her participation.

## Results

We analyzed the data following the processes suggested by Greenwald et al. (1998). The first two trials of each block were excluded since the response latencies for them were typically longer. Next, we recoded the latencies by excluding reaction times (RTs) that were below 300 ms or above 3000 ms, so that we could control for outlying trials where distraction and anticipation likely affected the trial. We disregarded any participant with an error rate above 30%. Thus, our final data analysis included 87 participants (46 female, 41 male, average age = 21.1 years,  $SD = 2.36$ , ranging from 18 to 25 years old).

In the IAT task, the compatible condition was composed of self-chosen objects with positive descriptions and other-chosen objects with negative descriptions, while the incompatible condition was composed of self-chosen objects with negative descriptions and other-chosen objects with positive descriptions. The choice-attitude valence compatible level (the compatible condition and the incompatible condition) was a within-subject variable. We conducted a one-way repeated ANOVA of choice-attitude valence compatibility level (compatible condition vs. incompatible condition), after controlling for gender and age. Results showed a significant main effect,  $F(1,86) = 4.023$ ,  $p < 0.05$ ,  $\eta^2 = 0.046$ . Participants' RT in the compatible condition ( $M_{RT} = 753$  ms,  $SD = 206$  ms) was faster than that in the incompatible condition ( $M_{RT} = 882$  ms,  $SD = 214$  ms). We suggest that participants preferred the self-chosen objects with positive descriptions and other-chosen objects with negative descriptions over other-chosen objects with positive descriptions and self-chosen objects with negative descriptions. In other words, compared with perceived other-chosen objects, perceived

**TABLE 1 | Task process of the IAT paradigm in Studies 1 and 2.**

Block	Task	Trials	Response key assignment	
			Left key	Right key
1	Attribution discrimination	24	Positive	Negative
2	Initial target discrimination	24	Objects chosen by the self	Objects chosen by the other
3	Initial combined task	24	Positive; objects chosen by the self	Negative; objects chosen by the other
4	Initial combined task	48	Positive; objects chosen by the self	Negative; objects chosen by the other
5	Reversed target discrimination	48	Objects chosen by the other	Objects chosen by the self
6	Reversed combined task	24	Positive; objects chosen by the other	Negative; objects chosen by the self
7	Reversed combined task	48	Positive; objects chosen by the other	Negative; objects chosen by the self

self-chosen objects were more strongly associated with positive than with negative words, indicating that people implicitly preferred self-chosen objects to other-chosen objects, despite their lack of actual experience of a choosing process, namely the mere choice effect. This result is consistent with the previous study of Huang et al. (2009).

### Choice Effect and Autonomy

We used the difference response time (d-RT). It is the RT in the incompatible condition (other-chosen objects that were implicitly linked with positive words and self-chosen objects that were implicitly linked with negative words) minus the RT in the compatible condition (self-chosen objects that were implicitly linked with positive words and other-chosen objects that were implicitly linked with negative words) as the indicator of the choice effect (*Mean different RT* = 129 ms, *SD* = 174 ms). Longer d-RT indicated a larger choice effect while shorter d-RT indicated a smaller choice effect. In the meantime, lower scorers on the five-item Choicefulness Subscale of the Self Determination Scale indicated lower level of trait autonomy, and higher scores indicated higher level of trait autonomy. The mean score of trait autonomy was 14.78, and standard deviation was 3.289.

Hypothesis 1 proposed that trait autonomy is positively correlated with choice effect. We examined the effect of trait autonomy on choice effect by controlling for gender and age in a hierarchical analysis. We conducted a hierarchical regression analysis by entering gender in a first block/model, age in a second block/model, and the trait autonomy as the independent variable in a third block/model. All variables were normalized as Z-scores for data analysis. The regression coefficients, standard error, 95% confidence interval [CI], the change in *F* statistic (including *p*-value), and the coefficient of determination change (delta *R*<sup>2</sup>) for each model are shown in **Table 2**. The results of regression analysis showed that after controlling for gender and age, the  $\beta$  of trait autonomy on choice effect represented by d-RT in the IAT task was 0.341, (*SE* = 0.105, *p* < 0.01, 95% confidence interval [CI] = [0.132, 0.551]), which suggested a significant direct effect. **Table 2** shows that trait autonomy explained incremental variance of d-RT in IAT (11.1%), *p* < 0.01, suggesting that people with a higher level of trait autonomy showed a larger choice effect.

It should be noted that only the effect of trait autonomy on choice effect was obtained; the other two variables (gender, age) were not significant predictors of the criteria. This above-mentioned result provides support for Hypothesis 1.

## STUDY 2: THE INFLUENCE OF AUTONOMY ON THE CHOICE EFFECT

Study 2 is designed to extend the results of Study 1. According to the findings in Study 1, there is a positive correlation between trait autonomy and the choice effect. To further investigate the nature of this relationship, we examined whether the choice effect would remain when autonomy was deprived in a between-subject design. We aimed to test whether experimentally manipulated autonomy affects the choice effect. We repeated the steps of Study 1, except that we did not measure trait autonomy by questionnaire but manipulated the level of autonomy. In this study, the perceived autonomy was manipulated by a priming task, which comprised: an autonomy fulfillment condition, an autonomy deprivation recall condition, and a control condition.

### Materials and Methods

#### Participants

Sixty-five participants (38 women, 27 men; *M<sub>age</sub>* = 22.3, *SD* = 1.9, range from 18 to 27 years old), all of them were university students. They were randomly assigned to the three experimental conditions. Twenty-two participants were assigned in the autonomy-fulfillment condition, 21 participants were assigned in the autonomy-deprived condition, and 21 participants were assigned in the control condition. All the participants were informed of conducting an object chosen task during the recruitment and before the experiment. Informed written consent was obtained from each participant before the test. They were told there would be no any dangers while they were doing the experiment in which they would see some texts in a computer screen and react by push some buttons on the keyboard. They were told their rights and they can decide to or not to participate in this experiment, and they had the right to quit the experiment at any time of the experiment. They were

**TABLE 2 |** The hierarchical regression of predictors on choice effect in Study 1.

Predictor	The d-RT in IAT									
	95% CI (Confidence interval)	<i>R</i>	<i>R</i> <sup>2</sup>	$\Delta R^2$	<i>F</i>	<i>p</i>	$\beta$	<i>SE</i>	<i>p</i>	
Model 1		0.105	0.011	0.011	0.945	0.334				
Gender	[-0.110, 0.319]						0.105	0.108	0.334	
Model 2		0.111	0.012	0.001	0.121	0.729				
Gender	[-0.108, 0.325]						0.109	0.109	0.332	
Age	[-0.254, 0.179]						-0.038	0.109	0.729	
Model 3		0.351	0.123	0.111**	10.501**	0.002				
Gender	[-0.129, 0.284]						0.078	0.104	0.456	
Age	[-0.313, 0.106]						-0.104	0.105	0.327	
Trait autonomy	[0.132, 0.551]						0.341**	0.105	0.002	

*N* = 87. \*\**p* < 0.01.

rewarded for about 2 US dollars for their participation. This study was in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Department of Psychology, Peking University.

## Procedure and Materials

Participants were told that the session consisted of two separate tasks. The first task was introduced as focusing on the recall of past events, while the real purpose of it was to prime the sense of autonomy in participants. For example, participants in the autonomy fulfillment condition were asked to write an essay about a particular incident in which they felt high level of autonomy. The introduction went as follows:

### Manipulating Materials of Autonomy Experience

#### Autonomy-fulfillment condition

*"The first part of this session is story collection. Please describe an event about personal autonomy. Here, the autonomy is defined as when an individual is able to make their own choices freely, and experiences a sense of control over their decisions. (If you have any questions about this definition, please ask the experimenter.)*

*"Now, please write down an event based on your real experience, in which your autonomy was satisfied. Please elaborate the details as much as possible, including the objective circumstances and your subjective feelings."*

#### Autonomy-deprivation condition

*"The first part of this session is story collection. Please describe an event about personal autonomy. Here, the autonomy is defined as when an individual is able to make their own choices freely, and experiences a sense of control over their decisions. (If any questions about this definition, please ask the experimenter.)*

*"Now, please write down an event based on your real experience, in which your autonomy was deprived. That is to say, your behaviors were not completely controlled by yourself and some decisions were not self-decided. Please elaborate the details as much as possible, including the objective circumstances and your subjective feelings."*

After completing the recall task, participants completed the Choicefulness subscale of the Self Determination Scale (Sheldon, 1995; Sheldon et al., 1996), as a manipulation check of the autonomy priming. For the participants in the control condition, they were given no priming materials and completed the scale directly. Then, all the participants were introduced to what ostensibly was a second task: the IAT task, which was the same task as in Study 1. At last, participants were fully debriefed, thanked, and paid for their participation.

## Results

After applying the same data protocol used in Study 1, the final data analysis of Study 2 included 64 participants (37 women, 27 men;  $M_{age} = 21.7$ ,  $SD = 1.95$ , range from 18 to 26 years old).

### Manipulation Checks

For the manipulation check of autonomy priming, a one-way ANOVA revealed that there was a significant main effect of the autonomy manipulation (the autonomy fulfillment condition, the autonomy deprivation condition, the control

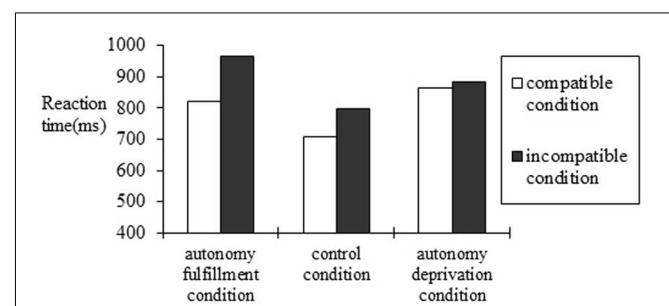
condition) on the Choicefulness Scale scores,  $F(2,61) = 3.190$ ,  $p = 0.048$ ,  $\eta^2 = 0.044$ . After controlling for gender and age, the ANOVA showed that there was no significant difference of Choicefulness Scale scores between the autonomy fulfillment condition ( $M = 16.86$ ,  $SD = 2.624$ ) and the control condition ( $M = 15.67$ ,  $SD = 2.850$ ),  $F(1,42) = 2.012$ ,  $p = 0.164$ ,  $\eta^2 = 0.048$ . While participants scored significantly higher ( $M = 15.67$ ,  $SD = 2.850$ ) on the Choicefulness Scale in the control condition than in the autonomy deprivation condition ( $M = 14.71$ ,  $SD = 2.918$ ),  $F(1,41) = 1.109$ ,  $p = 0.299$ ,  $\eta^2 = 0.028$ . Participants scored significantly higher ( $M = 16.86$ ,  $SD = 2.624$ ) in the autonomy fulfillment condition than those in the autonomy deprivation condition ( $M = 14.71$ ,  $SD = 2.918$ ),  $F(1,42) = 6.194$ ,  $p = 0.017$ ,  $\eta^2 = 0.137$ . This result confirmed the validity of autonomy manipulation in autonomy fulfillment condition and autonomy deprivation condition. That is, compared with those in autonomy deprivation condition, participants in autonomy fulfillment condition experienced a higher level of autonomy (see Table 3).

### Autonomy and the Choice Effect

To further identify how autonomy affects RT in compatible condition and incompatible condition, we analyzed a 3 (autonomy priming manipulation: the autonomy fulfillment condition, the control condition, and the autonomy deprivation condition) by 2 (choice-attitude valence compatibility level: the compatible condition and the incompatible condition) mixed design, in which the autonomy priming manipulation was a between-subject variable and the choice-attitude valence compatible level was a within-subject variable. A two-way repeated ANOVA of autonomy priming manipulation and choice-attitude valence compatible on response time was

**TABLE 3 | Autonomy scores in autonomy priming task's manipulation check in Study 2.**

	N	Mean scores	SD
Autonomy fulfillment	22	16.86	2.62
Control condition	21	15.67	2.85
Autonomy deprivation	21	14.71	2.92



**FIGURE 1 | Reaction time of compatible and incompatible evaluation conditions under the autonomy fulfillment, autonomy deprivation, and control condition in Study 2.**

conducted, controlling for gender and age (see **Figure 1**). The results showed a main effect: autonomy priming manipulation was significant,  $F(2,61) = 5.341, p = 0.007, \eta^2 = 0.15$ . The response time in the autonomy fulfillment group ( $M_{RT} = 892$  ms,  $SD = 45$  ms) was significantly longer than that in the control group ( $M_{RT} = 713$  ms,  $SD = 46$  ms),  $F(1,42) = 7.406, p = 0.01, \eta^2 = 0.155$ . The response time in the autonomy deprivation group ( $M_{RT} = 906$  ms,  $SD = 46$  ms) was also significantly longer than that in the control group ( $M_{RT} = 713$  ms,  $SD = 46$  ms),  $F(1,41) = 8.477, p = 0.006, \eta^2 = 0.174$ . The response time in the autonomy deprivation group ( $M_{RT} = 906$  ms,  $SD = 46$  ms) was not significantly longer than that in the autonomy fulfillment group ( $M_{RT} = 892$  ms,  $SD = 46$  ms),  $F(1,42) = 0.142, p = 0.709, \eta^2 = 0.003$ . These results indicated that autonomy priming (whether fulfillment or deprivation) led to slower participant RTs. A main effect of choice-attitude valence compatibility level on response time was significant,  $F(1,61) = 11.877, p = 0.001, \eta^2 = 0.15$ . The response time of incompatible trials ( $M = 882$  ms,  $SD = 28$  ms) was significantly longer than that of compatible trials ( $M = 793$  ms,  $SD = 28$  ms), indicating the conflict of objects and adjectives in the incompatible condition.

The interaction of priming manipulation and compatibility level was significant,  $F(2,61) = 4.550, p = 0.015, \eta^2 = 0.11$ , indicating that the response time was modulated by the priming manipulation. *Post hoc* analyses showed that, in the autonomy fulfillment condition, participants responded significantly faster in compatible condition ( $M = 819$  ms,  $SD = 237$  ms) than in the incompatible condition ( $M = 964$  ms,  $SD = 216$  ms),  $F(1,21) = 6.004, p = 0.024, \eta^2 = 0.219$ , indicating the existence of the mere choice effect. For participants in the control condition, they responded significantly faster in the compatible condition ( $M = 663$  ms,  $SD = 189$  ms) than in the incompatible condition ( $M = 766$  ms,  $SD = 251$  ms),  $F(1,20) = 8.787, p = 0.008, \eta^2 = 0.255$ , which also indicates the presence of the mere choice effect. These two findings demonstrated that participants preferred the self-chosen objects with positive descriptions (e.g., happiness, sunshine) and other-chosen objects with negative descriptions (e.g., death, war) over other-chosen objects with positive descriptions and self-chosen objects with negative descriptions. People implicitly preferred self-chosen objects to other-chosen objects even without owning those objects. The mere choice effect occurs, even without actually experiencing a choosing process.

For participants in the autonomy deprivation condition, their RT in the compatible condition ( $M = 897$  ms,  $SD = 259$  ms) was not significantly different from that in the incompatible condition ( $M = 915$  ms,  $SD = 198$  ms),  $F(1,20) = 2.111, p = 0.163, \eta^2 = 0.086$ , suggesting no mere choice effect. Here, the effect did not occur because there was no significant difference between compatible condition and the incompatible condition. The reference point is the participants' RT in the compatible condition, and we compare this reference RT with RT in the incompatible condition. The choice effect appeared in the autonomy fulfillment condition and the control condition, but not in the autonomy deprivation condition. The result is consistent with our Hypothesis 2.

We also used the difference response time (d-RT) as the indicator of the choice effect. The longer d-RT represents the larger choice effect. We conducted a one-way ANOVA of priming perceived autonomy on choice effect indicated by d-RT in the IAT task, controlling for gender and age. As hypothesized, the main effect of perceived autonomy was significant,  $F(2,61) = 4.550, p = 0.015, \eta^2 = 0.11$ .

We conducted the planned contrasts. By setting contrast coefficients, we can not only compare two means at once, but also combine multiple means from different levels to compute mean pair tests in these contrasts. Planned contrasts revealed that priming autonomy fulfillment ( $M_{d-RT} = 145$  ms,  $SD = 165$  ms) significantly increased the choice effect compared to priming autonomy deprivation ( $M_{d-RT} = 18$  ms,  $SD = 170$  ms),  $t(61) = 2.698, p = 0.009, d = 0.843$ , indicating a significantly larger choice effect in the autonomy fulfillment condition than in the autonomy deprivation condition. Compared with autonomy fulfillment, autonomy deprivation decreased the choice effect. This result is consistent with Hypothesis 3. Participants primed in the control group ( $M_{d-RT} = 103$  ms,  $SD = 123$  ms) did not have a significantly larger choice effect compared to participants primed with autonomy deprivation ( $M_{d-RT} = 18$  ms,  $SD = 170$  ms),  $t(61) = 1.785, p = 0.075, d = 0.564$ , the trend did not reach significance. Here, we did not find supporting evidence for Hypothesis 4, which proposed that autonomy deprivation would decrease the choice effect when compared with the control condition. Participants primed with autonomy fulfillment ( $M_{d-RT} = 145$  ms,  $SD = 165$  ms) did not show a significantly larger choice effect compared to participants in the control group ( $M_{d-RT} = 103$  ms,  $SD = 123$  ms),  $t(61) = 0.892, p = 0.376, d = 0.279$ ; the trend did not reach significance. Among the above effect sizes, the first one (i.e., the choice effect in the perceived autonomy fulfillment group compared to the autonomy deprivation group) is a fairly large effect. For Hypothesis 5, which stated that autonomy fulfillment would increase the choice effect when compared with the control condition, we neither found statistical support (see **Table 4**).

Similar to the data analysis procedure in Study 1, we examined the effect of autonomy priming on choice effect (indicated by the d-RT) in a regression model after controlling for gender and age. We entered gender in a first block/model, age in a second block/model, and the autonomy priming (score 3 represented autonomy fulfillment, score 2 represented control group, score 1 represented autonomy deprivation) as the independent variable in a third block/model. All variables were normalized as Z-scores for data analysis. The regression coefficients, standard error, 95% confidence interval [CI], the change in  $F$  statistic (including  $p$  value), and the coefficient of

**TABLE 4 |** The d-RT in IAT in autonomy priming conditions in Study 2.

	N	Mean d-RT (ms)	SD (ms)
Autonomy fulfillment	22	145	165
Control condition	21	103	123
Autonomy deprivation	21	18	170

determination change (delta  $R^2$ ) for each model are shown in **Table 5**. The results of regression analysis showed that after controlling for gender and age, the  $\beta$  of autonomy priming on choice effect represented by d-RT in the IAT task was .330, ( $SE = 0.116$ ,  $p < 0.01$ , 95% confidence interval [CI] = [0.097, 0.563]), which suggested a significant direct effect. **Table 5** shows that autonomy priming explained incremental variance of d-RT in IAT (10.8%),  $p = 0.006$ , suggesting that participants with autonomy fulfillment showed a larger choice effect and supporting Hypothesis 2 (see **Table 5**).

## GENERAL DISCUSSION

By using a modified illusory choice paradigm (adapted from Huang et al., 2009) to measure the mere choice effect the current research examined how autonomy would affect the choice effect even when the actual choice did not occur. Replicating the previous findings (Huang et al., 2009), the perceived choice, without involving a real choosing process, has also been found to enhance the attractiveness of an object in a autonomy-sufficient condition (Studies 1 and 2), which is termed as the choice effect. The sense of autonomy was measured not only as a trait by using questionnaire (Study 1), but also as a state by setting a priming task of recall writing (Study 2).

Our hypothesis that autonomy increases the choice effect was supported both when autonomy was measured as an individual-difference variable (Study 1) and when it was experimentally manipulated (Study 2). In Study 1, the level of trait autonomy was positively related with the choice effect. In Study 2, when state autonomy was enhanced, participants displayed a larger choice effect. When primed by the autonomy fulfillment recalling task, participants rated their chosen objects as more favorably than the objects chosen by others. That is to say, the choice effect occurred after one's state autonomy had been induced (see Study 2, in the autonomy fulfillment condition). Consistent with previous findings (Huang et al., 2009), we also found that the choice effect appeared without any autonomy related treatment (see Study 2, in the control condition). Interestingly, the choice effect disappeared when participants were primed with state

autonomy deprivation (see Study 2, in the autonomy deprivation condition). The two studies suggested that autonomy fulfillment is the premise of the choice effect, such that if people experience autonomy deprivation, their choice-induced preference would decrease or would even disappear.

Choice-induced preference has been a topic of longstanding interest in social psychology (Brehm, 1956; Steele, 1988; Lieberman et al., 2001; Gawronski et al., 2007; Huang et al., 2009; Egan et al., 2010). In the objects evaluation IAT task of the current study, the choice effect holds that people have a more positive attitude toward an object merely because they perceive choice of it. The perceived choice itself is sufficient to induce such effect. This evidence supports that choices influence preferences through a natural and automatic process, and the choice-induced preference is a byproduct of the choice (Leotti et al., 2010).

The occurrence of the mere choice effect is possibly related to the many aspects of the self-concept, such as self-serving or self-protecting biases (e.g., Sedikides and Strube, 1997), self-enhancement (e.g., Kurman, 2001), self-affirmation (e.g., Brown and Dutton, 1995), and self-verification (e.g., Chen et al., 2006). According to self-enhancement theory, people over evaluate self-related issues to maintain a positive self-image (e.g., Kurman, 2001). As "my choice" is a part of the self-concept, the positive words that describe the self-chosen objects represent the positive valence of self node (Greenwald et al., 2002). People experience the more positively self-image in choice effect, due to that "my choice" is given positive postchoice ratings. Faced with the need to maintain a positive self-image, participants would evaluate the "self-chosen" objects over the "non-self-chosen" objects, and that would result in the choice effect.

The most intriguing and main finding in the current study is that this choice effect was affected by the sense of autonomy. As showed in the results, a larger choice effect was elicited in the participants that experienced state autonomy fulfillment rather than in those with no priming treatment, but the trend did not reach a significant level. In addition, the choice effect disappeared when participants experienced autonomy deprivation. In the perceived choice-preference link, people's favorability on the self-chosen objects in the state autonomy fulfillment condition remains as high as in the control condition, whereas this

**TABLE 5 |** The hierarchical regression of predictors on choice effect in Study 2.

Predictor	The d-RT in IAT								
	95% CI (Confidence interval)	R	R <sup>2</sup>	ΔR <sup>2</sup>	F	p	β	SE	p
Model 1		0.136	0.019	0.019	1.174	0.283			
Gender	[-0.388, 0.115]						-0.136	0.126	0.283
Model 2		0.285	0.081	0.063	4.157	0.046			
Gender	[-0.323, 0.185]						-0.069	0.127	0.590
Age	[0.005, 0.513]						0.259*	0.127	0.046
Model 3		0.435	0.190	0.108**	8.025**	0.006			
Gender	[-0.332, 0.151]						-0.091	0.121	0.455
Age	[0.012, 0.493]						0.253*	0.110	0.040
Autonomy Priming	[0.097, 0.563]						0.330**	0.116	0.006

N = 64. \*p < 0.05; \*\*p < 0.01.

favorability would be weakened and would even disappeared if they experienced autonomy deprivation. The evidence that trait autonomy is positively correlated with the choice effect is consistent with this finding. In a word, autonomy moderated the relationship between the perceived choice and the induced preference.

The mere exercise of choice itself is assumed to provide a sense of autonomy (e.g., Iyengar and Lepper, 1999). People evaluate the chosen alternative as more desirable than the rejected alternative, in order to reassert their autonomy (Hammock and Brehm, 1966). The sense of autonomy, which has been treated as the expression or a result of actual choosing behavior, fulfills important psychological functions, such as enhancing happiness (Chekola, 2007; Demir et al., 2011) and increasing subjective well-being (Sheldon et al., 2004). Thus, people perceiving choice may experience the sense of autonomy, which will generate a positive feeling on the self-chosen objects, and that in turn will enhance the evaluation of the objects. Experiencing autonomy, which makes people feel free to act their own decisions, would improve individuals' feeling of the self-image. Besides, compared to those in who were merely aware of the choice, individuals who were primed with autonomy fulfillment displayed only a relatively, but not significantly larger trend in choice induced preference, because merely perceiving the choice could elicit the autonomy experience. The beforehand autonomy priming only contributes a little bit more on the basis of the autonomy experience induced by the choice.

One thing that needs to be pointed out in our objects evaluation IAT task is that participants were given the choosership and assigned to specific objects. That is, the perceived choice assigned to participants was not actually based on their free will. The autonomy induced by mere choice may be weaker than that induced by actual choice. Assuming that one's trait autonomy is stable, although the subsequent object evaluation task may elicit autonomy, this level could be canceled out by the previously primed autonomy deprivation. When the sense of autonomy has been deprived, one's intrinsic motivation and sense of control decreases (Zuckerman et al., 1978; Simon and McCarthy, 1982, Unpublished), and that generate a negative feeling on self-chosen objects, which in turn impairs the object-evaluation.

Taken together, this study found new evidence to explain the mechanism of the choice effect. That is, the sense of autonomy affects the choice effect, in other words, experiencing autonomy moderated the choice-preference link.

Although our study tapped on the mechanism underlying choice-induced preferences, the results still bear on some limitations. First, we did not directly test the positive and negative attitude on the self-chosen and the other-chosen objects separately. We just combined the attitudes to the positive-self-chosen objects with that to the negative-other-chosen ones, and the attitudes to the negative-self-chosen objects with that to the positive-other-chosen ones. In the future study, we could separate them and measure the attitude to one's positively or

negatively described objects separately by recording the real-time brain activities, which could also provide an implicit way measuring the attitude. Second, we did not record participants' explicit preference on objects, but only used the implicit attitudes as our indicator of preference. Although attitude on objects was evaluated in an implicit way by an IAT, which has the advantage of being immune to demanding characteristics and social desirability, it is necessary to replicate our findings using other explicit paradigm to confirm that the result can be generated in different kinds of situations. Third, we used the scenario in which participants were told which objects they have chosen, but not the actual choice action. A previous related study (Huang et al., 2009) using the same paradigm provided the evidence on the existence of a mere choice effect. Although this previous study has already verified that the virtual choice has the same efficacy as the actual choice. To be more carefully considered, we have to admit that the possible explanation is the vignette format. To fully verify the robust relationship of autonomy and choice effect, future research should investigate whether or not the actual choice actions provide a stronger relationship than that in the assigned choice. The relationship of autonomy deprivation and choice effect would be strengthened in the actual choice actions rather than assigned choice settings because of the more efforts in actions.

The findings of the current research reveal that autonomy affects the mere choice effect: (1) individual's autonomy trait is positively correlated with the mere choice effect; (2) the experience of autonomy deprivation decreases the mere choice effect, which results in that people do not value self-chosen objects more favorably than other-chosen objects anymore. Our research provides good insights in the relationship between autonomy and the mere choice effect, and contributes to the theoretical understanding of the mechanism in choice-induced preferences.

## AUTHOR CONTRIBUTIONS

LW proposed the main research idea. TT, ZS, and LW made the research design. TT ran the experiments. TT and ZS did the statistic analysis. ZS and LW wrote the manuscript.

## FUNDING

This work is supported by NSFC grant # 71021001, 91224008 and 91324201 and The Foundation of Beijing Key Laboratory of Behavior and Mental Health Grant #Z151100001615053.

## ACKNOWLEDGMENTS

We thank editor Aurora García-Gallego and two reviewers for their insightful comments on an earlier version of the manuscript.

## REFERENCES

- Ariely, D., and Norton, M. I. (2008). How actions create – not just reveal – preferences. *Trends Cogn. Sci.* 12, 13–16. doi: 10.1016/j.tics.2007.10.008
- Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *J. Abnorm. Psychol.* 52, 384–389. doi: 10.1037/h0041006
- Brown, J. D., and Dutton, K. A. (1995). Truth and consequences: the costs and benefits of accurate self-knowledge. *Pers. Soc. Psychol. Bull.* 21, 1288–1296. doi: 10.1177/01461672952112006
- Chekola, M. (2007). Happiness, rationality, autonomy and the good life. *J. Happiness Stud.* 8, 51–78. doi: 10.1007/s10902-006-9004-7
- Chen, S., English, T., and Peng, K. (2006). Self-verification and contextualized self-views. *Pers. Soc. Psychol. Bull.* 32, 930–942. doi: 10.1177/0146167206287539
- Coppin, G., Delplanque, S., Cayeux, I., Porcherot, C., and Sander, D. (2010). I'm no longer torn after choice: how explicit choices implicitly shape preferences of odors. *Psychol. Sci.* 21, 489–493. doi: 10.1177/0956797610364115
- deCharms, R. (1968). *Personal Causation*. San Diego, CA: Academic Press.
- Deci, E. L., Koestner, R., and Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychol. Bull.* 125, 627–668; discussion 692–700. doi: 10.1037/0033-2990.125.6.627
- Deci, E. L., and Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York, NY: Plenum Press.
- Deci, E. L., and Ryan, R. M. (2000). The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. *Psychol. Inq.* 11, 227–268. doi: 10.1207/S15327965PLI1104\_01
- Demir, M., Özdemir, M., and Marum, K. P. (2011). Perceived autonomy support, friendship maintenance, and happiness. *J. Psychol.* 145, 537–571. doi: 10.1080/00223980.2011.607866
- Egan, L. C., Bloom, P., and Santos, L. R. (2010). Choice-induced preferences in the absence of choice: evidence from a blind two choice paradigm with young children and capuchin monkeys. *J. Exp. Soc. Psychol.* 46, 204–207. doi: 10.1016/j.jesp.2009.08.014
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.
- Gawronski, B., Bodenhausen, G. V., and Becker, A. P. (2007). I like it, because I like myself: associative self-anchoring and post-decisional change of implicit evaluations. *J. Exp. Soc. Psychol.* 43, 221–232. doi: 10.1016/j.jesp.2006.04.001
- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., and Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychol. Rev.* 109, 3–25. doi: 10.1037/0033-295X.109.1.3
- Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480. doi: 10.1037/0022-3514.74.6.1464
- Gu, R., Ge, Y., Jiang, Y., and Luo, Y. (2010). Anxiety and outcome evaluation: the good, the bad and the ambiguous. *Biol. Psychol.* 85, 200–206. doi: 10.1016/j.biopsych.2010.07.001
- Hammock, T., and Brehm, J. W. (1966). The attractiveness of choice alternatives when freedom to choose is eliminated by a social agent. *J. Pers.* 34, 546–554. doi: 10.1111/j.1467-6494.1966.tb02370.x
- Huang, Y., Wang, L., and Shi, J. (2009). When do objects become more attractive? The individual and interactive effects of choice and ownership on object evaluation. *Pers. Soc. Psychol. Bull.* 35, 713–722. doi: 10.1177/0146167209333046
- Iyengar, S. S., and Lepper, M. R. (1999). Rethinking the value of choice: a cultural perspective on intrinsic motivation. *J. Pers. Soc. Psychol.* 76, 349–366. doi: 10.1037/0022-3514.76.3.349
- James, W. (1890). *The Principles of Psychology*. New York, NY: Henry Holt.
- Kurman, J. (2001). Self-enhancement: is it restricted to individualistic cultures? *Pers. Soc. Psychol. Bull.* 27, 1705–1716. doi: 10.1177/01461672012712013
- Lane, K. A., Banaji, M. R., Nosek, B. A., and Greenwald, A. G. (2007). "Understanding and using the Implicit Association Test: IV: procedures and validity," in *Implicit Measures of Attitudes: Procedures and Controversies*, eds B. Wittenbrink and N. Schwarz (New York, NY: Guilford Press), 59–102.
- Langer, E. J. (1975). The illusion of control. *J. Pers. Soc. Psychol.* 32, 311–328. doi: 10.1037/0022-3514.32.2.311
- Langer, E. J., and Rodin, J. (1976). The effects of choice and enhanced personal responsibility for the aged: a field experiment in an institutional setting. *J. Pers. Soc. Psychol.* 34, 191–198. doi: 10.1037/0022-3514.34.2.191
- Leotti, L. A., Iyengar, S. S., and Ochsner, K. N. (2010). Born to choose: the origins and value of the need for control. *Trends Cogn. Sci.* 14, 457–463. doi: 10.1016/j.tics.2010.08.001
- Lieberman, M. D., Ochsner, K. N., Gilbert, D. T., and Schacter, D. L. (2001). Do amnesics exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. *Psychol. Sci.* 12, 135–140. doi: 10.1111/1467-9280.00323
- Maison, D., Greenwald, A. G., and Bruin, R. H. (2004). Predictive validity of the implicit association test in studies of brands, consumer attitudes, and behavior. *J. Consum. Psychol.* 14, 405–415. doi: 10.1207/s15327663jcp1404\_9
- Moller, A. C., Deci, E. L., and Ryan, R. M. (2006). Choice and ego-depletion: the moderating role of autonomy. *Pers. Soc. Psychol. Bull.* 32, 1024–1036. doi: 10.1177/0146167206288008
- Olson, J. M., and Stone, J. (2005). "The influence of behavior on attitudes," in *The Handbook of Attitudes*, eds D. Albarracín, B. T. Johnson, and M. P. Zanna (Mahwah, NJ: Lawrence Erlbaum Press), 223–271.
- Patall, E. A., Cooper, H., and Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: a meta-analysis of research findings. *Psychol. Bull.* 134, 270–300. doi: 10.1037/0033-2990.134.2.270
- Ryan, R. M., and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* 55, 68–78. doi: 10.1037/0003-066X.55.1.68
- Sedikides, C., and Strube, M. J. (1997). Self-evaluation: to thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. *Adv. Exp. Soc. Psychol.* 29, 209–269. doi: 10.1016/S0065-2601(08)60018-0
- Sheldon, K. M. (1995). Creativity and self-determination in personality. *Creat. Res. J.* 8, 25–36. doi: 10.1207/s15326934crj0801\_3
- Sheldon, K. M., and Elliot, A. J. (1998). Not all personal goals are personal: comparing autonomous and controlled reasons for goals as predictors of effort and attainment. *Pers. Soc. Psychol. Bull.* 24, 546–557. doi: 10.1177/0146167298245010
- Sheldon, K. M., Ryan, R., and Reis, H. T. (1996). What makes for a good day? Competence and autonomy in the day and in the person. *Pers. Soc. Psychol. Bull.* 22, 1270–1279. doi: 10.1177/01461672962212007
- Sheldon, K. M., Ryan, R. M., Deci, E. L., and Kasser, T. (2004). The independent effects of goal contents and motives on well-being: it's both what you pursue and why you pursue it. *Pers. Soc. Psychol. Bull.* 30, 475–486. doi: 10.1177/0146167203261883
- Steele, C. M. (1988). The psychology of self-affirmation: sustaining the integrity of the self. *Adv. Exp. Soc. Psychol.* 21, 261–302. doi: 10.1016/S0065-2601(08)60229-4
- Swann, W. B., and Pittman, T. S. (1977). Initiating play activity of children: the moderating influence of verbal cues on intrinsic motivation. *Child Dev.* 48, 1128–1132. doi: 10.2307/1128374
- Van Overwalle, F., and Jordens, K. (2002). An adaptive connectionist model of cognitive dissonance. *Pers. Soc. Psychol. Rev.* 6, 204–231. doi: 10.1207/S15327957PSPR0603\_6
- van Prooijen, J. W. (2009). Procedural justice as autonomy regulation. *J. Pers. Soc. Psychol.* 96, 1166–1180. doi: 10.1037/a0014153
- Yang, Q., Tang, P., Gu, R., Luo, W., and Luo, Y. (2014). Implicit emotion regulation affects outcome evaluation. *Soc. Cogn. Affect. Neurosci.* 10, 824–831. doi: 10.1093/scan/nsu124
- Zuckerman, M., Porac, J., Lathin, D., Smith, R., and Deci, E. L. (1978). On the importance of self-determination for intrinsically motivated behavior. *Pers. Soc. Psychol. Bull.* 4, 443–446. doi: 10.1177/014616727800400317
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2016 Shang, Tao and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Fluid Intelligence and Cognitive Reflection in a Strategic Environment: Evidence from Dominance-Solvable Games

Nobuyuki Hanaki<sup>1</sup>, Nicolas Jacquemet<sup>2</sup>, Stéphane Luchini<sup>3</sup> and Adam Zylbersztein<sup>4\*</sup>

<sup>1</sup> Université Côte d'Azur, Centre National de la Recherche Scientifique, GREDEG, Valbonne, France, <sup>2</sup> CES, Paris School of Economics and University Paris 1 Panthéon-Sorbonne, Paris, France, <sup>3</sup> Aix-Marseille University (Aix-Marseille School of Economics), Centre National de la Recherche Scientifique and EHESS, Marseille, France, <sup>4</sup> Univ Lyon, Université Lumière Lyon 2, GATE L-SE UMR 5824, Ecully, France

## OPEN ACCESS

### Edited by:

Nikolaos Georgantzis,  
University of Reading, UK

### Reviewed by:

Adriana Breaban,  
Tilburg University, Netherlands  
Gerardo Sabater-Grande,  
Jaume I University, Spain

### \*Correspondence:

Adam Zylbersztein  
zylbersztein@gate.cnrs.fr

### Specialty section:

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
Frontiers in Psychology

Received: 31 May 2016

Accepted: 27 July 2016

Published: 10 August 2016

### Citation:

Hanaki N, Jacquemet N, Luchini S  
and Zylbersztein A (2016) Fluid  
Intelligence and Cognitive Reflection in  
a Strategic Environment: Evidence  
from Dominance-Solvable Games.

Front. Psychol. 7:1188.  
doi: 10.3389/fpsyg.2016.01188

Dominance solvability is one of the most straightforward solution concepts in game theory. It is based on two principles: dominance (according to which players always use their dominant strategy) and iterated dominance (according to which players always act as if others apply the principle of dominance). However, existing experimental evidence questions the empirical accuracy of dominance solvability. In this study, we study the relationships between the key facets of dominance solvability and two cognitive skills, cognitive reflection, and fluid intelligence. We provide evidence that the behaviors in accordance with dominance and one-step iterated dominance are both predicted by one's fluid intelligence rather than cognitive reflection. Individual cognitive skills, however, only explain a small fraction of the observed failure of dominance solvability. The accuracy of theoretical predictions on strategic decision making thus not only depends on individual cognitive characteristics, but also, perhaps more importantly, on the decision making environment itself.

**Keywords:** dominance solvability, cognitive skills, CRT, Raven's test, experiment

**JEL classification:** C72, D83

## 1. INTRODUCTION

Consider a game in which every decision maker is faced with a finite set of choices such that one specific choice always brings him higher monetary payoff than other choices, irrespective of the choices made by other players. In this situation, the individual choice boils down to going for either a higher or a lower monetary payoff. The straightforward response of a decision maker who cares about his monetary payoff is to disregard dominated actions—i.e., actions that may only deteriorate payoff relative to other actions. This dominance principle is the most basic solution concept of game theory (Camerer, 2003). It becomes very powerful when embedded in a strategic reasoning as a stepwise process. In each step, the dominance principle implies that dominated strategies should be eliminated from an agent's strategy space. In an important class of games—known as dominance-solvable games—this iterated elimination of dominated strategies leads to a unique solution.

Strikingly, the data collected from numerous experiments on dominance-solvable games raise important questions about the empirical accuracy of predictions derived from this principle. Subjects tend to display less strategic sophistication than is needed to justify many applications

of iterated dominance (and related refinements) to model human decision making in strategic environments (Crawford, 2004). The beauty contest game is one of the textbook examples of this issue<sup>1</sup>. A given set of players is asked to choose a number in the range [0, 100]. To win the game, a player should choose a number that is the closest to  $p = 2/3$  of the average of all chosen numbers. Any number above  $2/3 \times 100 \approx 66.7$  violates first-order dominance, because the average has to be lower than 100. Knowing this, players should all choose numbers no greater than 66.7, meaning that their average may not exceed  $2/3 \times 66.7 \approx 44.5$ . This reasoning lowers the target as the number of iterations increases, eventually leading to the unique Nash equilibrium in which all players choose 0. In many experimental studies of this game, the numbers chosen by players are used as a proxy of the depth of iterated reasoning.<sup>2</sup> A well replicated stylized fact is to observe 1/3 of subjects choosing a number higher than 67, and at least 1/3—a number between 44 and 67.

This paper focuses on one of the earliest and simplest example of such an empirical inaccuracy of dominance solvability, adapted from a  $2 \times 2$  game discussed in Rosenthal (1981) and first brought to the laboratory by Beard and Beil (1994)<sup>3</sup>. The normal-form representation of this game is given in **Table 1**. With  $L < S < H$ ,  $m < h$ , and  $s < h$ , the game is one-step dominance solvable: the elimination of player B's weakly dominated strategy  $l$  immediately leads to the Pareto-Nash equilibrium ( $R, r$ )<sup>4</sup>.

In line with observed behavior in other dominance solvable games, numerous studies (summarized in **Table 2**) find frequent failures to achieve the Pareto-Nash equilibrium. In spite of variations in the design (described in the table), deviations from the standard theoretical predictions are systematic and sizable. First, dominance is frequently violated by player Bs. Depending on the exact experimental setup, up to 27% column players choose a strictly dominated action. Second, player As violate

**TABLE 1 | Generic form of the normal representation of Rosenthal (1981) dominance solvable game.**

		Player B	
		<i>l</i>	<i>r</i>
Player A	<i>L</i>	(S; s)	(S; s)
	<i>R</i>	(L; m)	(H; h)

<sup>1</sup>This class of games has been first introduced by Moulin (1986) as the  $p$ -beauty contest games, where  $p$  (often equal 2/3) stands for the target fraction of all numbers' average.

<sup>2</sup>See Nagel (1995) and Ho et al. (1998) for early evidence from the laboratory, Costa-Gomes and Crawford (2006) for a laboratory experiment supporting a behavioral model of bounded rationality, and Bosch-Domenech et al. (2002) for related evidence from the field.

<sup>3</sup>Both Camerer (2003) and Crawford (2004) consider this game as a basic example of a dominance-solvable game, and a glaring case of a mismatch between theoretical predictions and actual behavior.

<sup>4</sup>If the game is played sequentially (so that player A moves first), the same solution can be obtained through backward induction. Note that if  $s > h$ , the solution does not change (since  $l$  remains player B's weakly dominated strategy), but the outcomes are no longer Pareto-rankable. Beard and Beil (1994), Schotter et al. (1994), and Goeree and Holt (2001) find that this environment also generates important violations of standard theoretical predictions.

iterated dominance, even in those cases in which player Bs commonly obey dominance. As an example, while only 6% of player Bs violate dominance in Jacquemet and Zylbersztein (2014)-ET2 and BT2, 26% of row players still contradict the predictions of dominance solvability by choosing  $L$  (and this figure may even attain 86% in other instances, see Beard, Beil – Tr. 5 in **Table 2**). As shown in the three middle columns of the table, both the absolute and the relative size of the stakes vary a great deal from one study to the other. Several lessons emerge from this accumulated evidence. First, both players react to their own monetary incentives. Second, in some cases player As also adjust their behavior to player Bs' incentives. Finally, as shown by Jacquemet and Zylbersztein (2014), players' inefficient behavior does not fade away with repetition and cannot be explained by inequality aversion (as framed by Fehr and Schmidt, 1999).

The aim of the present paper is to explore whether this empirical puzzle is related to players' cognitive skills. In this sense, our investigation belongs to a recent and growing body of experimental studies in both psychology and economics which investigate the relationship between strategic behavior and cognitive skills<sup>5</sup>. The main conclusion that can be drawn from these studies is that high cognitive skills predict strategic sophistication and efficient decision making. First, people with high cognitive skills make more accurate predictions about other people's intentions. Recent evidence from psychological research reveals the relationship between cognitive skills and the theory of mind. Using the "Reading the Mind in the Eyes" test (RMET, Baron-Cohen et al., 2001) to measure one's theory of mind, Ibanez et al. (2013) find that people with higher cognitive skills are better at inferring the internal emotional states of others<sup>6</sup>. Relatedly, the results of a neuroeconomic experiment on the  $p$ -beauty contest game by Coricelli and Nagel (2009) suggest that strategic thinking about other players' thoughts and behavior is implemented by medial prefrontal cortex (mPFC) – one of the brain areas commonly associated with theory of mind<sup>7</sup>. An economic experiment by Carpenter et al. (2013) also shows that people with higher cognitive ability make more accurate predictions of others' choices in a 20-player beauty contest game. Second, people with higher cognitive skills apply more sophisticated reasoning and are more apt in strategic adaptation. Burks et al. (2009) report that subjects with higher cognitive skills more accurately predict others' behavior in a sequential prisoners' dilemma game, and adapt their own behavior more strongly. In the context of the  $p$ -beauty contest game, subjects with higher cognitive skills are not only found to carry out more steps of reasoning on the equilibrium path (Burnham et al., 2009; Brañas-Garza et al., 2012), but also to adapt their behavior to their opponents' cognitive skills (Gill and Prowse, forthcoming) as well as to their beliefs about their opponents' cognitive skills (Fehr and

<sup>5</sup>Cognitive skills are often measured using (amongst others) the Cognitive Reflection Test (CRT, Frederick, 2005), the Raven's progressive matrices test (Raven, 2008), or both (like in this study). The details of these two measures are presented in Section 2.

<sup>6</sup>RMET consists of a series of photos of the area of the face involving the eyes. Subjects are asked to choose one of the four words that best describes what the person in the photo is thinking or feeling.

<sup>7</sup>See Hampton et al. (2008) for related evidence.

**TABLE 2 | Overview of existing experimental evidence.**

Experiment	Form	Payoff			Outcomes (%)				
		(L)	(R, r)	(R, I)	L	R, r	R, I	r R	r
Beard, Beil-Tr.1	Seq	(9.75; 3.0)	(10; 5.0)	(3; 4.75)	66	29	6	83	—
Beard, Beil-Tr.2	Seq	(9.00; 3.0)	(10; 5.0)	(3; 4.75)	65	35	0	100	—
Beard, Beil-Tr.3	Seq	(7.00; 3.0)	(10; 5.0)	(3; 4.75)	20	80	0	100	—
Beard, Beil-Tr.4	Seq	(9.75; 3.0)	(10; 5.0)	(3; 3.00)	47	53	0	100	—
Beard, Beil-Tr.5	Seq	(9.75; 6.0)	(10; 5.0)	(3; 3.00)	86	14	0	100	—
Beard, Beil-Tr.7	Seq	(58.50; 18.0)	(18.0; 28.50)	(60; 30.0)	67	33	0	100	—
Beard et al.–Tr.1	Seq	(1450; 450)	(1500; 750)	(450; 700)	79	18	3	83	—
Beard et al.–Tr.2	Seq	(1050; 450)	(1500; 750)	(450; 700)	50	32	18	64	—
Goeree, Holt–Tr.1	Ext	(80; 50)	(90; 70)	(20; 10)	16	84	0	100	—
Goeree, Holt–Tr.2	Ext	(80; 50)	(90; 70)	(20; 68)	52	36	12	75	—
Goeree, Holt–Tr.3	Ext	(400; 250)	(450; 350)	(100; 348)	80	16	4	80	—
Cooper, Van Huyck–Tr.9	Str	(4; 1)	(6; 5)	(2; 4)	27	—	—	—	86
Cooper, Van Huyck–Tr.9	Ext	(4; 1)	(6; 5)	(2; 4)	21	—	—	—	84
JZ, 2014–BT1	Str	(9.75; 3.0)	(3.0; 4.75)	(10; 5.0)	51	41	8	84	81
JZ, 2014–ET1	Str	(9.75; 5.0)	(5.0; 9.75)	(10; 10.0)	54	33	13	72	73
JZ, 2014–ET3	Str	(9.75; 5.5)	(5.5; 8.50)	(10; 10.0)	39	48	13	79	76
JZ, 2014–ET4	Str	(8.50; 5.5)	(5.5; 8.50)	(10; 10.0)	25	61	14	82	82
JZ, 2014–ET2	Str	(8.50; 8.5)	(6.5; 8.50)	(10; 10.0)	26	70	4	94	94
JZ, 2014–BT2	Str	(8.50; 7.0)	(6.5; 7.00)	(10; 8.5)	26	70	4	94	94

For each implementation in row, the first column describes the actual design of the experiment: simultaneous-move strategic-form game (Str), simultaneous-move extensive-form game (Ext), sequential-move game (Seq). The monetary payoffs of each outcome, displayed in columns 2–4, are in USD in Beard and Beil (1994) and Cooper and Van Huyck (2003), in cents of USD in Goeree and Holt (2001), in Yens in Beard et al. (2001), and in Euros in Jacquemet and Zylbersztein (2014). The game is repeated ten times in changing pairs in Jacquemet and Zylbersztein (2014), and one-shot in all other instances.

Huck, 2015). Third, cognitive skills may be associated with the economic efficiency of outcomes of both individual and group activities. Corgnet et al. (2015b) find that higher cognitive skills predict better performance and less shirking in an experimental labor task (summing up tables of 36 numbers without using a pen). Jones (2008), Al-Ubaydli et al. (in press), and Proto et al. (2014) report that groups with higher cognitive skills attain higher cooperation rates in repeated prisoner's dilemma games. On the other hand, Al-Ubaydli et al. (2013) do not find a relationship between group members' average cognitive skills and the efficiency of outcomes in a stag hunt coordination<sup>8</sup>.

Our contribution is two-fold. First, we provide new evidence on the relationship between strategic behavior and cognitive skills. We show that systematic mismatches between theoretical predictions and actual behavior in a classic  $2 \times 2$  dominance-solvable game have cognitive underpinnings. Subjects with higher cognitive skills are found to be more likely to play dominant strategy and to best respond to other's strategy. Furthermore, cognitive skills predict strategic sophistication: only those players with sufficiently high cognitive ability are found to display sensitivity to the presence of uncertainty about others' behavior. Our second contribution lies in

experimental methodology. We extend the recent body of laboratory experiments comparing the performance of different measures of cognitive skills in predicting economic behavior. Notwithstanding the previous results (see e.g., Brañas-Garza et al., 2012; Corgnet et al., 2015a), we report that the Raven's test score is a more general predictor of strategic behavior than the Cognitive Reflection Test score.

## 2. EXPERIMENTAL DESIGN

Our experiment is based on a  $2 \times 2$  factorial design that varies the payoff matrix and the nature of player B. Each of the four resulting experimental treatments is implemented through a between-subject procedure—each subject participates in only one experimental condition. This data come from a large dataset, part of which has been previously used by Hanaki et al. (2016). The main focus of that study is player A's behavior under strategic uncertainty and its relation to monetary incentives and fluid intelligence. Certain elements of their design (such as the use of Human and Robot conditions and interest in players' cognitive skills) inevitably needed to be adopted in the present study in order to address a much more general question of the empirical validity of the solution concept of dominance solvability. More precisely, we are interested in both players' behavior (so as to

<sup>8</sup>Al-Ubaydli et al. (2013, in press) also report that individual cognitive skills do not predict individual willingness to reach efficient outcomes in these two game.

measure the use of dominance by player Bs and the use of iterated dominance by player As under different information structures). We also make a methodological contribution, since in this paper we associate players' behavior with multiple facets of cognitive skills: fluid intelligence (measured by Raven's test) and cognitive reflection (measured by CRT).

Our first treatment variable is the size of the stakes, as represented by Game 1 and Game 2 in **Table 3**. Although they have the same strategic properties, these two game matrices differ in terms of the saliency of monetary incentives to use (iterated) dominance. In Game 2, player As may earn a surplus of only 0.25 when moving from  $L$  to  $(R, r)$  (with payoff going from 9.75 to 10), while ending up in  $(R, l)$  is relatively costly (yielding only 3). In Game 1, the potential gains and losses from action  $R$  relative to  $L$  are more balanced: the gain from moving from  $L$  to  $(R, r)$  increases to 1.5 (with payoff moving from 8.5 to 10), while the outcome  $(R, l)$  becomes less costly (now yielding 6.5). The incentives of player Bs, in turn, go in the opposite direction: the gain from using the dominant strategy  $r$  (and conditional on player As' choice  $R$ ) is lower in Game 1 [with payoff increasing from 4.75 to 5 between  $(R, l)$  and  $(R, r)$ ] than in Game 2 (where payoff increases from 8.5 to 10). In line with Jacquemet and Zylbersztein (2014) and Hanaki et al. (2016) (who report that both players only react to their own monetary incentives) and as discussed in Section 3.1, each of these games generates sizable yet diverse empirical violations of dominance solvability. These two games together thus provide a wide range of monetary incentives to use dominance solvability within a common strategic environment<sup>9</sup>.

Our second treatment variable is related to the nature of player B (the column player) who may be represented either by a human subject (Human condition) or a pre-programmed computer (Robot condition). The Human condition enables us to capture two cardinal breaches of dominance solvability: the failure to use the dominant strategy (player Bs' behavior) and the failure to best respond to others' dominant actions (player

As' behavior). However, the latter behavior occurs under strategic uncertainty and thus might stem from two distinct sources: bounded rationality and rational behavior under uncertainty. More precisely, player As may simply have a limited capability of best responding to dominant strategy, but may also intentionally refrain from best responding when in doubt about player Bs' use of dominant strategy. To separate these two effects, we introduce the Robot condition in which a human subject acting as player A interacts with a computerized player B who is pre-programmed to always choose  $r$ . We clearly inform the subjects in the Robot condition that they are interacting with a pre-programmed computer: "**the computer chooses  $r$  at each round, without exception**" (bold in the original instruction sheet). This is the only difference in the rules and procedures between Human and Robot conditions<sup>10</sup>. Thus, the key property of the Robot condition as compared to the Human condition is neutralizing strategic uncertainty player As face, while maintaining space for boundedly rational behavior.

The design of the experiment is otherwise the same in all four experimental conditions. We explore whether behavior is sensitive to learning by considering ten uniform, one-shot interactions. In order to homogenize incentives across rounds, the following rules are implemented: all games are played in strict anonymity, roles are fixed, and subjects' payoffs are computed based one randomly drawn round. In the Human condition, players are matched into pairs using a perfect stranger, round-robin scheme, which guarantees that subjects are involved in a series of one-shot interactions despite the repetition of the game<sup>11</sup>.

Our control variables also include two measures of cognitive skills. Both of them are introduced as part of a post-experimental supplementary task. Subjects' participation is rewarded with extra five Euros; otherwise, their answers are not incentivized<sup>12</sup>. The supplementary task starts with a debriefing question, where subjects are asked to "report any information they find relevant about how their decisions has been made." Then, we implement the following measures of cognitive skills.

The first task is the standard Cognitive Reflection Test based on Frederick (2005) which "measures cognitive reflectiveness or impulsiveness, respondents' automatic response versus more elaborate and deliberative thought" (Brañas-Garza et al., 2012, p. 255). It contains three questions:

1. *A notebook and a pencil cost 1.10 Euros in total. The notebook costs 1 Euro more than the pencil. How much does the pencil cost?*
2. *If it takes 5 machines 5 min to make 5 widgets, how long would it take 100 machines to make 100 widgets?*

<sup>10</sup>An English translation of the original instructions in French is provided as supplementary material.

<sup>11</sup>See Jacquemet and Zylbersztein (2013) for a detailed motivation and description of this design.

<sup>12</sup>Absence of monetary incentives for providing corrects answers is a standard procedure for both CRT and Raven's tests. Recent evidence on both tests suggests that monetary incentives do not *per se* affect people's performance. See Brañas Garza et al. (2015) for a metastudy on the determinants of CRT scores and Eckart et al. (2012) and Dessi and Rustichini (2015) for experimental evidence on the role of monetary incentives in Raven's test.

**TABLE 3 |** The experimental games.

		<b>B</b>	
		<i>l</i>	<i>r</i>
<b>A</b>	<i>L</i>	(8.50 ; 3.00)	(8.50 ; 3.00)
	<i>R</i>	(6.50 ; 4.75)	(10.00 ; 5.00)
		<b>B</b>	
<b>A</b>	<i>l</i>	(9.75 ; 8.50)	(9.75 ; 8.50)
	<i>r</i>	(3.00 ; 8.50)	(10.00 ; 10.00)

<sup>9</sup>Herein, we restrict our design to these two game matrices and do not seek to further investigate the effects of monetary incentives on both players' behavior. These effects are analyzed in detail in Jacquemet and Zylbersztein (2014) and Hanaki et al. (2016).

3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Subjects are informed that this set of three questions should be answered within 30 s (although we allow them to provide answers even after this time has elapsed). In this way, subjects can be classified according to their overall score (that is, the total number of correct answers) which can range from 0 to 3.

The second task is Raven's progressive matrix test (often called Raven's test), a picture based, non-verbal measure of fluid intelligence, that is "the capacity to think logically, analyze and solve novel problems, independent of background knowledge" (Mullainathan and Shafir, 2013, p. 48). It is widely used by, e.g., psychologists, educators, and the military (Raven, 2000). It consists of a series of tasks to be solved within a fixed amount of time. In each task, a subject should pick a single element (among eight options) that best fits a set of eight pictures. The level of difficulty increases from one question to the other<sup>13</sup>. In our experiment, each participant is given a series of 16 tasks to be solved within 10 min. Individual scores in Raven's test are computed as the number of correct answers to the 16 items of the test.

## 2.1. Experimental Procedures

For each game matrix, we run three Human sessions (involving 20 subjects per session: 10 player As interacting with 10 player Bs), and two Robot sessions (involving 20 player As per session interacting with automated player Bs). Subjects are given a fixed fee equal to five euros to compensate participation to the experiment.

Upon arrival to the laboratory, participants are randomly assigned to their computers and asked to fill in a short administrative questionnaire containing basic questions about their age, gender, education, etc. Experimental instructions are then read aloud: subjects are informed that they will play multiple rounds of the same game, each round with a different partner, and that their own role will remain unchanged throughout the experiment. Finally, subjects are asked to answer a short comprehension quiz. Once the quiz and any questions from participants are answered, the experiment begins. After each of the ten rounds of the game, subjects are only informed of their own payoffs. Information about past choices and payoffs is updated after each round and displayed at the bottom of the screen. Take-home earnings correspond to the outcome of a single round that is randomly drawn at the end of each experimental session.

In addition, the experimental game is followed by supplementary tasks. An additional five euros fee is paid to each subject for completing this part. Immediately after the end of the experimental game, participants are provided with a brief round-by-round summary of their decisions and outcomes, and are asked to provide in a blank space on their computer screens any relevant comments in particular about what might have affected their decisions during the experiment. Subjects are also

asked to solve the CRT test and a reduced-form Raven's test described above.

All the sessions were conducted in February and March 2014. Out of the 200 participants (94 males), 155 were students with various fields of specialization. The majority of subjects (65%) had already taken part in economic experiments. Participants' average age was 25.6 (st. dev. is 7.5). All sessions took place at the *Laboratoire d'Economie Experimentale de Paris* (LEEP) at Paris School of Economics. Subjects were recruited via an online registration system based on ORSEE (Greiner, 2015) and the experiment was computerized through software developed under REGATE (Zeiliger, 2000) and z-Tree (Fischbacher, 2007). Sessions lasted about 45–60 min, with an average payoff of roughly 18.83 euros (including a five euros show-up fee and five euros for completing the post-experimental tasks).

## 3. RESULTS

Our main experimental results can be summarized as follows. First, in line with the existing literature, we observe systematic and sizable deviations from standard predictions based on the principle of dominance solvability. This phenomenon persists across game matrices and despite repetition. Second, we associate strategic behavior with cognitive skills. We find that Raven's test score is a more reliable predictor of strategic behavior than CRT score: whenever the latter predicts behavior, the former does too, but not vice versa. Subjects with higher Raven's test scores are more likely to use the dominant strategy and to best respond to other player's dominant strategy. Unlike those with low Raven's test score, they also react to the presence of strategic uncertainty.

### 3.1. Aggregate Behavior in Experimental Games

**Table 4** outlines the main patterns of behavior in our experimental games. The statistical significance of the changes observed in this table is tested by Models 1–3 in **Table 5**. We first focus on the aggregate frequency of Pareto-Nash equilibrium ( $R, r$ ) – the sole outcome that survives the iterated elimination of (weakly) dominated strategies—found in the Human condition. In both games, we observe substantial deviations from the predictions of this solution concept: overall, players attain the  $(R, r)$  outcome 58% of times in Game 1 and 43% in Game 2 (Model 1,  $H_0 : \beta_1 = 0, p = 0.318$ ). We also observe that efficiency increases over time: in both games, we observe the lowest frequency of  $(R, r)$  in the initial round (0.333 in Game 1 and 0.200 in Game 2), whereas the highest frequency of  $(R, r)$  occurs in the final round (0.700 in Game 1 and 0.533 in Game 2).

To further explore the roots of these deviations, we turn to the aggregate patterns of both players' behavior in Human and Robot conditions. We focus on three behavioral dimensions of dominance solvability: the use of dominant strategy (captured by player Bs' behavior in the Human condition) and the ability to best respond to other player's dominant action with and without bearing the uncertainty about the latter (which is captured by player As' behavior in the Human and Robot conditions, respectively).

<sup>13</sup>See Raven (2008) for an overview.

**TABLE 4 | Aggregate results.**

	Round										Overall
	1	2	3	4	5	6	7	8	9	10	
<b>Pr (R,r) in the Human condition</b>											
Game 1	0.333	0.600	0.667	0.700	0.567	0.600	0.433	0.633	0.567	0.700	0.580
Game 2	0.200	0.333	0.400	0.400	0.433	0.500	0.500	0.433	0.467	0.533	0.420
<b>Pr (r) by player B in the Human condition</b>											
Game 1	0.767	0.800	0.867	0.900	0.800	0.800	0.700	0.833	0.867	0.800	0.813
Game 2	0.833	0.933	0.900	0.933	1.000	0.933	0.933	0.900	0.900	0.933	0.920
<b>Pr (R) by player A in the Human condition</b>											
Game 1	0.500	0.733	0.700	0.767	0.767	0.800	0.700	0.767	0.700	0.867	0.730
Game 2	0.300	0.333	0.400	0.400	0.433	0.533	0.533	0.500	0.500	0.533	0.447
<b>Pr (R) by player A in the Robot condition</b>											
Game 1	0.700	0.750	0.750	0.725	0.800	0.800	0.800	0.825	0.800	0.775	0.773
Game 2	0.500	0.575	0.725	0.575	0.800	0.700	0.700	0.775	0.775	0.775	0.690

Columns 1–10 summarize the frequencies of outcomes (defined in rows) as % of all outcomes observed in each round of a given experimental treatment. The last column provides overall results.

**TABLE 5 | Aggregate results: statistical support.**

	Model 1	Model 2	Model 3
	Pr (R, r)	Pr (r)	Pr (R)
Constant ( $\beta_0$ )	0.580*** (0.144)	0.813*** (0.032)	0.730*** (0.084)
1[Game 2] ( $\beta_1$ )	-0.160 (0.103)	0.107* (0.044)	-0.283** (0.140)
1[Robot] ( $\beta_2$ )		0.043 (0.103)	
1[Robot] × 1[Game 2] ( $\beta_3$ )		0.201 (0.161)	
N	600	600	1400
R <sup>2</sup>	0.026	0.025	0.066

Estimates of linear probability models on outcome (R, r) (Model 1), decision r by player B (Model 2) and decision R by player A (Model 3). Standard errors (in parentheses) are clustered at the session level in Human treatments (three clusters per game matrix, six in total) and individual level in the Robot condition (40 clusters per game matrix, 80 in total) and computed using the delete-one jackknife procedure. All models contain a dummy variable set to 1 for game matrix 2 (and 0 for game matrix 1). In Model 3, we also introduce an additional dummy variable set to 1 for Robot condition (and 0 for Human condition) and well as the interaction between these two variables. \*\*/\*\*\* indicate significance at the 10/5/1% level.

Inefficiency is caused by both players, although their roles differ from one game to another: the scope of inefficient behavior is similar for both players in Game 1, and highly asymmetric in Game 2. Overall, player As select action R with probability 0.730 in Game 1 and 0.447 in Game 2 (Model 3,  $H_0 : \beta_1 = 0, p = 0.047$ ). However, player As' behavior happens to be misaligned with player Bs' actual decisions which follow the opposite trend:

the total frequency of action r increases from 0.813 in Game 1 to 0.920 in Game 2 (Model 2,  $H_0 : \beta_1 = 0, p = 0.060$ ). Importantly, the data from Robot sessions suggest that the uncertainty about player Bs' behavior is not the only driver of player As' choices. Player As frequently and systematically fail to best respond to player Bs' dominant action even when the latter comes with certainty in the Robot condition, although their willingness to select action R increases in both games as compared to the Human condition (to 0.773 in Game 1 and 0.690 in Game 2)<sup>14</sup>. The fact that inefficient actions from player As prevail in the absence of strategic uncertainty may suggest that at least some of them are boundedly rational decision makers.

In the next section, we analyze how these three behavioral components of dominance solvability vary as a function of players' cognitive skills.

### 3.2. Cognitive Skills and Strategic Behavior

The average score in Raven's test (CRT) is 8.679 out of 16 with SD 3.117 (0.479 out of 3 with SD 0.852). Our experimental sample is properly randomized across treatments regarding both measures. We do not reject the null hypothesis that Raven's test scores have the same distributions in all treatments ( $p = 0.275$ , Kruskal-Wallis test). A Kruskal-Wallis test applied to the CRT scores leads to the same conclusion ( $p = 0.502$ ).

We also replicate several results from previous studies combining Raven's test and CRT regarding the relationship between both scores as well as gender differences (Brañas-Garza

<sup>14</sup>Model 3 suggests that these two proportions are not significantly different: testing  $H_0 : \beta_1 + \beta_3 = 0$  yields  $p = 0.303$ . The increase in the proportion of decisions R between Human and Robot conditions is insignificant for Game 1 ( $H_0 : \beta_2 = 0, p = 0.679$ ) and significant for Game 2 ( $H_0 : \beta_2 + \beta_3 = 0, p = 0.054$ ).

et al., 2012; Corgnet et al., 2015a). There is a moderate, yet highly significant correlation between Raven and CRT scores (Spearman's  $\rho = 0.306$ ,  $p < 0.001$ ) which suggests that they may have a common source, but do not capture the same cognitive skills. Furthermore, the average score of males is significantly higher than the average score of females (Raven's test: 9.382 with SD 0.341 vs. 8.014 with SD 0.384,  $p = 0.009$ ; CRT: 0.676 with SD 0.111 vs. 0.291 with SD 0.087,  $p = 0.007$ ; two-sided  $t$ -tests)<sup>15</sup>.

We also observe that many subjects (70%) of our 200 participants fail to provide at least one correct answer in our standard CRT. 16% provide exactly one, 8% – two, and 6% – three correct answers. This stands in line with Brañas-Garza et al. (2012) who report the respective frequencies of 67, 23, 9, and 1% for a similar sample size ( $N = 191$ ), and echoes the scores in the least performant sample reported in a seminal study by Frederick (2005): out of 138 students of the University of Toledo, 64% provide no correct answer, 21% provide one, 10% provide two, and 5% provide three corrects answers.

### 3.2.1. Cognitive Predictors of Strategic Behavior: Aggregate Results

In this part, we study the cognitive correlates of strategic behavior. Figures 1, 2 present the aggregate evolution of behavior as a function of cognitive skills, measured either by CRT score or by Raven's test score across roles (player A or player B) and experimental conditions (Human or Robot).

In Figure 1, the sample is divided into two subsamples: subjects who provided at least one correct answer to CRT (referred to as  $CRT > 0$ ) and those who did not (referred to as  $CRT = 0$ ). The aggregate patterns of behavior weakly differ between the two subsamples. Bootstrap proportion tests fail to reject the null hypothesis that the overall proportions of decision  $R$  are the same for both CRT categories in the Human condition ( $p = 0.126$ ) and in the Robot condition ( $p = 0.235$ )<sup>16</sup>. The aggregate proportions of decision  $r$ , in turn, are found to be statistically different ( $p = 0.037$ ), subjects with a CRT score zero being less likely to play  $r$  than subjects who gave at least one correct answer.

In Figure 2, we split our sample into three subsamples based on Raven's test score (1st tertile: less than 8 correct answers, 2nd tertile: between 8 and 10 correct answers, 3rd tertile: more than 10 correct answers). Although, bootstrap proportion tests suggest that player As' behavior in the Human condition does not vary significantly between these three subsamples (1st tertile vs. 2nd tertile:  $p = 0.255$ , 2nd vs. 3rd:  $p = 0.580$ , 1st vs. 3rd:  $p = 0.565$ ), significant differences arise for both player As in the Robot condition ( $p = 0.001$ ,  $p = 0.735$ ,  $p < 0.001$ , respectively) and for player Bs ( $p = 0.064$ ,  $p = 0.057$ ,  $p < 0.001$ , respectively). Raven's test score seems to have a more systematic association

<sup>15</sup>See also Frederick (2005) and Bosch-Domènec et al. (2014) for related evidence.

<sup>16</sup>We test the difference in proportion of a given outcome between two experimental conditions by carrying out a bootstrap proportion test that accounts for within-subject correlation, i.e., the fact that the same individual takes 10 decisions. The procedure consists of bootstrapping subjects and their corresponding decisions over all 10 rounds instead of bootstrapping decisions as independent observations (see e.g., Jacquemet et al., 2013, for a detailed description of the procedure).

with players' behavior than CRT score, although both measures fail to predict behavior under strategic uncertainty.

### 3.2.2. Cognitive Skills and Dominance Solvability: Regression Analysis

In what follows, we provide further econometric insights into these preliminary results. Following Brañas-Garza et al. (2012); Corgnet et al. (2015a), we use three individual characteristics discussed in the previous section – gender, Raven's test score and CRT score (kept as a dummy variable with value 1 if the subject gave at least one correct answer at the CRT test and 0 otherwise) – to explain behavior in our experimental games<sup>17</sup>. The econometric specification is based on the linear probability model and the estimation procedure is outlined in Jacquemet and Zylbersztein (2014). We also control for payoff scheme and repetition effects by including game matrix and round dummies. We consider three different outcome variables: player As' behavior in the Human and the Robot treatment, and player Bs' behavior in the Human treatment. Given the correlation between CRT and Raven's test scores, including both variables in the model might result in multicollinearity and lead to the under-rejection of the nullity of respective coefficients. For each outcome, we first include these two measures separately in Models 1 and 2, while Model 3 includes both variables. This evidence is summarized in Table 6.

We first turn to player Bs' behavior. Models 1 and 2 suggest that both the coefficient of  $CRT > 0$  dummy and the coefficient of Raven's test score are positive and significant ( $p = 0.067$  for  $CRT > 0$  and  $p = 0.015$  for Raven). In Model 3, the coefficient of Raven's test score remains highly significant ( $p = 0.014$ ), while the coefficient of CRT becomes insignificant ( $p = 0.253$ ). Their joint significance ( $p = 0.034$ ) implies that cognitive skills predict the use of dominant strategy.

We now turn to player As' behavior in the Human condition. Notwithstanding the previous set of results, cognitive skills are not found to explain player As' choices. The coefficient of  $CRT > 0$  dummy is insignificant ( $p = 0.226$ ) in Model 1, and so is the coefficient of Raven's test score ( $p = 0.633$ ) in Model 2. If we account for both, Model 3 reveals that the coefficients of both scores are neither individually ( $p = 0.226$  for  $CRT > 0$  and  $p = 0.550$  for Raven's test score) nor jointly significant ( $p = 0.503$ ). Finally, the behavior of player As in the Robot condition is only predicted by Raven's test score: unlike  $CRT > 0$  dummy, its coefficient remains positive and highly significant across models ( $p \leq 0.001$ ). Unsurprisingly, the joint insignificance of both coefficients in Model 3 is also rejected ( $p = 0.003$ ).

Altogether, the results presented in Table 6 suggest that cognitive skills predict certain components of strategic behavior: the use of dominant strategy (reflected in player Bs' behavior), as well as the ability to best respond to other player's dominant strategy (reflected in player As' behavior in the Robot condition). Moreover, in both cases Raven's test score is a more reliable predictor of behavior than CRT score. However, we also observe

<sup>17</sup>Given that most CRT scores in our sample are null and the higher the score, the less frequent it gets, dichotomizing the CRT score variable limits the impact of the outliers on the overall results.

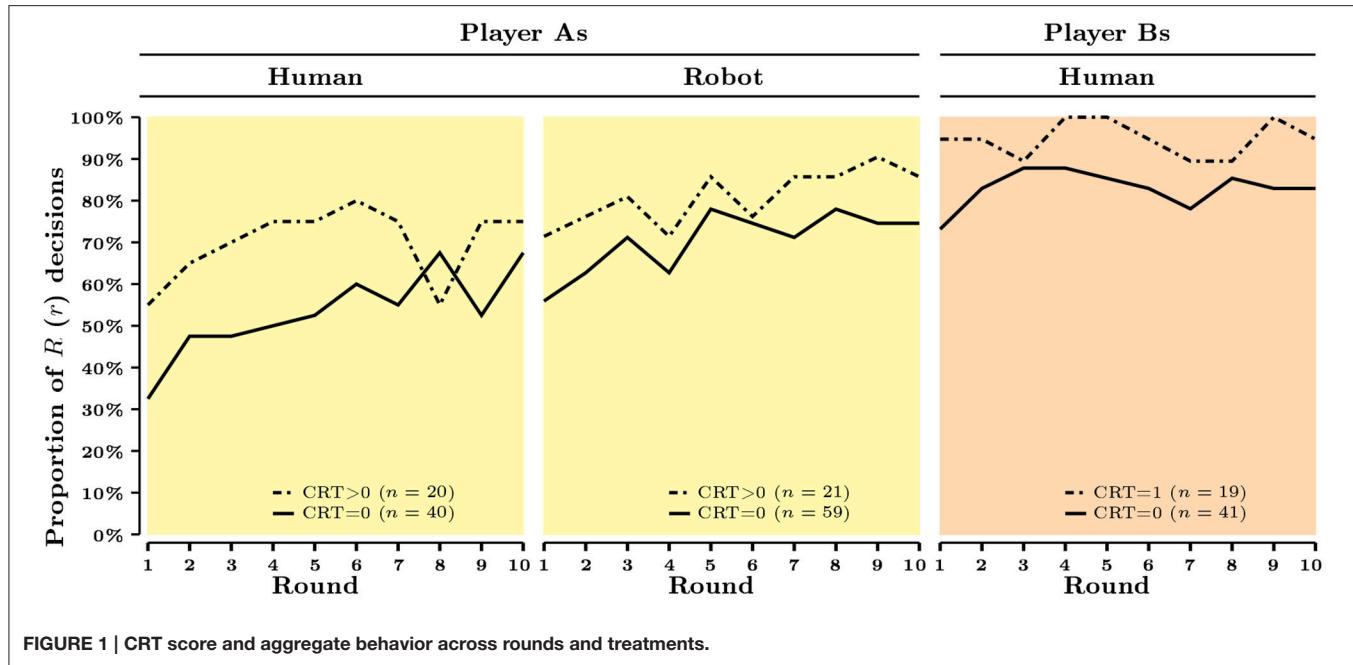


FIGURE 1 | CRT score and aggregate behavior across rounds and treatments.

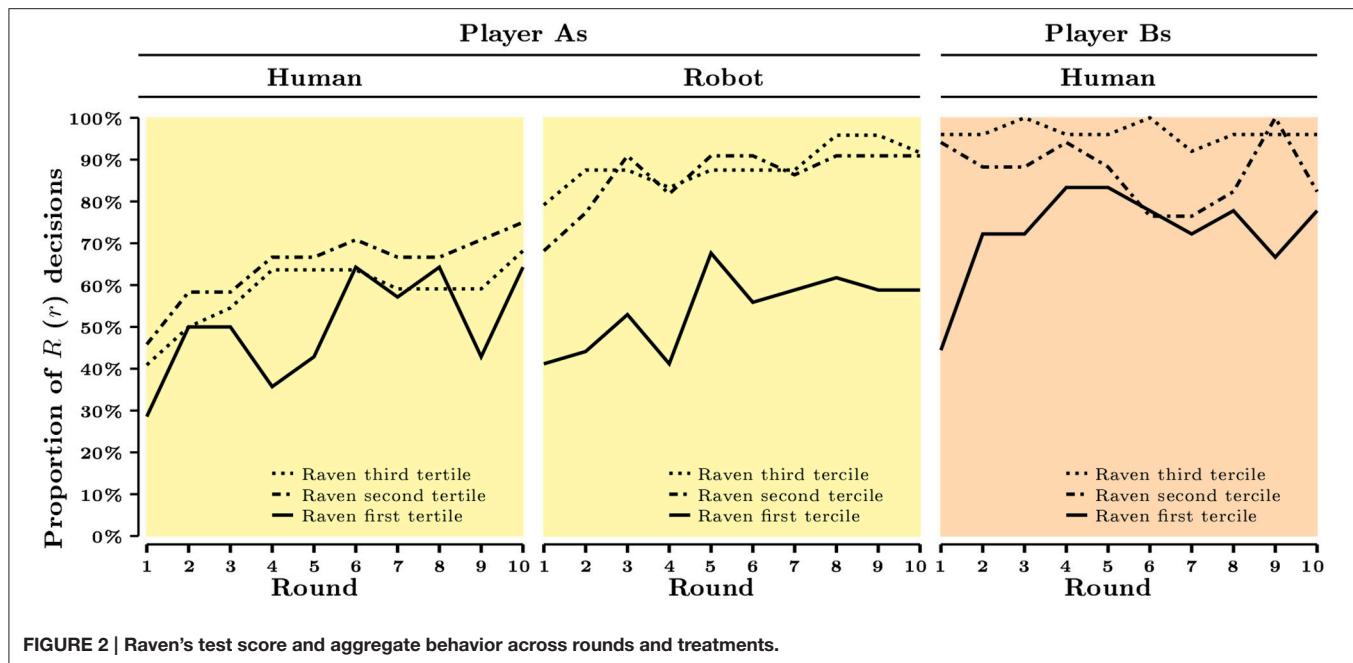


FIGURE 2 | Raven's test score and aggregate behavior across rounds and treatments.

that Raven's test score fails to predict player As' behavior once player Bs' behavior becomes uncertain, that is once we move from Robot to Human condition. This, in turn, points toward an interplay between the degree of strategic uncertainty, behavior in the experimental games, and individual cognitive skills. Importantly, the existence of such an interplay is also supported by **Figure 2** which shows that the aggregate levels of efficiency shift upwards between the Human condition and the Robot condition for the 2nd and 3rd Raven's score tertile, but not the 1st tertile.

In order to formally test this conjecture, we now look at the reaction of player As with different cognitive skills to the disappearance of strategic uncertainty. Splitting the data according to Raven's score tertile, for each of the three subsamples we compare player As' behavior in the Human condition to their behavior in the Robot condition by regressing player As' choice on the Robot dummy (set to 1 for the Robot and to 0 for the Human condition). We also include the previous set of independent variables (except for Raven's test score itself).

**TABLE 6 | Cognitive predictors of strategic behavior: regression analysis.**

	Pr ( <i>R</i> ) by player A						Pr ( <i>r</i> ) by player B		
	Human condition			Robot condition			Human condition		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Const.	0.423*** (0.080)	0.552** (0.176)	0.563** (0.197)	0.573*** (0.088)	0.242* (0.135)	0.240* (0.135)	0.705*** (0.027)	0.430*** (0.103)	0.444*** (0.099)
1[CRT>0]	0.131 (0.095)		0.152 (0.121)	0.062 (0.102)			0.109* (0.047)		0.046 (0.036)
Raven		0.013 (0.025)	0.018 (0.027)		0.0426*** (0.013)	0.0434*** (0.012)		0.0313** (0.009)	0.0287** (0.008)
1[Game 2]	-0.270* (0.129)	0.263 (0.139)	0.266 (0.136)	0.068 (0.083)	0.054 (0.076)	0.056 (0.079)	0.100 (0.052)	0.132* (0.056)	0.129* (0.056)
1[Male]	0.132 (0.126)	0.187 (0.100)	0.158 (0.107)	0.096 (0.090)	0.072 (0.076)	0.077 (0.089)	0.025 (0.046)	0.024 (0.046)	0.017 (0.047)
Round:									
2	0.133 (0.092)	0.133 (0.092)	0.133 (0.092)	0.063 (0.048)	0.063 (0.048)	0.063 (0.048)	0.067 (0.042)	0.067 (0.042)	0.067 (0.042)
3	0.150 (0.109)	0.150 (0.109)	0.150 (0.109)	0.138*** (0.050)	0.138*** (0.050)	0.138*** (0.050)	0.083 (0.048)	0.083 (0.048)	0.083 (0.048)
4	0.183** (0.070)	0.183** (0.070)	0.183** (0.070)	0.050 (0.047)	0.050 (0.047)	0.050 (0.047)	0.117* (0.048)	0.117* (0.048)	0.117* (0.048)
5	0.200* (0.089)	0.200* (0.089)	0.200* (0.089)	0.200*** (0.045)	0.200*** (0.045)	0.200*** (0.045)	0.100 (0.052)	0.100 (0.052)	0.100 (0.052)
6	0.267** (0.088)	0.267** (0.088)	0.267** (0.088)	0.150*** (0.054)	0.150*** (0.054)	0.150*** (0.054)	0.067 (0.049)	0.067 (0.049)	0.067 (0.049)
7	0.217* (0.098)	0.217* (0.098)	0.217* (0.098)	0.150*** (0.047)	0.150*** (0.047)	0.150*** (0.047)	0.017 (0.048)	0.017 (0.048)	0.017 (0.048)
8	0.233** (0.088)	0.233** (0.088)	0.233** (0.088)	0.200*** (0.057)	0.200*** (0.057)	0.200*** (0.057)	0.067 (0.056)	0.067 (0.056)	0.067 (0.056)
9	0.200 (0.113)	0.200 (0.113)	0.200 (0.113)	0.188*** (0.047)	0.188*** (0.047)	0.188*** (0.047)	0.083 (0.060)	0.083 (0.060)	0.083 (0.060)
10	0.300** (0.115)	0.300** (0.115)	0.300** (0.115)	0.175*** (0.056)	0.175*** (0.056)	0.175*** (0.056)	0.067 (0.049)	0.067 (0.049)	0.067 (0.049)
<i>R</i> <sup>2</sup>	0.151	0.141	0.160	0.050	0.139	0.140	0.060	0.108	0.111

Estimates of linear probability models explaining the likelihood of decision *R* by player A and decision *r* by player B. Standard errors (in parentheses) are clustered at the session level in the Human condition (three clusters per game matrix, six in total) and individual level in the Robot condition (40 clusters per game matrix, 80 in total) and computed using the delete-one jackknife procedure. Models 1 and 2 include a single measure of cognitive skills (a dummy set to 1 for a positive CRT score, or Raven's test score), while Model 3 combines both variables. Other independent variables include gender, game matrix and round dummies. The number of observations is  $N = 600$  for Human and  $N = 800$  for Robot conditions. \*\*/\*\*\* indicate significance at the 10/5/1% level.

These results are summarized in **Table 7**. The coefficient of the Robot dummy captures the effect of eliminating strategic uncertainty on player As' behavior for each of the three subsamples. This suggests that only player As with high enough cognitive skills are sensitive to the uncertainty about player Bs' behavior. The behavior of players with low Raven's test score (1st tertile) is unresponsive to the degree of strategic uncertainty: the coefficient of the Robot dummy is close to zero and insignificant ( $p = 0.822$ ). For players with medium scores (2nd tertile), we find a positive yet weakly significant effect ( $p = 0.087$ ) which becomes amplified and highly significant for those player As whose Raven's

test score belongs to the 3rd tertile of the experimental sample ( $p = 0.012$ ).

Finally, it is also worth noting that player As' reaction to the payoff scheme also varies as a function of Raven's test score. The coefficient of the Game 2 dummy is close to zero and highly insignificant in the 1st tertile regression ( $p = 0.890$ ). Then, it becomes negative in 2nd and 3rd tertile models (although it is only statistically significant in the former with  $p = 0.012$  and  $p = 0.271$ , respectively). This, in turn, stands in line with the previous finding that player As' willingness to play *R* increases as the safe choice *L* becomes less attractive relative to outcome

**TABLE 7 |** The effect of strategic uncertainty and cognitive skills: evidence from player As' behavior in Human and Robot conditions.

	Raven's test score tertile		
	1st	2nd	3rd
Constant	0.277*	0.592***	0.330**
	(0.147)	(0.065)	(0.135)
1[Robot]	0.044	0.158*	0.428**
	(0.195)	(0.090)	(0.155)
1[CRT>0]	0.002	0.038	0.016
	(0.262)	(0.066)	(0.188)
1[Male]	0.144	0.033	0.212
	(0.145)	(0.063)	(0.179)
1[Game 2]	0.034	-0.245**	-0.176
	(0.146)	(0.092)	(0.155)
Round dummies	Yes	Yes	Yes
N	480	610	310
R <sup>2</sup>	0.048	0.173	0.298

Estimates of linear probability models on decision R by player A. Standard errors (in parentheses) are clustered at the session level in the Human condition (three clusters per game matrix, six in total) and individual level in the Robot condition (40 clusters per game matrix, 80 in total) and computed using the delete-one jackknife procedure. Data from Human and Robot conditions are pooled and split into three subsamples based on Raven's test score tertiles. Other independent variables include a dummy set to 1 for a positive CRT score, as well as gender, game matrix and round dummies (omitted from the table). \*/\*\*/\*\*\* indicate significance at the 10/5/1% level.

(R, r). It also seems that the magnitude of this effect is mediated by player As' cognitive skills, although not in a monotone way.

## 4. CONCLUSION

This paper studies the relationship between strategic behavior and cognitive skills—cognitive reflection and fluid intelligence—in a classic  $2 \times 2$  dominance-solvable game. Our results show that subjects with higher fluid intelligence (measured by Raven's progressive matrices test) are more likely to play dominant strategy, and also more likely to best respond to other's strategy. Furthermore, fluid intelligence predicts strategic sophistication: only those players with sufficiently high Raven's test score are found to display sensitivity to the presence of uncertainty about others' behavior. Cognitive reflection (measured by CRT), in turn, lacks the power to predict behavior in our experimental setting. We see three main conclusions that stem from these findings.

First, these results contribute to the ongoing debate on the relationship between rationality and intelligence (see Stanovich, 2009, for a critical review). For instance, Stanovich and West (2014) distinguish between two aspects of rational behavior: instrumental rationality which is understood as the “ability to take appropriate action given one's goals and beliefs,” and epistemic rationality which enables agents to hold “beliefs that are commensurate with available evidence.” In the strategic environment investigated in this paper, instrumental rationality can be associated with the ability to solve the game, while epistemic rationality—with the ability to play it with others. Our

experimental data suggest an important relationship between fluid intelligence (rather than reflective thinking) and both of these facets of rationality in strategic settings. Both the ability to use dominance and iterated dominance to efficiently solve the game, as well as the responsiveness to the availability of strategic information, is found to be predicted by Raven's test score (but not by CRT score).

The second contribution is related to the experimental methodology. Despite the fact that CRT and Raven's test are both commonly used to measure cognitive skills in experimental subject pools, still very little is known about their relative performance in predicting different types of behavior. Therefore, the choice of one test over the other may happen to be at least as intuitive as evidence-based. As mentioned before, to the best of our knowledge only two experiments address this issue. Brañas-Garza et al. (2012) do so in a strategic environment (*p*-beauty contest game), while Corgnet et al. (2015a)—in a non-strategic one (individual choices on wealth distribution). Both studies find that CRT performs better than Raven's test in predicting subjects' behavior. The result of the present experiment points to the opposite conclusion. We believe that this difference is driven by the very nature of the experimental tasks which may involve different types of cognitive effort. In our view, this issue deserves attention in future research.

Finally, although we find evidence that behaving in accordance with dominance solvability is positively correlated with cognitive skills, we also substantiate that most of the variance in individual decision making cannot be explained by such skills. Thus, exploring factors alongside cognitive skills that generate strategic behavior remains an open and important empirical question. An interesting avenue is to disentangle individual determinants, e.g., personal characteristics (such as cognitive skills) that are associated with appropriate behavior, from environmental determinants, that is, those features of the decision making environment that lead decision makers to take certain types of actions.

## 5. AUTHOR CONTRIBUTIONS

NH, NJ, SL, and AZ all contributed equally to this work. Authors are listed in an alphabetical order.

## 6. FUNDING

This project has received funding from JSPS-ANR bilateral research grant BECOA (ANR-11-FRJA-0002), as well as the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, and LABEX OSE of the Paris School of Economics (ANR-10-LABX\_93-01), both within the program “Investissements d'Avenir” (ANR-11-IDEX-007) operated by the French National Research Agency (ANR). Ivan Ouss provided efficient research assistance. We thank Juergen Bracht, Colin Camerer, Guillaume Fréchette, Haoran He, Asen Ivanov, Frédéric Koessler, Rosemarie Nagel, Ariel Rubinstein, Jason F. Shogren, Jean-Marc Tallon, Antoine Terracol, and Marie Claire Villeval for their comments. NH and NJ gratefully acknowledge the Institut Universitaire de France. SL thanks the School of Business at the University of

Western Australia for hospitality and support. A major part of this work was conducted while NH was affiliated with Aix-Marseille University (Aix-Marseille School of Economics, AMSE) and NJ was affiliated with Université de Lorraine (BETA). NH and NJ thank both institutions for their various supports.

## REFERENCES

- Al-Ubaydli, O., Jones, G., and Weel, J. (2013). Patience, cognitive skill, and coordination in the repeated stag hunt. *J. Neurosci. Psychol. Econ.* 6, 71. doi: 10.1037/npe0000005
- Al-Ubaydli, O., Jones, G., and Weel, J. (in press). Average player traits as predictors of cooperation in a repeated prisoner's dilemma. *J. Behav. Exp. Econ.*
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J. Child Psychol. Psychiatry* 42, 241–251. doi: 10.1186/2040-2392-4-33
- Beard, T. R., Beil, R. O. J., and Mataga, Y. (2001). Reliant behavior in the United States and Japan. *Econ. Inquiry* 39, 270–279. doi: 10.1111/j.1465-7295.2001.tb00065.x
- Beard, T. R., and Richard O. B. (1994). Do people rely on the self-interested maximization of others? An experimental test. *Manage. Sci.* 40, 252–262.
- Bosch-Domènech, A., Brañas-Garza, P., and Espín, A. M. (2014). Can exposure to prenatal sex hormones (2d: 4d) predict cognitive reflection? *Psychoneuroendocrinology* 43, 1–10. doi: 10.1016/j.psyneuen.2014.01.023
- Bosch-Domènech, A., Montalvo, J. G., Nagel, R., and Satorra, A. (2002). One, two,(three), infinity,...: newspaper and lab beauty-contest experiments. *Am. Econ. Rev.* 92, 1687–1701. doi: 10.1016/j.psyneuen.2014.01.023
- Brañas-Garza, P., García-Muñoz, T., and Hernán González, R. (2012). Cognitive effort in the Beauty Contest Game. *J. Econ. Behav. Organ.* 83, 254–260. doi: 10.1016/j.jebo.2012.05.018
- Brañas Garza, P., Kujal, P., and Lenkei, B. (2015). *Cognitive Reflection Test: Whom, How, When*. MPRA Working Paper 68049. University Library of Munich.
- Burks, S. V., Carpenter, J. P., Goette, L., and Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7745–7750. doi: 10.1073/pnas.0812360106
- Burnham, T. C., Cesarin, D., Johannesson, M., Lichtenstein, P., and Wallace, B. (2009). Higher cognitive ability is associated with lower entries in a p-beauty contest. *J. Econ. Behav. Organ.* 72, 171–175. doi: 10.1016/j.jebo.2009.05.015
- Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.
- Carpenter, J., Graham, M., and Wolf, J. (2013). Cognitive ability and strategic sophistication. *Games Econ. Behav.* 80, 115–130. doi: 10.1016/j.geb.2013.02.012
- Cooper, D. J., and Van Huyck, J. B. (2003). Evidence on the equivalence of the strategic and extensive form representation of games. *J. Econ. Theory* 110, 290–308. doi: 10.1016/S0022-0531(03)00040-1
- Corgnet, B., Espín, A. M., and Hernán-González, R. (2015a). The cognitive basis of social behavior: cognitive reflection overrides antisocial but not always prosocial motives. *Front. Behav. Neurosci.* 9:287. doi: 10.3389/fnbeh.2015.00287
- Corgnet, B., Hernán González, R., and Mateo, R. (2015b). Cognitive reflection and the diligent worker: an experimental study of Millennials. *PLoS ONE* 10:e0141243. doi: 10.1371/journal.pone.0141243
- Coricelli, G., and Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9163–9168. doi: 10.1073/pnas.0807721106
- Costa-Gomes, M. A., and Crawford, V. P. (2006). Cognition and behavior in two-person guessing games: an experimental study. *Am. Econ. Rev.* 96, 1737–1768. doi: 10.1257/aer.96.5.1737
- Crawford, V. (2004). "Theory and experiment in the analysis of strategic interaction Chapter 12," in *Advances in Behavioral Economics*, eds C. F. Camerer, G. Loewenstein, and M. Rabin (Princeton, NJ: Princeton University Press), 344–373.
- Dessi, R., and Rustichini, A. (2015). *Strong Intrinsic Motivation*. IDEI Working Papers 847. Toulouse: Institut d'Économie Industrielle.
- Eckartz, K., Kirchkamp, O., and Schunk, D. (2012). *How do Incentives Affect Creativity?* CESifo Working Paper.
- Fehr, D., and Huck, S. (2015). Who knows it is a game? On strategic awareness and cognitive ability. *Exp. Econ.* doi: 10.1007/s10683-015-9461-0. [Epub ahead of print].
- Fehr, E., and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Q. J. Econ.* 114, 817–868. doi: 10.1162/00335399556151
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* 10, 171–178. doi: 10.1007/s10683-006-9159-4
- Frederick, S. (2005). Cognitive reflection and decision making. *J. Econ. Perspect.* 19, 25–42. doi: 10.1257/089533005775196732
- Gill, D., and Prowse, V. (forthcoming). Cognitive ability, character skills, and learning to play equilibrium: a level-k analysis. *J. Polit. Econ.*
- Goeree, J. K., and Holt, C. A. (2001). Ten little treasures of game theory and ten intuitive contradictions. *Am. Econ. Rev.* 91, 1402–1422. doi: 10.1257/aer.91.5.1402
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *J. Econ. Sci. Assoc.* 1, 114–125. doi: 10.1007/s40881-015-0004-4
- Hampton, A. N., Bossaerts, P., and O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6741–6746. doi: 10.1073/pnas.071199105
- Hanaki, N., Jacquemet, N., Luchini, S., and Zylbersztein, A. (2016). Cognitive ability and the effect of strategic uncertainty. *Theory Decis.* 81, 101–121. doi: 10.1007/s11238-015-9525-9
- Ho, T.-H., Camerer, C., and Weigelt, K. (1998). Iterated dominance and iterated best response in experimental "p-Beauty Contests." *Am. Econ. Rev.* 88, 947–969.
- Ibanez, A., Huepe, D., Gempp, R., Gutiérrez, V., Rivera-Rei, A., and Toledo, M. I. (2013). Empathy, sex and fluid intelligence as predictors of theory of mind. *Pers. Individ. Diff.* 54, 616–621. doi: 10.1016/j.paid.2012.11.022
- Jacquemet, N., Joule, R.-V., Luchini, S., and Shogren, J. F. (2013). Preference elicitation under oath. *J. Environ. Econ. Manage.* 65, 110–132. doi: 10.1016/j.jeem.2012.05.004
- Jacquemet, N., and Zylbersztein, A. (2013). Learning, words and actions: experimental evidence on coordination-improving information. *The BE J. Theor. Econ.* 13, 215–247. doi: 10.1515/bejte-2012-0018
- Jacquemet, N., and Zylbersztein, A. (2014). What drives failure to maximize payoffs in the lab? A test of the inequality aversion hypothesis. *Rev. Econ. Design* 18, 243–264. doi: 10.1007/s10058-014-0162-5
- Jones, G. (2008). Are smarter groups more cooperative? Evidence from prisoner's dilemma experiments, 1959–2003. *J. Econ. Behav. Organ.* 68, 489–497. doi: 10.1016/j.jebo.2008.06.010
- Moulin, H. (1986). *Game Theory for the Social Sciences*. New York , NY: New York University Press.
- Mullainathan, S., and Shafir, E. (2013). *Scarcity: Why Having Too Little Means So Much*. New York, NY: Times Books; Henry Holt and Company; LLC .
- Nagel, R. (1995). Unraveling in guessing games: an experimental study. *Am. Econ. Rev.* 85, 1313–1326.
- Proto, E., Rustichini, A., and Sofianos, A. (2014). *Higher Intelligence Groups Have Higher Cooperation Rates in the Repeated Prisoner's Dilemma*. IZA Working Paper DP, 8499.
- Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cogn. Psychol.* 41, 1–48. doi: 10.1006/cogp.199.0735

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01188>

- Raven, J. (ed.). (2008). "The Raven progressive matrices tests: their theoretical basis and measurement model Chapter 1," in *Uses and Abuses of Intelligence. Studies Advancing Spearman and Raven's Quest for Non-arbitrary Metrics*, (Cambridge University Press), 17–68.
- Rosenthal, R. W. (1981). Games of perfect information, predatory pricing and the chain-store paradox. *J. Econ. Theory* 25, 92–100.
- Schotter, A., Weigelt, K., and Wilson, C. (1994). A laboratory investigation of multiperson rationality and presentation effects. *Games Econ. Behav.* 6, 445–468.
- Stanovich, K. E. (2009). *What Intelligence Tests Miss: The Psychology of Rational Thought*. New Haven, CT: Yale University Press.
- Stanovich, K. E., and West, R. F. (2014). What intelligence tests miss. *Psychologist* 27, 80–83. Available online at: <https://thepsychologist.bps.org.uk/volume-27-edition-2/what-intelligence-tests-miss>
- Zeiliger, R. (2000). *A Presentation of Regate, Internet based Software for Experimental Economics*. Available online at: <http://regate-ng.gate.cnrs.fr/sferriol/>

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Hanaki, Jacquemet, Luchini and Zylbersztein. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Moderating Effects of Social Value Orientation on the Effect of Social Influence in Prosocial Decisions

Zhenyu Wei<sup>1,2</sup>, Zhiying Zhao<sup>3</sup> and Yong Zheng<sup>4\*</sup>

<sup>1</sup> Center for Studies of Education and Psychology of Ethnic Minorities in Southwest China, Southwest University, Chongqing, China, <sup>2</sup> Faculty of Psychology, Southwest University, Chongqing, China, <sup>3</sup> Key Laboratory for NeuroInformation of Ministry of Education, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China, <sup>4</sup> Key Laboratory of Cognition and Personality (MOE), Southwest University, Chongqing, China

## OPEN ACCESS

### Edited by:

Manuel Ignacio Ibáñez,  
Jaume I University, Spain

### Reviewed by:

Antonio M. Espín,  
Middlesex University, UK  
Gerardo Sabater-Grande,  
Universitat Jaume I, Spain

### \*Correspondence:

Yong Zheng  
zhengy@swu.edu.cn

### Specialty section:

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

Received: 04 April 2016

Accepted: 08 June 2016

Published: 21 June 2016

### Citation:

Wei Z, Zhao Z and Zheng Y (2016)  
Moderating Effects of Social Value Orientation on the Effect of Social Influence in Prosocial Decisions.  
*Front. Psychol.* 7:952.  
doi: 10.3389/fpsyg.2016.00952

Prosocial behaviors are susceptible to individuals' preferences regarding payoffs and social context. In the present study, we combined individual differences with social influence and attempted to discover the effect of social value orientation (SVO) and social influence on prosocial behavior in a trust game and a dictator game. Prosocial behavior in the trust game could be motivated by strategic considerations whereas individuals' decisions in the dictator game could be associated with their social preference. In the trust game, prosocials were less likely than proselves to conform to the behavior of other group members when the majority of group members distrusted the trustee. In the dictator game, the results of the three-way ANOVA indicated that, irrespective of the type of offer, in contrast to proselves, prosocials were influenced more by others' generous choices than their selfish choices, even if the selfish choices were beneficial to themselves. The overall results demonstrated that the effect of social influence appears to depend on individuals' SVO: that is, prosocials tend to conform to prosocial rather than proself behaviors.

**Keywords:** social value orientation, social influence, prosocial decision, trust, generosity

## INTRODUCTION

People often face mixed-motive social dilemmas in which their self-interest is at variance with what is best for their community (Balliet et al., 2009). Previous studies have shown that people differ in fundamental ways in how they approach and interact in social dilemmas (Van Lange et al., 2013a,b). Social value orientation (SVO) has been defined as a personal trait that reflects how people resolve social dilemmas (Messick and McClintock, 1968; Kelley and Stahelski, 1970; Kuhlman and Marshello, 1975; Liebrand, 1984; McClintock and Liebrand, 1988; Van Lange, 1999). The implications of individual differences in SVO refer to people's self-regarding versus other-regarding preferences (Van Lange, 2000). The most common manner of assessing SVO is by means of decomposed games (Liebrand, 1984; McClintock and Liebrand, 1988). Researchers have noted that three SVOs are common (Messick and McClintock, 1968): individuals can be classified as prosocials, individualists, and competitors. Prosocials are defined as individuals who attempt either to maximize the welfare of others or to choose joint gain. Individualists prefer to maximize their own welfare, showing little concern with others' outcomes. Finally, competitors attempt to maximize the difference between their own welfare and others' outcomes (Messick and McClintock, 1968; Kuhlman and Marshello, 1975; Van Lange, 1999). Because competitors show non-cooperative behavior similar to individualists' and the proportion of competitors is quite small,

previous studies have combined individualists and competitors into a category called “proselfs” (Van Lange and Liebrand, 1991; Van Lange et al., 1998; Bogaert et al., 2008).

Previous studies have attempted to link SVO with individuals’ behavior in prosocial decisions (McClintock and Allison, 1989; Van Lange et al., 1998, 2007; Kanagaretnam et al., 2009). Behavior is considered to be prosocial when it benefits others (Batson and Powell, 2003; Twenge et al., 2007; Piff et al., 2010; Zaki and Mitchell, 2011). Most cultures encourage or even require prosocial behavior because it is vital to the social system. People often perform prosocial behaviors because doing so enables them to belong to their community or society and to enjoy the social reward (i.e., a good reputation). Prior studies have demonstrated that prosocials are more generous in their helping responses than proselfs and more engaged in donating money to organizations aimed at helping the poor and the ill (McClintock and Allison, 1989; Van Lange et al., 2007). Prosocials also exhibit greater trust than individualists in the trust game (Kanagaretnam et al., 2009).

During the past decade, researchers have been interested in understanding how SVO interacts with features of a social situation to predict behavior (Balliet et al., 2009). Social influence plays an important role in our daily lives. We live in a highly complex social environment where social information continuously affects our perception and decision-making. Previous studies have shown that individuals tend to change their opinions and behaviors in order to align with group norms (Cialdini and Goldstein, 2004). This phenomenon, known as “social conformity”, refers to the action of changing one’s initial choices or opinions to match those of the group majority (Turner, 1991). Following the work of Asch (1951), psychologists have extensively examined the causes and underlying mechanisms of social conformity. Three motivations relate to conforming behavior: a desire to be correct, a desire to obtain social approval from others, and a desire to maintain a positive self-concept (Cialdini and Goldstein, 2004). Previous studies have shown that social influence can motivate people to behave prosocially (Shang and Croson, 2009; Nook et al., in press). However, they leave important questions unanswered because they say little about the individual differences in prosocial conformity. Some studies have demonstrated that conformity behavior could be modulated by personality traits (Steiner and Vannoy, 1966; DeYoung et al., 2002). From this perspective, SVO, which has been defined as a personal trait that reflects individuals’ social preferences, could affect individuals’ willingness to follow the majority in prosocial behavior. To address this question, we designed two tasks to investigate how SVO influences individuals’ conformity behaviors in trusting behavior and generous behavior.

In Study 1, we investigated the interaction between SVO and social influence in trusting behavior using the trust game. There are two players in the original trust game: an investor and a trustee (Berg et al., 1995). Both players are endowed with \$10. First, the investor decides whether to give the endowment to the trustee. Then, the amount given is multiplied by the experimenter. Finally, the trustee chooses whether to keep the amount he/she received or pass any portion of the money back to the investor. The amount passed by the investor is used to capture trust. Trust refers to a willingness to bet

that the other will reciprocate a risky move even at a cost to themselves (Camerer, 2003). Prosocial behavior in the trust game could emanate from strategic considerations (Espin et al., 2016). In the present study, we developed a variant of the trust game in which participants, who were able to see other group members’ choices before making a decision, were asked to decide whether to send the endowment to a stranger or to keep the endowment. We predicted that participants’ rate of trust in the trust game is dependent on their SVO. Compared with proselfs, prosocials should be more trusting in the trust game (Kanagaretnam et al., 2009). In addition, we hypothesized that the choices of the majority would affect participants’ trusting behavior: that is, subjects would trust the trustee when they see that the majority of the group does so. Further, previous studies have shown that conformity behavior could be affected by personality traits (Steiner and Vannoy, 1966; DeYoung et al., 2002). With the assumption that SVO, which has been defined as a fundamental personal trait that reflects how people resolve social dilemmas, could influence individuals’ decisions in the trust game (Kanagaretnam et al., 2009), we expected that the effect of social influence in the trust game would be modulated by SVO.

In Study 2, we were interested in the interaction between SVO and social influence in the dictator game. We used a modified dictator game, which was designed by Zaki and Mitchell (2011), to investigate the effect of social influence on generous behavior, free of strategic considerations. Participants made iterated choices about whether to allocate varying amounts of money to themselves or to another person (see *Experimental design and procedure in Study 2* for details). This task yields a behavioral measure of generosity (giving to the receiver at a cost to one’s self) (Zaki and Mitchell, 2011). Participants’ decision in this task could be motivated by their social preference rather than strategic considerations because the second player is passive. We assumed that participants’ choices would be dependent on their SVO: that is, compared with proselfs, prosocials would tend to make more generous choices in the dictator game. We also hypothesized that the choices of the majority would affect subjects’ choices: that is, subjects would make more generous choices when they saw that the majority of the group allocated money to the receiver. In the end, as prosocials show a natural willingness to help others and they are more generous than proselfs (McClintock and Allison, 1989; Van Lange et al., 2007), we predicted that the effect of social influence in the dictator game would be modulated by SVO and that prosocials might be less likely to be influenced by the selfish choices of group members.

## STUDY 1

### Materials and Methods

#### Participants

One hundred thirty-six healthy right-handed participants completed study 1. All were native Mandarin speakers, with no neurological or psychological disorders, and with normal color vision. Written informed consent was obtained after detailed explanation of the experiment. This study was conducted in

accordance with the Declaration of Helsinki and approved by the Ethics Committee of Southwest University.

### Measurement of Social Value Orientation

We used a questionnaire including a series of nine decomposed games to assess a participant's SVO (Van Lange and Kuhlman, 1994; Van Lange et al., 1997). This questionnaire is an efficient and easy-to-administer instrument (Van Lange et al., 1997). Subjects were classified as prosocial, individualistic, or competitive if at least six of nine decisions were consistent with a particular value orientation (Van Lange and Kuhlman, 1994; Van Lange et al., 1997). One hundred sixteen participants fell into one of three SVO. We identified 52 prosocials (35 females), 56 individualists (23 females) and 8 competitors (1 females). Following prior research on SVO, we combined the individualists and competitors to form a group of proselves (Van Lange and Liebrand, 1991; Van Lange et al., 1998).

### Experimental Design and Procedure

After arrived at the laboratory, participants were told that they would perform the experiment with four other participants, who were in separate rooms, but that they would see the choices of the other group members on the computer screen during the experiment. In the experiment, participants would play an on-line monetary game as an investor independently with 70 different strangers (trustees). The strangers were randomly selected from the university and played the game with participants through a local network. Other group members also did not know anything about trustees.

In each trial, both players were endowed with ¥2. The investor was restricted to the options of keeping the endowment or sending all ¥2. If the investor decided to send ¥2 to the trustee, this money would be tripled. Then the trustee was restricted to either send nothing back or send half of the tripled amount back (¥3). However, the investor would not know the outcome (i.e., trustee's choice) during the task. Subjects were told that they will receive ¥10 for participating in the experiment plus the additional money earned from ten of their decisions, chosen at random, during the trust game. Actually, subjects earned a show-up fee (¥15) and a bonus (¥4). We asked participants whether s/he believed the existence of trustees and group members after they finished the task. Six participants reported that they did not believe the existence of trustees. Therefore, their data were excluded in the analysis.

The hypotheses of this study were tested in a  $2 \times 3$  (SVO: Prosocial Orientation vs. Prosself Orientation  $\times$  Social Influence: No influence vs. Trust influence vs. Distrust influence) factorial design. The experiment contained one block (70 trials). The duration of a trial is approximately 11 s. In 10 of the trials, two peers decided to send the endowments to the trustee while the other two peers decided to keep the endowments. These trials were not included in the final analysis because they were used solely to maintain believability of the interaction between participant and the four peers. In one-third of the remaining trials (20 trials), the group's choices were hidden from the subject (the no information, or baseline condition; we told participants that they would not see their peers' choices in these trials because

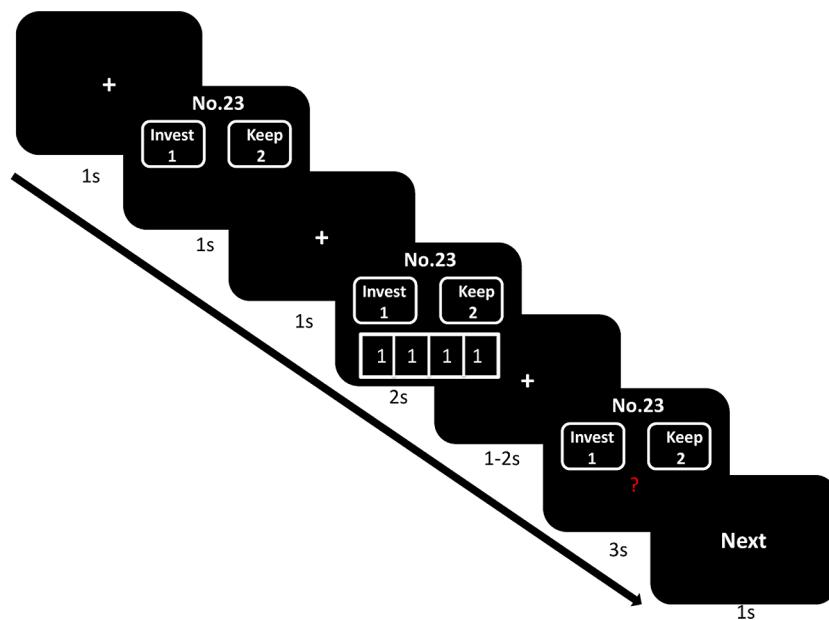
the decisions in these trials were not made by all of the four other members). In this situation, they would see four "x" symbols. For the 20 trials of the trust influence condition, three or four group members decided to send the endowments to the trustee. For the 20 trials of the distrust influence condition, one or none of the group members decided to send the endowments to the trustee. These trials were presented in a random order.

Participants then received details about the procedure of the experiment. At the beginning of each trial, the participants were presented with a fixation point for a 1 s duration. The offer would be shown on the screen for 1 s, followed by a fixation point for duration of 1 s. They could see the number of the trustee in the top of the offer screen. Then the choices of group members would be presented for 2 s under the offer, followed by a fixation point for duration of 1–2 s. Subsequently, the decision phase was shown on the screen for 3 s. Participant used the index and middle fingers of their right hand to respond to the offer by pressing keyboard ("1" to invest and "2" to keep the endowment). In the end, the word "next" displayed for 1 s, which indicated that the next trial was about to begin. The sequence of events in a trial is illustrated in **Figure 1**. Before performing the task, participants completed a training session. We told participants that the computer used to conduct the pre-experiment training was not connected to the local network, therefore the choices of the other group members would remain hidden. A PC running E-Prime 2.0 was used to display the stimuli and acquire the responses of the participants.

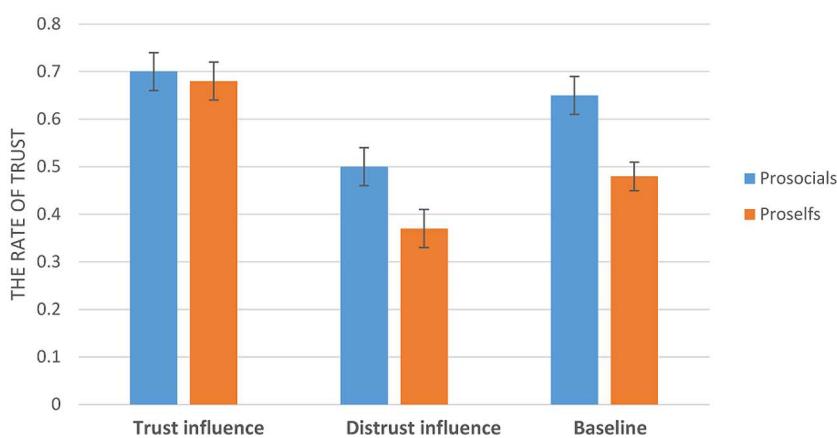
### Results

Trials in which the subjects did not respond in the decision stage were excluded from further data analyses. 5.3% of total trials were rejected to enter the following data analyses. Social influence effect was measured by the rate of trust of participants. A  $2$  (SVO: proselves, prosocials)  $\times$   $3$  (social influence: trust, distrust, baseline) repeated measure ANOVA revealed a significant main effect of the factor social influence,  $F(2,113) = 14.31, p < 0.001$ . Participants trusted the trustee at a significantly higher rate in the trust condition ( $M = 0.69, SD = 0.3$ ) than in the distrust condition ( $M = 0.43, SD = 0.32$ ) and baseline ( $M = 0.56, SD = 0.26$ ). The main effect of SVO was significant,  $F(1,114) = 10.74, p < 0.001$ . Prosocial individuals ( $M = 0.62, SD = 0.31$ ) trusted the trustee at a significantly higher rate than proselves ( $M = 0.51, SD = 0.31$ ).

The interaction between SVO and social influence was significant,  $F(2,113) = 4.23, p < 0.05$  (**Figure 2**). The results indicated that prosocial individuals ( $M = 0.65, SD = 0.24$ ) trusted the trustee at a significant higher rate than proselves ( $M = 0.48, SD = 0.25$ ) in the baseline condition,  $F(1,114) = 13.91, p < 0.001$ . In addition, prosocial individuals also ( $M = 0.5, SD = 0.33$ ) trusted the trustee at a significant higher rate than proselves ( $M = 0.37, SD = 0.3$ ) in the distrust condition,  $F(1,114) = 4.49, p < 0.05$ . The difference between prosocials ( $M = 0.7, SD = 0.31$ ) and proselves ( $M = 0.68, SD = 0.29$ ) in the trust condition was not significant,  $F(1,114) = 0.1, p = 0.75$ .



**FIGURE 1 | Demonstration of sequence of events in a trial (take trust influence as an illustration).**



**FIGURE 2 | The rate of trust (Error bars represent standard errors of the mean).**

## Discussion

A prior study found that genetics explain about 20% of the cross-sectional variation in trust game behavior (Cesarini et al., 2008), thus suggesting stable individual differences in trust. Our results, like those of Kanagaretnam et al. (2009), suggest that SVO may partially underlie such individual differences. However, the findings of Cesarini et al. (2008) also indicate that about 80% of variation must be explained by unknown environmental factors (Ahern et al., 2014). According to the present findings, social conformity might be one such factor since individuals' behavior in the trust game could indeed be influenced by the opinions of peers (as in other environments; see Cialdini and Goldstein, 2004). The present study showed that prosocials were less likely than proselfs to conform to group

members when the majority of group members did not trust the trustee.

Prosocials tend to consider the impact of their behavior on others and strive to maximize joint outcomes (De Cremer and Van Lange, 2001). They prefer to seek win-win situations in a disagreement (Van Lange et al., 1997). In contrast to prosocials, proselfs strive to maximize their own outcomes. Therefore, prosocials show a higher level of prosocial behavior than proselfs in the trust game (Kanagaretnam et al., 2009). In the present study, prosocials showed a lower level of conformity behavior than proselfs when group members distrusted the trustee. We infer that prosocials are less influenced by group members' distrust behavior because they are naturally prosocial and trusting individuals. In this vein, it might be argued that peers' choices in

the trust game serve as a cue of expected trustworthiness, which could affect individuals' emotional systems in decision making. Because participants were asked to make decisions under time pressure, the emotional reactions could guide their decisions. As a previous study showed, some people trust the trustee due to strategic self-interest whereas other people trust the trustee because of social efficiency reasons (Espín et al., 2016). Prosocials care more about the social efficiency whereas proselves tend to be self-interested. Therefore, prosocials still trust the trustee when they perceive that the trustee will not reciprocate (i.e., the distrust condition).

## STUDY 2

### Materials and Methods

#### Participants

One hundred three healthy right-handed participants completed study 2. All were native Mandarin speakers, with no neurological or psychological disorders, and with normal color vision. Written informed consent was obtained after detailed explanation of the experiment. This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Southwest University.

#### Measurement of Social Value Orientation

We used a questionnaire including a series of nine decomposed games to assess a participant's SVO (Van Lange and Kuhlman, 1994; Van Lange et al., 1997). Subjects were classified as prosocial, individualistic, or competitive if at least six of nine decisions were consistent with a particular value orientation (Van Lange and Kuhlman, 1994; Van Lange et al., 1997). Ninety-five participants fell into one of three SVO. We identified 47 prosocials (29 females), 42 individualists (23 females), and 6 competitors. Following prior research on SVO, we combined the individualists and competitors to form a group of proself individuals (Van Lange and Liebrand, 1991; Van Lange et al., 1998).

#### Experimental Design and Procedure

The hypotheses of this study were tested in a  $2 \times 2 \times 3$  (SVO: Prosocial Orientation vs. Proself Orientation  $\times$  Offer Type: Selfish vs. Generous  $\times$  Social Influence: No influence vs. Selfish influence vs. Generous influence) factorial design. Each trial began with two monetary offers, one associated with the participant and the other with the receiver. Participants made iterated choices about whether to allocate varying amounts of money to themselves or to the receiver. For example, if the offer assigned ¥1.00 to the participant and ¥3.00 to the receiver, participants should choose between ¥1.00 for themselves and ¥3.00 for the receiver. The amounts that each person stood to gain varied across trials but always adhered to one of a set of six ratios specifying the relationship between the self vs. other monetary amounts: 3:1, 2:1, 3:2, 4:3, 5:4, and 1:1. Each ratio could produce two relationships between the amounts that the participant and the receiver stood to gain. Thus there were eleven ratios in present experiment (3:1, 2:1, 3:2, 4:3, 5:4, 1:1, 4:5, 3:4, 2:3, 1:2, and 1:3). For each trial, a random value between ¥0.00

and ¥3.00 was chosen, and a second value was determined by transforming the first value according to the ratio that applied during that trial. For example, if the amount of ¥1.00 was selected and the ratio was 2:1, the other one was ¥0.5. The maximum amount that either the participant or receiver stood to gain in one trial was ¥9.00.

The experiment contained 120 trials. On average, a trial lasted 13 s. There were five types of offers in the present study. If the ratio were larger than 1:1 (e.g., 2:1), the offer was a selfish offer. If the ratio were smaller than 1:1 (e.g., 1:2), the offer was a generous offer. If the ratio was 1:1, the offer was an equal offer. Besides, we also added "pure-self" and "pure-other" offers in the experiment. During pure-self trials the participant was presented with offers of a non-zero amount of money (e.g., ¥1.00) for herself/himself and ¥0.00 for the receiver, while in the pure-other condition, the participant was presented with offers of ¥0.00 for herself/himself and a non-zero amount of money for the receiver. Finally, we also added non-reward trials in which participants chose between ¥0.00 for herself/himself and ¥0.00 for the receiver. Overall, there were fifty selfish offers, fifty generous offers, ten equal offers, ten pure-self offers, ten pure-other offers and ten non-reward offers. The selfish offer condition and generous offer condition each comprised 15 selfish influence trials, 15 generous influence trials, 15 baseline trials and 5 mediate influence trials (these trials were not included in the final analysis because they were used solely to maintain believability of the interaction between participant and the four peers). We only included trials in which the participant and receiver stood to gain unequal, non-zero amounts of money.

When participants arrived at the laboratory, they were told individually that they would participate in the experiment with another four subjects, who were in separate rooms. In the experiment, they would independently play a monetary game with a human receiver, who would be in the other room. Group members and participants knew nothing about each other and they were told that they would also not meet each other after the experiment. Participants were told that they would be making repeated decisions about whether to allocate money to themselves, or to the receiver. Five of their decisions, chosen at random by the system, would be enacted and added to the final payment. To minimize the influence of reputation motives on the participant's choices, participants were told that the receiver would not know that the participant had completed the distribution game, and that the additional compensation would simply be included in the receiver's payment after the experiment. The participant could observe the choices of the other four group members through a local network on the computer during the experiment, but group members would not know the participant's choices. In addition, because participants used different computers, and because the order of offer presentation is random, they would sometimes not see the choices of group members if a group member had not responded to the offer. In this situation, they would see four "x" symbols. These trials were classified as the baseline condition. These instructions allowed participants to believe in the existence of the other group members. We asked participants whether s/he believed the existence of group members and the receiver after they finished

the task. One participant reported that he did not believe the existence of group members and the receiver. Therefore, his data were excluded in the analysis.

Participants then received instructions about how the experiment would proceed. At the beginning of each trial, the participants were presented with a fixation point for 1 s. The offer would be shown on the screen for 2 s, followed by a fixation point for duration of 1–2 s. Then, the choices of group members would be presented for 2 s underneath the offer, followed by a fixation point that would last for duration of 1–2 s. In the end, the decision phase was shown on the screen for 3 s. Participants used the index and middle fingers of their right hand to respond to the offer by pressing one of the two buttons on the keyboard ("1" to allocate to self and "2" to allocate to the receiver). The decision phase was followed by the word "Next", which was displayed for 1 s and indicated that the next trial was about to begin. The sequence of events in a trial is illustrated in **Figure 3**. Before performing the task, participants completed a training session. We told participants that the computer used to conduct the pre-experiment training was not connected to the local network, therefore the choices of the other group members would remain hidden. A PC running E-Prime 2.0 was used to display the stimuli and acquire the responses of the participants.

## Results

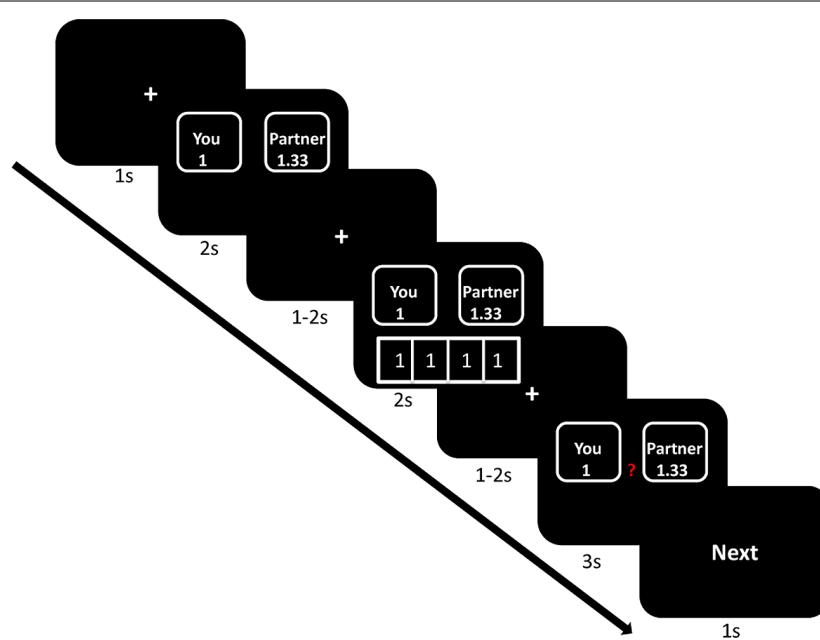
Trials in which the subjects did not respond in the decision stage were excluded from further data analyses. 4.6% of total trials were rejected to enter the following data analyses.

Social influence effect was measured by the rate of allocate money to the receiver. A 2 (SVO: proselves, prosocials)  $\times$  2 (offer: selfish, generous)  $\times$  3 (social influence: selfish, generous, baseline) repeated measure ANOVA revealed a significant main

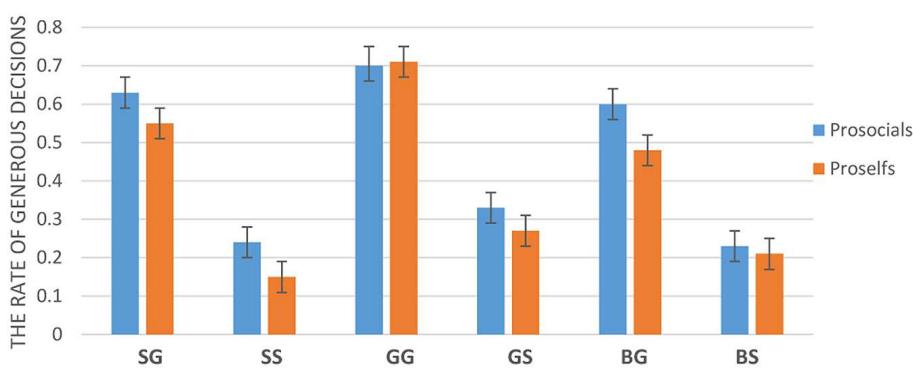
effect of the factor social influence,  $F(2,92) = 21.27, p < 0.001$ . Participants allocated money to receiver at a significantly higher rate in the generous influence condition ( $M = 0.5, SD = 0.35$ ) than in the selfish influence condition ( $M = 0.39, SD = 0.34$ ). The main effect of offer was significant,  $F(1,93) = 93.87, p < 0.001$ . Participants allocated money to the receiver at a significantly higher rate in generous offer condition ( $M = 0.61, SD = 0.31$ ) than in selfish offer condition ( $M = 0.24, SD = 0.26$ ). This result is consistent with a previous study (Zaki and Mitchell, 2011). The main effect of SVO was close to significance ( $M_{proselfs} = 0.39, SD_{proselfs} = 0.22; M_{prosocials} = 0.45, SD_{prosocials} = 0.23, F(1,93) = 3.47, p = 0.067$ ).

The interaction between offer and social influence was significant,  $F(2,92) = 8.49, p < 0.001$ . The result indicated that participants allocated money to the receiver at a significantly higher rate ( $M = 0.7, SD = 0.3$ ) in generous offer-generous influence condition than in generous offer-selfish influence condition ( $M = 0.59, SD = 0.3$ ),  $p < 0.001$ , and in generous offer-baseline condition ( $M = 0.54, SD = 0.3$ ),  $p < 0.01$ . In the selfish offer condition, participants allocated money to the receiver at a significantly higher rate in the generous influence condition ( $M = 0.3, SD = 0.25$ ) than in the selfish influence condition ( $M = 0.19, SD = 0.25$ ),  $p < 0.001$ , and in baseline ( $M = 0.22, SD = 0.25$ ),  $p < 0.001$ . The difference between selfish influence condition and baseline condition was not significant,  $p = 0.081$ . The interaction between SVO and social influence was not significant,  $F(2,92) = 3.47, p = 0.478$ .

The interaction between SVO, offer type and social influence was significant,  $F(2,92) = 4.97, p < 0.01$  (**Figure 4**). Regardless of the type of offer, proselves allocated money to the receiver at a significantly higher rate in the generous influence condition ( $M_{selfish\ offer} = 0.27, SD_{selfish\ offer} = 0.19$ ;



**FIGURE 3 |** Demonstration of sequence of events in a trial (take generous offer and selfish influence as an illustration).



**FIGURE 4 |** The rate of generous decisions (SG: selfish influence-generous offer; SS: selfish influence-selfish offer; GG: generous influence-generous offer; GS: generous influence-selfish offer; BG: baseline-generous offer; BS: baseline-selfish offer. Error bars represent standard errors of the mean).

$M_{\text{generous offer}} = 0.71$ ,  $SD_{\text{generous offer}} = 0.26$ ) than in selfish influence condition ( $M_{\text{selfish offer}} = 0.15$ ,  $SD_{\text{selfish offer}} = 0.16$ ,  $p < 0.001$ ;  $M_{\text{generous offer}} = 0.55$ ,  $SD_{\text{generous offer}} = 0.25$ ,  $p < 0.001$ ), and in baseline condition ( $M_{\text{selfish offer}} = 0.21$ ,  $SD_{\text{selfish offer}} = 0.2$ ,  $p < 0.05$ ;  $M_{\text{generous offer}} = 0.48$ ,  $SD_{\text{generous offer}} = 0.26$ ,  $p < 0.001$ ). For prosocials, they allocated money to the receiver at a significantly higher rate in generous influence condition ( $M = 0.7$ ,  $SD = 0.35$ ) than in the baseline ( $M = 0.6$ ,  $SD = 0.33$ ) when the offer is generous offer,  $p < 0.05$ . However, the difference between selfish influence condition ( $M = 0.63$ ,  $SD = 0.34$ ) and baseline was not significant,  $p = 0.268$ . In addition, the difference between selfish influence condition and generous influence condition was also not significant,  $p = 0.126$ . In the selfish offer condition, prosocials allocated money to the receiver at a significantly higher rate in the generous influence condition ( $M = 0.33$ ,  $SD = 0.31$ ) than in selfish influence condition ( $M = 0.24$ ,  $SD = 0.32$ ,  $p < 0.01$ ) and in baseline condition ( $M = 0.23$ ,  $SD = 0.3$ ,  $p < 0.001$ ). The difference between selfish influence condition and baseline was not significant,  $p = 0.926$ .

## Discussion

Study 2 set out to investigate the effects of SVO and social influence in generous decisions. People often change their decisions and judgments to conform to normative group behavior (Cialdini and Goldstein, 2004; Klucharev et al., 2009; Wei et al., 2013). The present study showed that individuals' generous decisions can be influenced by the group members' choices; however, this effect can be modulated by individuals' SVO. Results of the three-way ANOVA showed that no matter the offers were selfish or generous, proselves were influenced by others' selfish choices and generous choices. However, when it comes to prosocials, they were influenced by others' generous choices rather than their selfish choices.

Generosity is defined as helping another at a cost to oneself; therefore, generosity is a kind of prosocial behavior (Zak et al., 2007). Prosocials have a stable preference for maximizing joint outcomes, but proselves prefer to maximize their own benefits (Van Lange, 2000). Additionally, prior studies have

demonstrated that prosocials are more generous in their helping responses than proselves and more engaged in donating money to organizations aimed at helping the poor and the ill (McClintock and Allison, 1989; Van Lange et al., 2007). We infer that selfish choices are conflict with prosocials' social preference and prosocials know that selfish behavior is not encouraged by social norms. Therefore, in both offer conditions, prosocials were influenced by others' generous choices rather than their selfish choices, even if the selfish choices were beneficial to themselves.

## GENERAL DISCUSSION

Social value orientation is regarded as a stable personality trait that reflects how people evaluate outcomes for self and others (Messick and McClintock, 1968). Individual SVO can determine and predict individuals' choice behavior in a wide variety of decisions (Messick and McClintock, 1968; Kuhlman and Marshello, 1975; Van Lange, 1999), including prosocial decisions (McClintock and Allison, 1989; Van Lange et al., 1998, 2007; Kanagaretnam et al., 2009). According to previous studies, prosocials tend to trust others, and they are more generous than proselves (McClintock and Allison, 1989; Van Lange et al., 2007; Kanagaretnam et al., 2009). In the present study, in agreement with previous ones, we found that people tend to conform to the choices of group members in prosocial decisions. However, our study also found that individuals' SVO could modulate the effect of social influence in prosocial decisions. Relative to proselves, prosocials were less likely to conform to proself behaviors. We infer that prosocials know that proself behavior is not accepted by general social norms and they can resist the proself choices of other group members. Proselves, as well, know that prosocial behavior is encouraged by social norms. Therefore, they would experience group pressure when they realized that the majority was choosing the prosocial option (Asch, 1951; Strickland and Crowne, 1962; Becker et al., 1964) and would then be more likely to conform to the choices of group members, even when these choices conflicted with their own preferences.

## CONCLUSION

Prior experimental studies have provided evidence that prosocial behaviors are susceptible to individuals' preferences for payoffs and social context (McCabe et al., 2003; Boone et al., 2008; Declerck et al., 2010). In the present studies, we combined individual differences with social influence in an attempt to discover the effect of SVO and social influence on prosocial behavior in the trust game and the dictator game. Our results extend our current understanding of prosocial conformity by showing that the effect of social influence on prosocial behavior depends on a person's SVO. Prosocials tend to follow prosocial choices rather than proself behaviors. Prosocials have a natural willingness to behave prosocially and they know that prosocial behavior is encouraged by social norms (McClintock and Allison, 1989; Van Lange et al., 2007). Therefore, they can resist the proself influence that conflicts with their own social preference.

## REFERENCES

- Ahern, K. R., Duchin, R., and Shumway, T. (2014). Peer effects in risk aversion and trust. *Rev. Financ. Stud.* 27, 3213–3240. doi: 10.1093/rfs/hhu042
- Asch, S. (1951). *Effects of Group Pressure upon the Modification and Distortion of Judgments*. Pittsburgh, PA: Carnegie Press.
- Balliet, D., Parks, C., and Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: a meta-analysis. *Group Process. Intergroup Relat.* 12, 533–547. doi: 10.1177/1368430209105040
- Batson, C. D., and Powell, A. A. (2003). "Altruism and prosocial behavior: part III," in *Handbook of Psychology: Personality and Social Psychology*, Vol. 5, eds T. Millon and M. J. Lerner (New York, NY: John Wiley & Sons, Inc.), 463–484.
- Becker, S. W., Lerner, M. J., and Carroll, J. (1964). Conformity as a function of birth order, payoff, and type of group pressure. *J. Abnorm. Soc. Psychol.* 69, 318–323. doi: 10.1037/h0044843
- Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games Econ. Behav.* 10, 122–142. doi: 10.1006/game.1995.1027
- Bogaert, S., Boone, C., and Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: a review and conceptual model. *Br. J. Soc. Psychol.* 47, 453–480. doi: 10.1348/014466607X244970
- Boone, C., Declerck, C. H., and Suetens, S. (2008). Subtle social cues, explicit incentives and cooperation in social dilemmas. *Evol. Hum. Behav.* 29, 179–188. doi: 10.1016/j.evolhumbehav.2007.12.005
- Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: University Press, Princeton.
- Cesarini, D., Dawes, C. T., Fowler, J. H., Johannesson, M., Lichtenstein, P., and Wallace, B. (2008). Heritability of cooperative behavior in the trust game. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3721–3726. doi: 10.1073/pnas.0710069105
- Cialdini, R. B., and Goldstein, N. J. (2004). Social influence: compliance and conformity. *Annu. Rev. Psychol.* 55, 591–621. doi: 10.1146/annurev.psych.55.090902.142015
- De Cremer, D., and Van Lange, P. A. M. (2001). Why prosocials exhibit greater cooperation than proselfs: the role of social responsibility and reciprocity. *Eur. J. Pers.* 15, 5–18. doi: 10.1002/per.418
- Declerck, C. H., Boone, C., and Kiyonari, T. (2010). Oxytocin and cooperation under conditions of uncertainty: the modulating role of incentives and social information. *Horm. Behav.* 57, 368–374. doi: 10.1016/j.yhbeh.2010.01.006
- DeYoung, C. G., Peterson, J. B., and Higgins, D. M. (2002). Higher-order factors of the big five predict conformity: are there neuroses of health? *Pers. Individ. Dif.* 33, 533–552. doi: 10.1016/S0019-8869(01)00171-4
- Espin, A. M., Exadaktylos, F., and Neyse, L. (2016). Heterogeneous motives in the trust game: a tale of two roles. *Front. Psychol.* 7:728. doi: 10.3389/fpsyg.2016.00728
- Kanagaretnam, K., Mestelman, S., Nainar, K., and Shehata, M. (2009). The impact of social value orientation and risk attitudes on trust and reciprocity. *J. Econ. Psychol.* 30, 368–380. doi: 10.1016/j.jeop.2008.12.003
- Kelley, H. H., and Stahelski, A. J. (1970). Social interaction basis of cooperators' and competitors' beliefs about others. *J. Pers. Soc. Psychol.* 16, 66–91. doi: 10.1037/h0029849
- Kharchev, V., Hytönen, K., Rijpkema, M., Smidts, A., and Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron* 61, 140–151. doi: 10.1016/j.neuron.2008.11.027
- Kuhlman, D. M., and Marshello, A. F. (1975). Individual differences in game motivation as moderators of preprogrammed strategy effects in prisoner's dilemma. *J. Pers. Soc. Psychol.* 32, 922–931. doi: 10.1037/0022-3514.32.5.922
- Liebrand, W. B. G. (1984). The effect of social motives, communication and group size on behaviour in an N-person multi-stage mixed-motive game. *Eur. J. Soc. Psychol.* 14, 239–264. doi: 10.1002/ejsp.2420140302
- McCabe, K., Rigdon, M. L., and Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *J. Econ. Behav. Organ.* 52, 267–275. doi: 10.1016/S0167-2681(03)00003-9
- McClintock, C. G., and Allison, S. T. (1989). Social value orientation and helping behavior. *J. Appl. Soc. Psychol.* 19, 353–362. doi: 10.1111/j.1559-1816.1989.tb00060.x
- McClintock, C. G., and Liebrand, W. B. G. (1988). Role of interdependence structure, individual value orientation, and another's strategy in social decision making: a transformational analysis. *J. Pers. Soc. Psychol.* 55, 396–409. doi: 10.1037/0022-3514.55.3.396
- Messick, D., and McClintock, C. G. (1968). Motivational bases of choice in experimental games. *J. Exp. Soc. Psychol.* 4, 1–25. doi: 10.1016/0022-1031(68)90046-2
- Nook, E. C., Ong, D. C., Morelli, S. A., Mitchell, J. P., and Zaki, J. (in press). Prosocial conformity: prosocial norms generalize across behavior and empathy. *Pers. Soc. Psychol. Bull.* doi: 10.1177/0146167216649932
- Piff, P. K., Kraus, M. W., Côté, S., Cheng, B. H., and Keltner, D. (2010). Having less, giving more: the influence of social class on prosocial behavior. *J. Pers. Soc. Psychol.* 99, 771–784. doi: 10.1037/a0020092
- Shang, J., and Croson, R. (2009). Field experiments in charitable contributions: the impact of social influence on the voluntary provision of public goods. *Econ. J.* 119, 1422–1439.
- Steiner, I. D., and Vannoy, J. S. (1966). Personality correlates of two types of conformity behavior. *J. Pers. Soc. Psychol.* 4, 307–315. doi: 10.1037/h0023707
- Strickland, B. R., and Crowne, D. P. (1962). Conformity under conditions of simulated group pressure as a function of the need for social approval. *J. Soc. Psychol.* 58, 171–181. doi: 10.1080/00224545.1962.9712366
- Turner, J. C. (1991). *Social Influence*. London: Open University Press.

## AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: ZW and YZ. Program the task: ZW and ZZ. Performed the experiments: ZW. Analyzed the data: ZW. Wrote the paper: ZW, ZZ, and YZ.

## FUNDING

This work is supported by the MOE Project of Key Research Institute of Humanities and Social Sciences at Universities, China (#15JJDZONGHE022).

## ACKNOWLEDGMENT

The authors thank the two reviewers for their constructive comments that improved the manuscript considerably.

- Twenge, J. M., Baumeister, R. F., DeWall, C. N., Ciarocco, N. J., and Bartels, J. M. (2007). Social exclusion decreases prosocial behavior. *J. Pers. Soc. Psychol.* 92, 56–66. doi: 10.1037/0022-3514.92.1.56
- Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: an integrative model of social value orientation. *J. Pers. Soc. Psychol.* 77, 337–349. doi: 10.1016/j.actpsy.2008.01.005
- Van Lange, P. A. M. (2000). Beyond self-interest: a set of propositions relevant to interpersonal orientations. *Eur. Rev. Soc. Psychol.* 11, 297–331. doi: 10.1080/14792772043000068
- Van Lange, P. A. M., Balliet, D. P., Parks, C. D., and Van Vugt, M. (2013a). *Social Dilemmas: Understanding Human Cooperation*. Oxford: Oxford University Press.
- Van Lange, P. A. M., Bekkers, R., Schuyt, T. N., and Vugt, M. V. (2007). From games to giving: social value orientation predicts donations to noble causes. *Basic Appl. Soc. Psychol.* 29, 375–384. doi: 10.1080/0197353070165223
- Van Lange, P. A. M., De Bruin, E., Otten, W., and Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: theory and preliminary evidence. *J. Pers. Soc. Psychol.* 73, 733–746. doi: 10.1037/0022-3514.73.4.733
- Van Lange, P. A. M., Joireman, J., Parks, C. D., and Van Dijk, E. (2013b). The psychology of social dilemmas: a review. *Organ. Behav. Hum. Decis. Process.* 120, 125–141. doi: 10.1016/j.obhdp.2012.11.003
- Van Lange, P. A. M., and Kuhlman, D. M. (1994). Social value orientations and impressions of partner's honesty and intelligence: a test of the might versus morality effect. *J. Pers. Soc. Psychol.* 67, 126–141. doi: 10.1037/0022-3514.67.1.126
- Van Lange, P. A. M., and Liebrand, W. B. G. (1991). Social value orientation and intelligence: a test of the goal-prescribes-rationality principle. *Eur. J. Soc. Psychol.* 21, 273–292. doi: 10.1002/ejsp.2420210402
- Van Lange, P. A. M., Vugt, M. V., Meertens, R. M., and Ruiter, R. A. (1998). A social dilemma analysis of commuting preferences: the roles of social value orientation and trust. *J. Appl. Soc. Psychol.* 28, 796–820. doi: 10.1111/j.1559-1816.1998.tb01732.x
- Wei, Z., Zhao, Z., and Zheng, Y. (2013). Neural mechanisms underlying social conformity in an ultimatum game. *Front. Hum. Neurosci.* 7:896. doi: 10.3389/fnhum.2013.00896
- Zak, P. J., Stanton, A. A., and Ahmadi, S. (2007). Oxytocin increases generosity in humans. *PLoS ONE* 2:1128. doi: 10.1371/journal.pone.0001128
- Zaki, J., and Mitchell, J. P. (2011). Equitable decision making is associated with neural markers of intrinsic value. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19761–19766. doi: 10.1073/pnas.1112324108
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- The reviewer GG and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.
- Copyright © 2016 Wei, Zhao and Zheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Pay What You Want! A Pilot Study on Neural Correlates of Voluntary Payments for Music

Simon Waskow<sup>1,2†</sup>, Sebastian Markett<sup>1,3\*†</sup>, Christian Montag<sup>4,5\*</sup>, Bernd Weber<sup>3,6,7</sup>, Peter Trautner<sup>3,7</sup>, Volkmar Kramarz<sup>8</sup> and Martin Reuter<sup>1,3</sup>

<sup>1</sup> Department of Psychology, University of Bonn, Bonn, Germany, <sup>2</sup> Department of Philosophy, University of Bonn, Bonn, Germany, <sup>3</sup> Center for Economics and Neuroscience, University of Bonn, Bonn, Germany, <sup>4</sup> Institute of Psychology and Education, Ulm University, Ulm, Germany, <sup>5</sup> Key Laboratory for NeuroInformation, Center for Information in Medicine, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China, <sup>6</sup> Department of Epileptology, University Hospital Bonn, Bonn, Germany, <sup>7</sup> Department of NeuroCognition, Life and Brain Center Bonn, Bonn, Germany, <sup>8</sup> Department of Sound Studies, University of Bonn, Bonn, Germany

## OPEN ACCESS

### Edited by:

Nikolaos Georgantzis,  
University of Reading, UK

### Reviewed by:

Stephen Garcia,  
University of Michigan, USA  
Sören Enge,  
Technische Universität Dresden,  
Germany

### \*Correspondence:

Sebastian Markett  
sebastian.markett@uni-bonn-diff.de  
Christian Montag  
christian.montag@uni-ulm.de

<sup>†</sup>These authors have contributed  
equally to this work and share the  
first-authorship.

### Specialty section:

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

Received: 26 March 2016

Accepted: 21 June 2016

Published: 06 July 2016

### Citation:

Waskow S, Markett S, Montag C, Weber B, Trautner P, Kramarz V and Reuter M (2016) Pay What You Want!  
A Pilot Study on Neural Correlates of Voluntary Payments for Music.  
*Front. Psychol.* 7:1023.  
doi: 10.3389/fpsyg.2016.01023

Pay-what-you-want (PWYW) is an alternative pricing mechanism for consumer goods. It describes an exchange situation in which the price for a given good is not set by the seller but freely chosen by the buyer. In recent years, many enterprises have made use of PWYW auctions. The somewhat contra-intuitive success of PWYW has sparked a great deal of behavioral work on economical decision making in PWYW contexts in the past. Empirical studies on the neural basis of PWYW decisions, however, are scarce. In the present paper, we present an experimental protocol to study PWYW decision making while simultaneously acquiring functional magnetic resonance imaging data. Participants have the possibility to buy music either under a traditional “fixed-price” (FP) condition or in a condition that allows them to freely decide on the price. The behavioral data from our experiment replicate previous results on the general feasibility of the PWYW mechanism. On the neural level, we observe distinct differences between the two conditions: In the FP-condition, neural activity in frontal areas during decision-making correlates positively with the participants’ willingness to pay. No such relationship was observed under PWYW conditions in any neural structure. Directly comparing neural activity during PWYW and the FP-condition we observed stronger activity of the lingual gyrus during PWYW decisions. Results demonstrate the usability of our experimental paradigm for future investigations into PWYW decision-making and provides first insights into neural mechanisms during self-determined pricing decisions.

**Keywords:** pay what you want, decision-making and neuroeconomics, music cognition, emotional utility, pricing mechanism

## INTRODUCTION

In October 2007, the critically acclaimed band Radiohead provided the most prominent example of the use of Pay-What-You-Want (PWYW) to date, when they offered their fans to pay whatever they wanted for the electronic version of the band’s album *In Rainbows* (Benkler, 2011). In the meantime other bands followed this example in similar ways. Implementations of PWYW are not limited to music distribution but can be found across other economic fields, such as gastronomy and the hotel industry<sup>1</sup>. The somewhat paradoxical

<sup>1</sup>[www.pay-what-you-want.net](http://www.pay-what-you-want.net)

success of the PWYW pricing mechanism has triggered a fair amount of scientific research that has demonstrated PWYW's profitability in various settings (Kim et al., 2009; Regner and Barria, 2009; Belsky et al., 2010; Gneezy et al., 2012; Riener and Traxler, 2012). The question raised by the fact that people pay voluntarily under PWYW conditions is: Why do they do it? Rephrased in economic terms, this means: How do they derive utility from paying a fair amount of money for something that they can get for free? Researchers commonly assert that the buyer's motivation to pay for a product that they could – in principle – also get for free, is due to the power of social norms, which may outweigh explicit market norms. This, however, is incongruent with the traditional economic view of man as a homo economicus (Persky, 1995; Heyman and Ariely, 2004; Kim et al., 2009). It is assumed, that peoples' preferences to these social norms are enacted through some kind of *emotional utility*, that can compensate for reduced *monetary utility* (Kim et al., 2009). This view is in line with the dominant neuroeconomic approach to the problem of social preferences (Fehr and Camerer, 2007) and thus, PWYW research may also contribute to new insights in this field.

For the present study, we created a paradigm, in which PWYW payments can be directly compared to payments in a control condition. This control condition was designed to enable us to isolate the unique aspect of PWYW payments, namely that they are given on a voluntary basis, while holding all other factors constant. While other studies compared PWYW payments with a fixed price condition in a between subjects design (Gneezy et al., 2012), the present study is the first to our knowledge to implement PWYW and control condition on the same subjects and in a laboratory environment, which allows a much higher degree of control of confounding variables. We further designed our paradigm to be suitable for fMRI to investigate the neural correlates of PWYW decisions. Our experimental design involved repeated decisions to buy digital music albums. After listening to song snippets (the "listening stage"), participants were asked whether they wanted to obtain this album and how much they were willing to pay ("decision stage"). Crucially, the experimental manipulation included two different contexts at this point. In our control condition, the fixed-price condition (FP condition), participants made a bid on a product with an unknown, randomly determined selling price (the fixed price) and only received the album if their bid was higher than the unknown FP. The PWYW condition in contrast allowed the participant to pay whatever they wanted for the album.

The *FP-condition* required participants to make a rather rational purchase decision, based on only two preferences, one for purchasing a product (music) and an opposing one for keeping the money this product would cost. Multiple studies suggest that the fronto-mesolimbic reward system plays a crucial role in product valuation by responding with increasing activity to increasing valuation (Erk et al., 2002; Knutson et al., 2007; Plassmann et al., 2007). Since the FP (control) condition of our study represents a product-evaluation-decision paradigm we expect a correlation of neural activity in fronto-mesolimbic regions during the *decision stage* (i.e., when confronted with the

instruction screen after listening to the song) and the prices paid by the participant.

During the *listening stage*, the neuronal response to our product (i.e., music) has to be taken into account. Converging evidence suggests that the striatal reward system responds to music pleasing to the listener (Blood and Zatorre, 2001; Koelsch et al., 2006; Montag et al., 2011; Salimpoor et al., 2011). Additionally, higher activations in the OFC have been reported in response to attractive products (Erk et al., 2002). We therefore expect that higher striatal and orbito-frontal response to music during the listening stage relates to higher willingness to pay (WTP) in both conditions.

The *PWYW condition* was designed to match the FP condition as closely as possible with the one exception that participants were able to choose any price for the music, including 0.00€. To our knowledge PWYW pricing has only been studied on the behavioral level. In the present study, we will first seek to replicate the findings from these pioneering studies before we proceed to interpret neuronal contrasts between PWYW and FP-condition. In an early study, Kim et al. (2009) delivered evidence for the profitability of PWYW in various real life settings. Their model explains a buyer's WTP under PWYW conditions ( $WTP_{PWYW}$ ) as a function of his or her internal reference price, which is the price last paid for the same or a similar product. The reference price is assumed to represent a buyer's WTP for a given product under general (fixed-price) conditions. A higher general WTP will lead to a higher  $WTP_{PWYW}$ . However, buyers are expected to go for some monetary profit, so their  $WTP_{PWYW}$  will be smaller than their general WTP. We therefore expect that buyers pay more than 0,00€ on average in the PWYW condition. We also expect payments in the PWYW to be positively correlated with participants' general WTP as assessed in the FP-condition and PWYW payments to be smaller than the WTP. Furthermore, we expect that buyers will refuse to buy an album more often in the PWYW condition because of the following rationale: In a fixed price condition, prices are generally assumed to be fixed by the seller and it is not of the buyers' concern whether the set price is appropriate for the product or not. In voluntary payments as in PWYW, on the contrary, the sole responsibility for determining the price is placed upon the buyer who will not only consider the subjective value of the product (how much they like the album) but also the objective value (e.g., the appropriate price for any music album) and the perspective of the seller (who, e.g., wants to make a living from selling). Thus, voluntary payments may signal a prosocial identity and buyers may tend to avoid to purchase at all when they feel that their WTP might be "too low," presumably in order to maintain their positive self-image (Gneezy et al., 2012).

By contrasting the two experimental conditions, we are able to isolate the rational aspects of the purchase decision from its social aspects. Two different approaches to the neural correlates of PWYW pricing are conceivable, given the literature. The first approach will focus on the rewarding properties of pro-social behavior. The fronto-mesolimbic reward network represents the key regions that encode social preferences. Following a reward-oriented approach to social preferences (Fehr and Camerer, 2007), it has been shown, that fair actions correlate with

greater response of reward related striatal areas in both the beneficiary (Tabibnia et al., 2008) and the donor (Moll et al., 2006). Furthermore, a study by King-Casas et al. (2005) suggests, that activity in the Nucleus accumbens (NAcc) could predict cooperation in a repeated trust game. Thus, when studying the neural correlates of purchase decisions in a PWYW paradigm, we should assume that paying a fair price triggers a response of the buyers reward system. The decision should then result from a trade-off between monetary and non-monetary (an therefore social or emotional) reward (see Kim et al., 2009). The dorsolateral prefrontal cortex (DLPFC) and the ventromedial prefrontal cortex (VMPFC) are likely to be crucially involved in this balancing of competing rewards (Fehr and Camerer, 2007). During the decision stage of the PWYW condition, we therefore expect a correlation between prices and activity in regions of the fronto-mesostriatal network, also activated in the FP condition. However, we expect a stronger activation in PWYW since in the case of high payments, there is an additional source of reward next to the product, namely the reward of having committed a pro-social act. During the listening stage, on the contrary, we expect a less pronounced correlation between prices paid and neural activity in fronto-mesostriatal areas, because in a PWYW situation, prices do not solely depend on product preference, but also on social concerns.

The second approach is less reward based: Although reward is arguably an important aspect of social decision making, a property that *exclusively* applies to social cognition is its relatedness to the intentions of others, namely Theory of Mind (ToM; see Amodio and Frith, 2006; Behrens et al., 2009; Young and Dungan, 2012). In the particular context of PWYW, it is vital that the buyer recognizes the sellers intentions, as pointed out by Regner and Barria (2009). Two adjoining regions have repeatedly been associated with tasks that are related to intentions and mental states of others: First, the right temporo-parietal junction (rTPJ; Saxe and Kanwisher, 2003; Saxe and Wexler, 2005), and second, the cortex areas of the superior temporal gyrus and sulcus, which we will refer to as the *STS-Region*. The STS-Region reacts to visually perceived social cues like eye-, head-, and hand-movements (Allison et al., 2000) but is also involved in moral cognition that require high amounts of cognitive control (Borg et al., 2006; Emonds et al., 2011). For the contrast of activity in PWYW and FP-condition *during the decision stage*, we therefore expect that compared to the FP-condition, the PWYW condition should trigger greater activity (a) in frontal regions, associated with social and non-social reward processing, (b) in regions associated with processing of intentions and mental states of others (ToM), like rTPJ and STS, and (c) in regions that respond to emotional content of stimuli, like amygdala and VMPFC.

## MATERIALS AND METHODS

### Participants

We tested healthy participants ( $N = 25$ , 12 female, 13 male, mean age  $M = 35.08$ ,  $SD = 17.71$ ), who gave written consent to participate in the study. All analyses were controlled for age

and sex. The study was approved by the Ethics Committee of the University Clinics Bonn, Germany (ethics statement: 276/11).

### Stimulus-Material

The stimulus-material consisted of music-albums that were downloaded from Bandcamp.com, an internet platform that allows bands to distribute their music in a variety of pricing formats including PWYW (which, on this site, is called "Name Your Own Price"). In the experiment, we only used albums that were made available by the artists under PWYW conditions or as a free download. The experiment therefore resembles the real life buying conditions for the offered products.

Albums were taken from the Bandcamp.com charts for different musical genres between December 2011 and January 2012. We included music from the following genres: rock, metal, hip hop, country/folk, indie, and pop (Berns et al., 2010). We used the most popular albums that were available via PWYW or free download, and featured at least five songs. Participants were informed that the albums might vary in length. We did not include albums without vocals, compilations with music of different artists, cover- or theme-albums (like Christmas-albums or soundtracks).

We downloaded 14 albums in each of the six genres. At the beginning of the scanning session, participants chose the three genres they liked best. This was done to insure that the general appeal of the songs to the participants would be relatively high. Within each of these genres, seven albums were randomly assigned to both the PWYW- and the FP-condition, which makes for a total of 42 buying decisions, 21 in each condition. All 42 trials were put into a random order. This procedure ensured that the treatment variable "buying condition" is independent of genre, liking or order of the songs.

We selected one song from each album that was played to the participants during the scanning session. This was always the first song, except when the first song was an intro, in which case we used the second one. During scanning, we only presented 30 s excerpts, that would ideally include parts of a verse and a chorus. Note, however, that in order to increase the variance of prices paid, we only allowed participants to bid on an entire album and not on single songs (so they had to infer their buying decision of an album by listening to one sample track).

We emphasized in the instructions, that participants would make real life buying decisions in the experiment and that the artists actually offered their music under PWYW-conditions and would receive the money participants paid. It was not suggested, however, that due to these reasons, there was a moral obligation to pay for the music. After completion of the study, all payments made by participants were transferred to the corresponding artists.

### Buying-Conditions

In the *PWYW-condition*, participants could obtain any album for 0.00€ or any price they chose, up to 10.00€. Whenever participants proposed to pay 0.00€, they were additionally asked, if they wanted to obtain the album for 0.00€ or not, to reduce the ambiguity of this response. In the *FP-condition*, the subjects' WTP

was determined via a classic Becker–DeGroot–Marschak (BDM)-auction (Becker et al., 1964). In this condition, every album had a price that was randomly fixed to any positive value up to 10.00€ and unknown to the participants (the only thing they knew about the price was, that it would never be 0.00€. Therefore, in the FP-condition, a bidding of 0.00€ meant unambiguously that the participant had no interest in obtaining this album). Participants could bid any price they wanted to pay for the album, however, they were informed, that they would only get the album, if their bidding was greater than the randomly fixed price. Participants were informed that in this case, they would buy the record for the randomly fixed price even if they have offered to pay more. Under these circumstances, the participants bidding determines the maximum value of the price he or she may have to pay, and his optimal strategy therefore is to bid his or her true value – the WTP – for any given album.

## Experimental Design

The fMRI-experiment consisted of 42 trials in which subjects had to decide how much to pay for the digital version of a music album (on the behavioral level, we thus studied a total of 1050 buying decisions, distributed over our  $N = 25$  participants). Each trial started with a listening stage in which a snippet from a representative song of a given album was played to the participants via headphones. To control for participants' prior knowledge of the artists and to avoid that the record was already owned by the participant, we obtained a set of records from Bandcamp.com, a website devoted to the distribution of professionally produced records from lesser known amateur bands. Each trial consisted of a listening-stage, a decision-stage and a response-stage. Since a large amount of motion related brain activity was to be expected during this stage, participants were instructed to complete their pricing decision before the beginning of the response-stage and we made no hypotheses about this stage. There were 21 trials in the PWYW- and 21 trials in the FP-condition, that were compared in a within-subject design, with the FP-condition serving as a control condition in which the subjects' WTP was determined for the present study. Both conditions were identical in every aspect, except for the consequences that the subjects' pricing-decisions had for their own and the sellers' pay-off. Participants were initially endowed with a budget of 10.00€, which they could spend fully in every trial. Participants were instructed that this was their money from now on that they were also free to keep the amount partially or entirely for themselves by not spending it all or by not making any purchase. At the end of the experiment, one trial was randomly selected and the transaction was completed depending on the participant's decision in this trial. By this approach, we made sure that each decision in each trial had the potential to result in a real consequence.

Every album was presented only once in the course of the experiment, to ensure a novelty aspect of the product. This means, however, that we can only compare average prices between the two buying conditions. Stimulus timing was 30 s for the listening stage, followed by a 3 s instruction slide indicating the condition (PWYW or Fixed-Price) and a 5 s time window that allowed participants to think about the price they were willing to

pay, which makes for a total of 8 s for the decision stage. This stage was followed by the input stage, in which participants were asked to enter the amount they had decided on (the stages of a prototypical trial are shown in **Table 1**). The input stage lasted as long as it took the participant to enter the price in each trial and thus served as a temporal jitter. A similar timing structure was used in Knutson et al. (2007).

## Image Acquisition

fMRI data was recorded on a 1.5T Scanner (Avanto, Siemens, Erlangen, Germany) with a standard 8 channel Siemens head coil. We collected about 800–1000 T2\*-weighted, gradient echo EPI-scans, depending on how much time participants needed to type in their prices. The following parameters were used: 31 slices per volume; slice thickness: 3 mm; inter-slice gap: 0.3 mm; matrix size:  $64 \times 64$ ; echo time: 45 ms; repetition time: 2500 ms; flip angle: 90°. Structural images were obtained by collecting 160 T1-weighted volumes (repetition time: 1660 ms; echo time: 3.09 ms; flip angel: 15°; slice thickness: 1 mm).

## Image Processing

Functional images were preprocessed using SPM8. Preprocessing included the following steps in the given order: (a) slice timing (b) realignment for motion correction (c) co-registration with the high resolution spatial images (d) spatial normalization using SPM's unified segmentation routine and (e) smoothing with a Gaussian spatial filter with 8 mm full width at half maximum.

Preprocessed data were analyzed using a general linear model, fitted using SPM8's canonical hemodynamic response function and a high pass filter of 128 seconds. Each stage of the experiment (listening, decision, and input) in each condition (PWYW and FP) was modeled as a separate regressor (i.e., six orthogonal regressors). We also included additional regressors parametrically modulating the listening- and decision-stage by the prices participants paid on these trials. The modulators were included to investigate linear dependencies between brain activity and prices paid. On each trial in either condition (PWYW and FP), participants had the option to indicate that they were not willing to pay any money at all. In the FP-condition, this would have inevitably revoked the chance to obtain the album at all. Because it cannot be ruled out that the decision to not buy an album is qualitatively different from paying even a little sum, we decided to exclude trials with 0.00€ payments and model them as separate regressors. In five participants (three men and two women), this led to a reduction of valid experimental trials by more than 50%. We therefore excluded these participants from the analysis of imaging data. In the PWYW-condition, however, it was possible to obtain an album for free by entering a price of 0€. To distinguish this situation from occasions where the participant rejected the album even if it was free, we interrogated participants each time they entered 0.00€ if they wanted to obtain the album for free. This, however, did only occur five times across all participants and trials. All other trials, in which participants entered 0.00€ were modeled with separate regressors as in the FP-condition. Six motion parameters were added as regressors of no interest to the model to account for residual head motion not corrected during preprocessing.

**TABLE 1 | Stages of each trial.**

Stage	Listening		Decision		Response
	Duration	30 s	3 s	5 s	
Display/Stimulus	Album-Cover + Music	PWYW		PWYW Think about your price!	5.00€

The table shows a trial of the PWYW condition. Condition names were not abbreviated but spelled out during the actual experiment. In the FP-condition, the PWYW display was replaced by "Fixed Price."

Contrast estimates from the GLM analyses were submitted to a second-level analysis that treated subjects as random effects and modeled participants age and sex as covariates of no interest. Resulting statistical parametric maps were initially thresholded at  $p < 0.001$  and then corrected for the family-wise error at the cluster level to keep the probability of false-positive results beneath  $p < 0.05$  at the whole brain level.

## RESULTS

We will start with presenting behavioral results to demonstrate the effectiveness of our experimental manipulation and the consistency of our results with previous work. We will then present the neuroimaging results from the FP-condition that reflect neural correlates of a "traditional" ("rational") exchange situation. After this, we present the neuroimaging results regarding the PWYW condition, and differences in brain activation between the two buying conditions. We will conclude the results section by a meta-analysis to provide a solid ground for the interpretation of our main finding.

### Behavioral Results

The analyses of the behavioral data include the entire sample of 25 participants. The same results, however, were obtained for the reduced sample of 20 participants who were included in the imaging analyses. All analyses are based on four descriptive measures: The average prices paid for an album in the PWYW and FP-condition, their difference (priceFP – pricePWYW) and their ratio (pricePWYW/priceFP). These measures were computed in two ways: uncorrected measures comprise the price inputs of all trials. As not all trials actually involved the purchase of an album in all participants (like an input of "0.00€" in the FP-condition or input of 0.00€ and answering "no" to the question of whether they wanted to have the album for free in the PWYW condition), we corrected the measures for these "non-transactions" by excluding these trials from the analyses. Descriptive statistics for all four measures in their uncorrected and uncorrected versions are presented in Table 2.

In line with our hypotheses and previous behavioral findings, participants took advantage of the PWYW offer and paid less in this condition. However, prices paid in the PWYW condition were significantly higher than 0.00€ [ $t(24) = 11.85$ ,  $p < 0.001$  for corrected,  $t(24) = 9.43$ ,  $p < 0.001$  for uncorrected prices], replicating previous results that demonstrated the general feasibility of the PWYW pricing system. Even though prices paid in the FP and in the PWYW condition were highly correlated ( $r = 0.73$ ;  $p < 0.001$  for the corrected means of

**TABLE 2 | Descriptive statistics of corrected and uncorrected prices (means in €).**

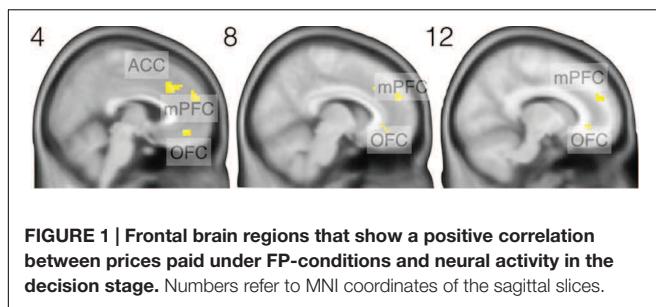
	Corrected	Uncorrected
Price PWYW (WTP <sub>PWYW</sub> )	3.10 ( $SD = 1.31$ )	2.37 ( $SD = 1.26$ )
Price FP (WTP)	3.49 ( $SD = 1.06$ )	2.96 ( $SD = 1.27$ )
Difference FP-PWYW	0.39 ( $SD = 0.89$ )	0.59 ( $SD = 0.86$ )
Quotient PWYW/FP	0.87 ( $SD = 0.29$ )	0.83 ( $SD = 0.43$ )

both buying conditions,  $r = 0.77$ ,  $p < 0.001$  for uncorrected means) reflecting that participants were guided by their general WTP when determining how much they would pay in the PWYW condition (Kim et al., 2009), there was still a significant difference in the amounts participants paid in the two conditions [ $t(24) = -2.15$ ;  $p = 0.042$  for the corrected,  $t(24) = -3.41$ ;  $p = 0.002$  for the uncorrected measures]. In line with results from Gneezy et al. (2012), participants in our study refused to buy an album more often in the PWYW condition (that is, out of 525 trials in each condition, participants decided 128 times that they would not buy an album, not even for a price of 0.00€, in the PWYW condition, but only refrained from a purchase 91 times in the FP condition;  $p < 0.01$ ;  $\chi^2 = 7.90$ ;  $df = 1$ ;  $N = 1050$ ). This could be interpreted as a tendency to rather not buy the record before paying a price that might be "too low." In sum, behavioral results indicate that participants distinguished between the two conditions as expected from the previous literature, and adjusted their decisions accordingly (Kim et al., 2009; Regner and Barria, 2009; Riener and Traxler, 2012). PWYW payment was the only variable that showed associations with sex and age: Men payed higher prices than women, but only for the corrected prices [ $M = 3.66$ ,  $SD = 1.18$  vs.  $M = 2.49$ ,  $SD = 1.20$ ,  $t(23) = 2.454$ ,  $p = 0.022$  for corrected prices] and age was positively correlated with the corrected PWYW prices (non-parametric  $r = 0.579$ ,  $p = 0.002$  for corrected and non-parametric  $r = 0.44$ ,  $p = 0.028$  for uncorrected means). Controlling for age and sex, however, did not affect the behavioral results.

### Imaging Results

#### Fixed-Price

Our first analysis focused on the parametrically modulated regressor of the *decision stage*, to investigate whether participants' WTP was reflected in neural activity at this stage. Figure 1 shows the statistical parametric map of the respective second level analysis: Neural activity in three clusters correlated positively with the prices paid in the fixed price condition: One cluster in



the orbitofrontal cortex (OFC; peak coordinate  $x = 0, y = 41, z = -11, Z = 4.35, p < 0.05$ , corrected,  $k = 40$  voxels), one cluster in medial prefrontal cortex (mPFC;  $x = 9, y = 53, z = 22, Z = 4.00, p < 0.05$ , corrected)  $k = 38$  voxels, and one cluster in the anterior cingulate (ACC;  $x = 0, y = 29, z = 31, Z = 3.78, p < 0.05$ , corrected  $k = 55$  voxels). These results indicate that brain regions implicated in reward processing encode participants' WTP. Corresponding results can be found in Knutson et al. (2007) and Plassmann et al. (2007).

The participants' WTP should be related to their preference for the product they are about to purchase. It is likely that this product preference emerges during the *listening stage* while the participant makes first contact with the music. We therefore investigated next, whether a similar linear relationship between WTP and neural activity existed during the listening stage. No significant results were obtained at the whole brain level. A more focused search around the peak coordinates of the significant clusters from the previous analysis on the linear dependencies between participants' WTP and neural activity during the decision stage, however, revealed a similar relationship between WTP and neural activity in the orbitofrontal cortex during the listening stage as well ( $Z = 4.25, p < 0.05$ , small-volume corrected in a 10-mm-radius sphere around  $x = 0, y = 41, z = -11$ ). This result is in line with previous findings of a response of the OFC to product-attractiveness (Erk et al., 2002; Plassmann et al., 2007). No results were obtained when small-volume-corrected searches were conducted around the clusters in the medial frontal cortex and the ACC.

### Pay-What-You-Want

Our analysis of neural activity during the different stages in the FP-condition has revealed a linear relationship between participants' WTP and neural activity in brain areas involved in reward-processing. Next, we investigated whether a similar relationship existed in the PWYW condition. Focusing on the linear relationship between prices paid and neural activity in the *decision-stage*, we were not able to find a significant relationship in any brain regions, neither at the whole brain level, nor when focusing on the peak coordinates found in the fixed price condition, nor when lowering the threshold to  $p < 0.001$ , uncorrected. Similarly, no relationship between prices paid and neural activity was detected during the *listening stage* of the PWYW condition. While neural activity in reward-related brain areas appears to be predictive for the WTP

under traditional fixed-price exchange conditions, this type of relationship does not seem to exist under PWYW conditions. This is remarkable, since the two conditions did only differ with respect to the pricing mechanism, and our finding that prices paid in both conditions were correlated, indicates that participants' pricing decisions in the PWYW condition were not random.

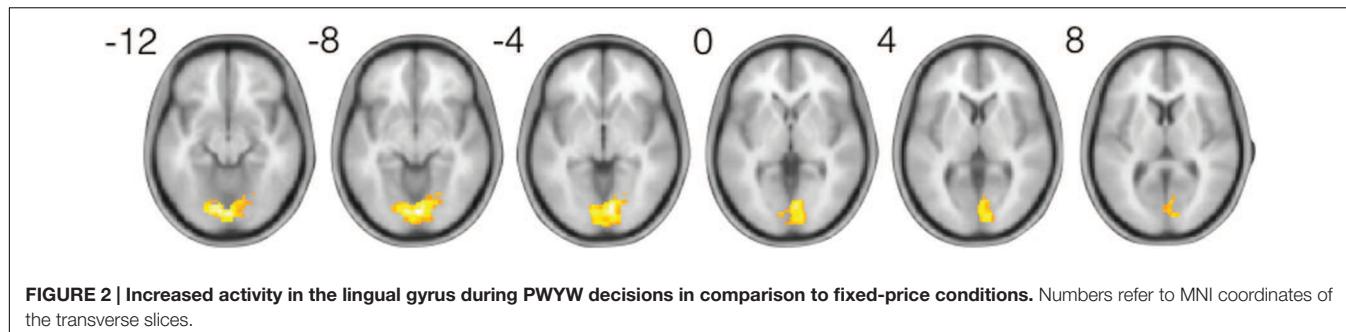
We next investigated *differences* in neural activity between the two conditions by directly contrasting the two pricing conditions during the decision stage, the time when participants were first confronted with the pricing context in this experimental trial. Note that this contrast did not make use of parametrically modulated regressors. At this stage, we found increased neural activity when participants were confronted with the PWYW condition, compared to the FP context, in the occipital lobe, peaking in the lingual gyrus (see Figure 2,  $x = 3, y = -85, z = -8, Z = 4.85, k = 389$  voxels,  $p < 0.05$ , corrected). No results were obtained for the reverse contrast (FP > PWYW) at the selected threshold.

### Meta-Analysis

In contrast to reward-, risk-, and higher cognition-related brain areas, the lingual gyrus has not received much attention in the neuroeconomic literature. Together with the behavioral results, our experimental design, however, suggests an implication of this brain area in PWYW decision making. We conducted an automated meta-analysis within the NeuroSynth-framework (Yarkoni et al., 2011) to obtain a quantitative reverse inference on the peak activation coordinate in the lingual gyrus. We queried the NeuroSynth database<sup>2</sup> that encompassed 413,429 activation coordinates from 11,406 studies in August 2015 to obtain information on the probability that studies contained a certain search term given activation at this specific location. The posterior probability measure from the NeuroSynth database is a measure for selective activation of a brain region and can allow for inferences on psychological states from brain imaging results. As the posterior probabilities derived from NeuroSynth are not corrected for uncertainty, we report only results with a significant  $z$ -statistic.

While all main associations were reported for visual processing ( $z = 4.57$ , posterior probability 0.68), visual attention ( $z = 4.11$ , posterior probability 0.82), or simply anatomical location, one search result suggested an implication of this region in emotional information processing ( $z = 4.1$ , posterior probability 0.84). In a next step, we queried the database for two additional coordinates that were local activation maxima within the activated cluster obtained from the PWYW > FP contrast. According to NeuroSynth, the local maximum at MNI ( $-9, -82, -11$ ) is associated with the terms "memory" ( $z = 3.55$ , posterior probability 0.64), "autobiographical" ( $z = 4.84$ , posterior probability 0.83), and "retrieved" ( $z = 4.03$ , posterior probability 0.83). The second local maximum at MNI ( $9, -88, 4$ ) was only associated with search terms related to visual processing.

<sup>2</sup>[www.neurosynth.org](http://www.neurosynth.org)



## DISCUSSION

The present study's experimental paradigm was designed to elucidate the relationship between neural activity and participants' willingness-to-pay under two different pricing regimes: a FP-condition resembling a traditional exchange between a seller and a consumer, and a PWYW condition that resembled the fixed-priced condition in every way, with the only exemption that in the PWYW condition, consumers were given the option to pay any price they wanted. On the behavioral level, we replicated previous findings on the feasibility of the PWYW approach: Even though participants decided to pay significantly less when the pricing decision was in their hands, they still offered to pay amounts significantly greater than zero. Also, as in the study of Gneezy et al. (2012), participants refused to buy an album more often in the PWYW than in the FP-condition. On the neural level, we found supporting evidence for our hypothesis on the relationship between mesolimbic-frontal activity and willingness-to-pay. Such a relationship, however, was only present in the fixed price condition. In the following, we will discuss this finding and the absence of such a relationship in the PWYW condition and seek possible explanations for apparent differences in neural activation between the two pricing conditions.

### FP-Condition

During the *decision stage*, we found three areas in which activity was positively correlated with prices paid on a trial basis: the mPFC, the OFC, and the ACC. All of these areas are known candidates for higher cognitive function and decision making in economic contexts. Our finding in the MFC is in line with results from a study of Knutson et al. (2007; -4, 59, -3 and 4, 46, -6), who found this region to respond to price information and to be more active in cases in which participants found the price to be appropriate and purchased the presented product.

For our result from the OFC, we find corresponding evidence in Plassmann et al. (2007) who could show, that the medial OFC (as well as the right DLPFC) correlated with participants' WTP (similar results were obtained by Erk et al., 2002). While Plassmann et al. (2007) presented primary rewards such as food stimuli to hungry participants, our results show that the mOFC also reacts to other rewards such as music. Again, this is in line with the results from Erk et al. (2002) who used pictures of more or less attractive cars as stimulus material. The ACC has also been implicated in decision making, especially with respect

to action selection, as discussed by Rushworth et al. (2007). The ACC and its interconnectivity with the mPFC has also been positioned in a framework of evaluation, appraisal, and conflict-resolution (Etkin et al., 2011). Our present design does not allow to disentangle reward- and conflict-based accounts. This will be an interesting endeavor for future research.

During the *listening stage*, we observed a similar relationship between neural activity and WTP at the same location in OFC as during the decision stage. Previous research on the neuronal response to music has primarily focused on a different structure during music reception by showing that a positive response to music correlates with higher activity in the Striatum. However, these studies were either based on the presentation of reported favorites of the participants (Blood and Zatorre, 2001; Montag et al., 2011; Salimpoor et al., 2011) or contrasted "pleasant" with heavily dissonant music (Koelsch et al., 2006). In contrast, our study made use of musical material that was previously unknown to the participants and also generally pleasing since it consisted of the top albums of the Bandcamp.com charts for each of our musical categories.

We found a correlation between activity in the OFC during music presentation with the prices later paid. Especially, since this ROI corresponds precisely to the cluster that also correlates with the price during the decision stage, we should assume that the OFC is involved in product (music) valuation. This is consistent with the repeated findings of striatal activity in response to pleasant music, due to strong anatomic connections between striatal areas and the OFC (Plassmann et al., 2007). We should note that this correlation has predictive value, since the participants in our study were only informed about the condition under which they had to decide their price *after* the listening stage. Further, because this correlation is calculated by a parametric modulated regressor, our result is sensitive to the shared variance of price and neural activity on the individual level.

### PWYW Condition

Even though the behavioral data on PWYW decisions was in line with the previous literature, we found no correlations between BOLD-signal and prices paid in the PWYW condition. Even though null-findings are difficult to interpret, we can conclude that the straight forward translation of product preference into prices that we found in the FP-condition does not exist in the PWYW condition in the same way. The high correlation between

the mean prices in the two buying conditions across participants suggest some systematic behind the pricing decisions, which should be reflected in neural activation data. A possible explanation for the apparent null finding might therefore be a higher degree of between-subject variability in the decision making mechanisms under PWYW-conditions that preclude robust activations at the group level. A possible avenue for future studies might therefore be the application of multivoxel pattern classification analyses that have been shown to decode signals from sub-threshold activation data on the single-subject level (Riggall and Postle, 2012).

Instead of a correlation between neural activity and prices paid, or significant activation in Theory of Mind related areas like the STS-Region, we found an unexpected, but highly robust contrast of activity between the two buying conditions in the Lingual Gyrus during the decision stage. How can we explain this result? The lingual gyrus is part of the secondary visual cortex. At this point, it should be stressed once more, that our two experimental conditions only differed in respect to which condition-name was presented and in which options this resulted for the participants pricing decision. The condition-name therefore served as a visual cue for different sets of options to be considered in determining the price, and considering this, it is not surprising, that visual discrimination plays a crucial role in our setting. Hence, we find that in both conditions, activity of the occipital cortex is higher than baseline. This effect is significantly greater for the PWYW condition and comprises a bilateral cluster of 389 voxels (see **Figure 2**,  $x = 3$ ,  $y = -85$ ,  $z = -8$ ,  $Z = 4.85$ ,  $k = 389$  voxels,  $p < 0.05$ , corrected). It is unlikely that the activation difference is a merely perceptual in response to the cue. We will argue that the activation difference reflects an affective response to the indication of different pricing schemes. Previous findings concerning the lingual gyrus can support this interpretation.

Adolphs (2002) has argued that the visual cortex plays a role in the early processing of emotional stimuli and the lingual gyrus has repeatedly been associated with reaction to emotionally relevant stimuli (Taylor et al., 1998; Critchley et al., 2000b; Moll et al., 2002; Rilling et al., 2008; Fusar-Poli et al., 2009; Premkumar et al., 2012).

A meta-analysis of fMRI studies (Fusar-Poli et al., 2009) was able to show that areas of the visual cortex like lingual, inferior occipital and fusiform gyrus react to the emotional expression of faces. Furthermore, they showed, that the lingual gyrus reacted more strongly to sad than to neutral faces and especially played a role in the implicit processing of facial expressions. In a study, in which participants had to discriminate between scenes showing social acceptance or rejection, Premkumar et al. (2012) found, that the lingual gyrus reacted more strongly to rejection, which is in line with findings of Rilling et al. (2008). The authors conclude that the lingual gyrus plays a role in discriminating different qualities of social information. In addition, participants in this study, who *reported* lower emotional arousal in response to the scenes displayed, showed a more pronounced contrast in the lingual gyrus than others, suggesting a connection of this area to an important peripheral marker of emotional experience. The possible involvement of the lingual gyrus in

emotional information processing is further corroborated by our own NeuroSynth analysis which showed a possible association with our peak activation location and affective processing. It should be noted, that the point of peak activation of the lingual gyrus within the 8 s of the decision-stage is unknown. A minimal interpretation of our findings could be, that the norm-related PWYW condition requires a “special attention,” because decisions under this condition may be relevant for emotional regulation, which is not the case under FP conditions.

We have to ask, however, why other areas associated with emotional processing, like the amygdala and VMPFC, don't show up in our contrast of PWYW and FP-condition. Visual cortex and STS-region are both connected to these areas, and especially the amygdala is believed to influence processing in the visual cortex and the STS-region via feedback projections (Allison et al., 2000; Critchley et al., 2000a). Our findings in the visual cortex may support the assumption of early discrimination between stimuli with and without social components, possibly before they are assigned an emotional valence. The absence of any straightforward correlations between BOLD-signal in the PWYW condition, and the amount of money paid to the seller could suggest that after an early discrimination between the two buying conditions, participants used different strategies to determine their prices in the PWYW condition. We should still assume that these are influenced by the early discrimination.

Finally, the possibilities that our PWYW paradigm offers for future research on social and economic decision making should be noted. The paradigm implements two frames, in which decisions can take place, one organized primarily by rational considerations (the explicit rules of the market), and one that additionally includes the consideration of implicit social norms. The paradigm manages to implement this contrast while maintaining a degree of external validity, which is unusually high for fMRI experiments: A usual purchase of music via Bandcamp.com or other internet platforms does in fact happen via a screen (which is not the case for chocolate bars) and money is (or is not) paid to a real artist, who is only present in form of the music and cover information on the platform. Our design offers a contrast of two conditions which differ only in TOM aspects, except for one cue about the pricing condition. Potential uses of this paradigm should be investigated further: First, cross-modality validation, for example by replacing the visual cue about the buying condition with an auditory one, should yield insight in the role of the lingual gyrus. A combination of fMRI and measurements of electrodermal activity as in Critchley et al. (2000b), could also be considered, to gain insights into arousal processes and the embodiment of decision process as proposed in the somatic marker hypothesis. In the present study, we did not obtain independent measures of preference for each song. This could be helpful, however, to achieve clearer results on neuronal activity in the PWYW condition. Also, it should be investigated, how exactly the “socialness” of the PWYW condition is perceived, for example by distinguishing two conditions in which the degree to which social information is made salient varies. Personality should be taken into account as well to account for between-participant variation in WTP under both conditions. Interesting traits would be those that either relate to social behavior such

as “cooperativeness” (Cloninger et al., 1993) or “openness to experience” (Costa and McCrae, 1992) which could underly the willingness to explore unusual or innovative ideas such as PWYW system.

To maintain statistical power for the comparison of neuronal activity in different subgroups, a bigger sample and/or more experimental trials would be needed. A point that warrants discussion is the use of an BDM-auction in the fixed-price conditions. While this approach has been popularized by auction enterprises such as ebay.com, it differs from the usual experience of buyers in a traditional shop where tags provide information about prices. The BDM-auction, however, is usually used in neuroeconomic research to obtain a direct and bias free measure for participants’ WTP (Weber et al., 2007) which was the desired goal in our experiment. Another concern in our study is the amount of trials in both conditions. We tried to include as many trials as possible to obtain sufficient statistical power but at the same time tried to limit the number of trials to maintain enough ecological validity as buyers in real exchange situations tend not to make too many consecutive buying decisions. During analysis, we excluded trials in which no transaction took place as they might have been qualitatively different from actual transactions. Even though, we excluded participants with too many dropped trials, we cannot rule out entirely that the null result regarding the correlation between neural activity and prices payed in the PWYW condition resulted from a reduced statistical power due to unequal trial numbers in both conditions. Trials in the PWYW condition were excluded when participants bid 0.00€ and indicated that they did not want to obtain the album for free in response to a subsequent question. While this was effective to distinguish between qualitatively different decisions that both resulted in a 0.00€ bid, future studies may want to make use of a

## REFERENCES

- Adolphs, R. (2002). Neural systems for recognizing emotion. *Curr. Opin. Neurobiol.* 12, 169–177. doi: 10.1016/S0959-4388(02)00301-X
- Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* 4, 267–278. doi: 10.1016/S1364-6613(00)01501-1
- Amodio, D. M., and Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277. doi: 10.1038/nrn1884
- Becker, G. M., DeGroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behav. Sci.* 9, 226–232. doi: 10.1002/bbs.3830090304
- Behrens, T. E. J., Hunt, L. T., and Rushworth, M. F. S. (2009). The computation of social behavior. *Science* 324, 1160–1164. doi: 10.1126/science.1169694
- Belsky, L., Kahr, B., Berkelhammer, M., and Benkler, Y. (2010). *Everything in Its Right Place: Social Cooperation and Artist Compensation*, 17 Mich. Telecomm. Tech. L. Rev. 1. Available at: <http://www.mttr.org/volseventeen/belsky.pdf>
- Benkler, Y. (2011). *Voluntary Payment Models in Rethinking Music: A Briefing Book*. Available at: [http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/Rethinking\\_Music\\_Briefing\\_Book\\_April-25-2011.pdf](http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/Rethinking_Music_Briefing_Book_April-25-2011.pdf)
- Berns, G. S., Capra, C. M., Moore, S., and Noussair, C. (2010). Neural mechanisms of the influence of popularity on adolescent ratings of music. *Neuroimage* 49, 2687–2696. doi: 10.1016/j.neuroimage.2009.10.070
- Blood, A. J., and Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11818–11823. doi: 10.1073/pnas.191355898
- Borg, J. S., Hynes, C., Van Horn, J., Grafton, S., and Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: an fMRI investigation. *J. Cogn. Neurosci.* 18, 803–817. doi: 10.1162/jocn.2006.18.5.803
- Cloninger, C. R., Svrakic, D. M., and Przybeck, T. R. (1993). A psychobiological model of temperament and character. *Arch. Gen. Psychiatry* 50, 975–990. doi: 10.1001/archpsyc.1993.01820240059008
- Costa, P. T., and McCrae, R. R. (1992). Normal personality assessment in clinical practice: the NEO Personality Inventory. *Psychol. Assess.* 4, 5–13. doi: 10.1037/1040-3590.4.1.5
- Critchley, H. D., Daly, E. M., Bullmore, E. T., Williams, S. C. R., Van Amelsvoort, T., Robertson, D. M., et al. (2000a). The functional neuroanatomy of social behavior: changes in cerebral blood flow when people with autistic disorder process facial expressions. *Brain* 123, 2203–2212. doi: 10.1093/brain/123.11.2203
- Critchley, H. D., Elliott, R., Mathias, C. J., and Dolan, R. J. (2000b). Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study. *J. Neurosci.* 20, 3033–3040.
- Emonds, G., Declerck, C. H., Boone, C., Vandervliet, J. M., and Parizel, P. M. (2011). Comparing the neural basis of decision making in social dilemmas of people with different social value orientations, a fMRI study. *J. Neurosci. Psychol. Econ.* 4, 11–24. doi: 10.1037/a0020151
- Erk, S., Spitzer, M., Wunderlich, A. P., Galley, L., and Walter, H. (2002). Cultural objects modulate reward circuitry. *Neuroreport* 13, 2499–2503. doi: 10.1097/00001756-200212200-00024
- “reject” button that allows buyers to reject undesired albums right away.
- To our knowledge the present study is the first neuroeconomic imaging study to investigate neural correlates of the PWYW pricing system. PWYW has recently received a lot of attention in the behavioral literature, presumably sparked by reports on its surprising success in the marketing of music albums (Benkler, 2011). Alongside first evidence on neural underpinnings of PWYW buying decisions, we present an experimental design that allows the study of different pricing mechanisms in a neuroimaging setting. We would like to encourage further studies on this topic, for instance with machine learning techniques that aim at the prediction of PWYW decisions from multivariate patterns in neural activity. In general, our paradigm is also suitable for the study of other products than music albums. A replication of the present finding using different products/stimuli would provide further valuable insights.

## AUTHOR CONTRIBUTIONS

SW, SM, MR, CM, and BW designed the research; SW and SM conducted the experiment, analyzed the data, and wrote the manuscript; VK, PT provided protocols and technical advise; MR, CM, BW, VK, and PT edited the manuscript.

## ACKNOWLEDGMENT

CM is supported by a Heisenberg-grant awarded to him by the Deutsche Forschungsgemeinschaft (German Research Foundation; DFG MO 2363/3-1).

- Etkin, A., Egner, T., and Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn. Sci.* 15, 85–93. doi: 10.1016/j.tics.2010.11.004
- Fehr, E., and Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* 11, 419–427. doi: 10.1016/j.tics.2007.09.002
- Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., et al. (2009). Functional atlas of emotional faces processing: a voxelbased meta-analysis of 105 functional magnetic resonance imaging studies. *J. Psychiatry Neurosci.* 34, 418–432.
- Gneezy, A., Gneezy, U., Riener, G., and Nelson, L. D. (2012). Pay-what-youwant, identity, and self-signaling in markets. *Proc. Natl. Acad. Sci. U.S.A.* 109, 7236–7240. doi: 10.1073/pnas.1120893109
- Heyman, J., and Ariely, D. (2004). Effort for payment: a tale of two markets. *Psychol. Sci.* 15, 787–793. doi: 10.1111/j.0956-7976.2004.00757.x
- Kim, J., Natter, M., and Spann, M. (2009). Pay-What-You-Want – a new participative pricing mechanism. *J. Mark.* 73, 44–58. doi: 10.1509/jmkg.73.1.44
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., and Montague, P. R. (2005). Getting to know you: reputation and trust in a twoperson economic exchange. *Science* 308, 78–83. doi: 10.1126/science.1108062
- Knutson, B., Rick, S., Wimmer, G. E., Prelec, D., and Loewenstein, G. (2007). Neural predictors of purchases. *Neuron* 53, 147–156. doi: 10.1016/j.neuron.2006.11.010
- Koelsch, S., Fritz, T., von Cramon, D. Y., Müller, K., and Friederici, A. D. (2006). Investigating emotion with music: an fMRI study. *Hum. Brain Mapp.* 27, 239–250. doi: 10.1002/hbm.20180
- Moll, J., de Oliveira-Souza, R., Bramati, I. E., and Grafman, J. (2002). Functional networks in emotional moral and nonmoral social judgments. *Neuroimage* 16, 696–703. doi: 10.1006/nimg.2002.1118
- Moll, J., Krueger, F., Zahn, R., Pardini, M., Oliveira-Souza, R., and Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15623–15628. doi: 10.1073/pnas.0604475103
- Montag, C., Reuter, M., and Axmacher, N. (2011). How one's favorite song activates the reward circuitry of the brain: personality matters! *Behav. Brain Res.* 225, 511–514. doi: 10.1016/j.bbr.2011.08.012
- Persky, J. (1995). Retrospectives: the ethology of homo economicus. *J. Econ. Perspect.* 9, 221–231. doi: 10.1257/jep.9.2.221
- Plassmann, H., O'Doherty, J., and Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.* 27, 9984–9988. doi: 10.1523/JNEUROSCI.2131-07.2007
- Premkumar, P., Ettinger, U., Inchley-Mort, S., Sumich, A., Williams, S. C. R., Kuipers, E., et al. (2012). Neural processing of social rejection: the role of schizotypal personality traits. *Hum. Brain Mapp.* 33, 695–706. doi: 10.1002/hbm.21243
- Regner, T., and Barria, J. A. (2009). Do consumers pay voluntarily? The case of online music. *J. Econ. Behav. Organ.* 71, 395–406. doi: 10.1016/j.jebo.2009.04.001
- Riener, G., and Traxler, C. (2012). Norms, moods, and free lunch: longitudinal evidence on payments from a Pay-What-You-Want restaurant. *J. Socio Econ.* 41, 476–483. doi: 10.1016/j.socloc.2011.07.003
- Riggall, A. C., and Postle, B. R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *J. Neurosci.* 32, 12990–12998. doi: 10.1523/JNEUROSCI.1892-12.2012
- Rilling, J. K., Goldsmith, D. R., Glenn, A. L., Jairam, M. R., Elfenbein, H. A., Dagenais, J. E., et al. (2008). The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia* 46, 1256–1266. doi: 10.1016/j.neuropsychologia.2007.11.033
- Rushworth, M. F. S., Behrens, T. E. J., Rudebeck, P. H., and Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behavior. *Trends Cogn. Sci.* 11, 168–176. doi: 10.1016/j.tics.2007.01.004
- Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., and Zatorre, R. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat. Neurosci.* 14, 257–262. doi: 10.1038/nn.2726
- Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people: fMRI investigations of theory of mind. *Neuroimage* 19, 1835–1842. doi: 10.1016/S1053-8119(03)00230-1
- Saxe, R., and Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia* 43, 1391–1399. doi: 10.1016/j.neuropsychologia.2005.02.013
- Tabibnia, G., Satpute, A. B., and Lieberman, M. D. (2008). The sunny side of fairness: preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychol. Sci.* 19, 339–347. doi: 10.1111/j.1467-9280.2008.02091.x
- Taylor, S. F., Liberzon, I., Fig, L. M., Decker, L. R., Minoshima, S., and Koeppe, R. A. (1998). The effect of emotional content on visual recognition memory: a PET activation study. *Neuroimage* 8, 188–197. doi: 10.1006/nimg.1998.0356
- Weber, B., Aholt, A., Neuhaus, C., Trautner, P., Elger, C. E., and Teichert, T. (2007). Neural evidence for reference-dependence in real-market-transactions. *Neuroimage* 35, 441–447. doi: 10.1016/j.neuroimage.2006.11.034
- Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* 8, 665–670. doi: 10.1038/nmeth.1635
- Young, L., and Dungan, J. (2012). Where in the brain is morality? Everywhere and maybe nowhere. *Soc. Neurosci.* 7, 1–10. doi: 10.1080/17470919.2011.569146

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Waskow, Markett, Montag, Weber, Trautner, Kramarz and Reuter. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Commentary: Fairness is intuitive

**Kristian O. R. Myrseth**<sup>1\*</sup> and **Conny E. Wollbrant**<sup>2\*</sup>

<sup>1</sup> School of Management, University of St Andrews, St Andrews, Scotland, <sup>2</sup> Department of Economics, School of Business, Economics and Law, University of Gothenburg, Gothenburg, Sweden

**Keywords:** fairness, self-control, intuition, decision times, dictator game

## A commentary on

### Fairness is intuitive

Cappelen, A. W., Nielsen, U. H., Tungodden, B., Tyran, J.-R., and Wengström, E. (2015). *Exp. Econ.* doi: 10.1007/s10683-015-9463-y. [Epub ahead of print].

Cappelen et al. (2015) open their paper, “Fairness is intuitive,” with the observation, “A key question in the social sciences is whether it is intuitive to behave in a fair manner or whether fair behavior requires active self-control” (p. 2). They purport to offer “evidence showing that fair behavior is intuitive to most people” (p. 1). Their premise is that deciding by intuition is faster than deciding by deliberation. While this premise in and on itself is rather uncontroversial—the conclusion that they draw from it is not: “Since a decision that relies on intuition is typically made faster than a decision that relies on deliberation, the response time of a fair decision relative to a selfish decision provides an important indication of the intuitiveness of fair behavior” (p. 2). This reasoning, in fact, amounts to a reverse inference fallacy<sup>1</sup>. “Intuitive” may mean “fast,” but this would not imply that “fast” means “intuitive.”

However, we may ask, under which empirical conditions might we be allowed to draw the inference of “intuitive” from “fast”? Naturally, these conditions would require that “fast” rule out “deliberative.” To achieve this, we would need information beyond relative response speed alone—such as absolute decision times. And this begs the question, which range of decision times would rule out “deliberative”—or at the very least, render it improbable? Although the precise cut-off for deliberative decisions may be difficult to establish (see e.g., Schneider and Shiffrin, 1977; Posner and Rothbart, 1998), it is clear that an individual, if given a few seconds, may have sufficient time to reflect consciously—and ample time, if given more than thirty. Responses made at those speeds ought thus not be taken as “intuitive” *prima facie*, on the basis of the response time data alone. Unfortunately, the authors make just this mistake.

Cappelen et al. (2015) find that “fair” decisions in a dictator game are faster than are “selfish” decisions, from which they infer that the fair decision is the more intuitive (e.g., Figure 2, p. 4). However, fair decisions took on average 38.4 s, and unfair decisions on average 48.5. It would seem, then, that both decision categories are fairly slow—and neither would appear unlikely to be characterized by deliberative processes. We may speculate about sources of the difference in mean response times, but intuitive as opposed to deliberative decision making is but one out of multiple possible explanations. Another explanation, for example, could be differences in degrees of deliberation. That is, individuals who deliberated more extensively might have reached a selfish decision, whereas individuals who deliberated less—but who did deliberate nonetheless—might have arrived at a fair choice. It is even possible, in this scenario, that the impulsive response is selfish—as some prior literature has suggested (e.g., Martinsson et al., 2012; Achtziger et al., 2015). The spontaneous response may then have been overruled by controlled deliberation, which might

## OPEN ACCESS

### Edited by:

Nikolaos Georgantzis,  
University of Reading, UK

### Reviewed by:

Daniel Västfjall,  
Linköping University, Sweden  
Carlos Alos-Ferrer,  
University of Cologne, Germany

### \*Correspondence:

Kristian O. R. Myrseth  
kom@st-andrews.ac.uk;  
Conny E. Wollbrant  
conny.wollbrant@economics.gu.se

### Specialty section:

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
*Frontiers in Psychology*

**Received:** 14 February 2016

**Accepted:** 19 April 2016

**Published:** 09 May 2016

### Citation:

Myrseth KOR and Wollbrant CE  
(2016) Commentary: Fairness is  
intuitive. *Front. Psychol.* 7:654.  
doi: 10.3389/fpsyg.2016.00654

<sup>1</sup> Another term for “reverse inference fallacy,” is the “fallacy of affirming the consequent”, as defined by Dowden (2016).

have been overturned yet again by even more extensive deliberation. In other words, individuals might have experienced an initial proclivity, changed their mind, and then changed their mind once again. As this possible scenario shows, it would be very difficult, to assign “fair” as opposed to “selfish” responses to intuition over deliberation.

Although Cappelen et al. (2015) make the valuable point of distinguishing conceptually between actual decision time and overall measured response time—which encompasses also reading time and decision implementation time—their distinction does not salvage their conclusion. Indeed, their measured response times include the time spent on reading and comprehending the instructions, but any such activity—by its very nature—would require some degree of deliberation. Therefore, it would not be possible subsequently—on the basis of relative response times alone—to distinguish between intuitive and deliberative decision processes<sup>2</sup>. A very fast decision, for example, may be the product of deliberation during the preceding reading and comprehension steps.

Cappelen et al. (2015) build on the work by Rand et al. (2012, 2014), who fall into similar traps. Rand et al. (2012, 2014) argue that time-pressure promotes “cooperation,” and that this amounts to evidence for the notion that cooperation is intuitive<sup>3</sup>.

<sup>2</sup>A reviewer pointed out that our argument could be construed as a blanket dismissal of the utility of response times as a process measure. We would stress, however, that our argument applies to empirical and theoretical contexts similar to that of the target paper, and we recognize that response time measurement has its uses. Examples of insightful application of response times in economic decision-making include Rubinstein (2007) and Achtziger and Alós-Ferrer (2014).

<sup>3</sup>As Cappelen et al. (2015) note, the empirical stability of the pattern obtained by Rand et al. (2012, 2014) is contested. Tinghög et al. (2013) and Verkoeijen and Bouwmeester (2014) fail to reproduce the pattern. Moreover, Tinghög et al. (2013) and Recalde et al. (2015) argue that the original pattern may have arisen from analytical and methodological artifacts, respectively.

## REFERENCES

- Achtziger, A., and Alós-Ferrer, C. (2014). Fast or rational? A response-times study of Bayesian updating. *Manage. Sci.* 60, 923–938. doi: 10.1287/mnsc.2013.1793
- Achtziger, A., Alós-Ferrer, C., and Wagner, A. (2015). Money, depletion, and prosociality in the dictator game. *J. Neurosci. Psychol. Econ.* 8, 1–14. doi: 10.1037/npe0000031
- Cappelen, A. W., Nielsen, U. H., Tungodden, B., Tyran, J.-R., and Wengström, E. (2015). Fairness is intuitive. *Exp. Econ.* doi: 10.1007/s10683-015-9463-y. [Epub ahead of print].
- Dowden, B. (2016). *Fallacies. The Internet Encyclopedia of Philosophy*. Available online at: <http://www.iep.utm.edu/> (March 27, 2016).
- Martinsson, P., Myrseth, K. O. R., and Wollbrant, C. (2012). Reconciling pro-social vs. selfish behavior: on the role of self-control. *Judgm. Decis. Mak.* 7, 304–315.
- Myrseth, K. O. R., and Wollbrant, C. (2015). *Intuitive cooperation Refuted: Commentary on Rand et al. (2012) and Rand et al. (2014)*. University of Gothenburg, Working Papers in Economics No. 617.
- Posner, M., and Rothbart, M. K. (1998). Attention, self-regulation and consciousness. *Philos. Trans. R. Soc. B* 353, 1915–1927. doi: 10.1098/rstb.1998.0344
- Rand, D. G., Greene, J. D., and Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature* 489, 427–430. doi: 10.1038/nature11467
- Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., et al. (2014). Social heuristics shape intuitive cooperation. *Nat. Commun.* 5:3677. doi: 10.1038/ncomms4677
- Recalde, M. P., Riedl, A., and Vesterlund, L. (2015). *Error Prone Inference from Response Time: The Case of Intuitive Generosity in Public-Good Games*. University of Pittsburgh, Department of Economics. Working Paper Series 15/004.
- Rubinstein, A. (2007). Instinctive and cognitive reasoning: a study of response times. *Econ. J.* 117, 1243–1259. doi: 10.1111/j.1468-0297.2007.02081.x
- Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. detection, search, and attention. *Psychol. Rev.* 84, 1–66. doi: 10.1037/0033-295X.84.1.1
- Tinghög, G., Andersson, D., Bonn, C., Böttiger, H., Josephson, C., Lundgren, G., et al. (2013). Intuition and cooperation reconsidered. *Nature* 498, E1–E2. doi: 10.1038/nature12194
- Verkoeijen, P. P. J. L., and Bouwmeester, S. (2014). “Does intuition cause cooperation?” *PLoS ONE* 9:e96654. doi: 10.1371/journal.pone.0096654

However, subjects in their time-pressure treatments had adequate time to deliberate—median response times were 6–13 s, across studies. As Myrseth and Wollbrant (2015) argue, this calls into question the meaning of the time-pressure treatments. Although Rand et al. (2012, 2014) also show that cooperation is negatively associated with response time, a closer examination of their data, in which average cooperation rates are plotted against response times, reveals that the pattern is non-linear and generally unclear (Myrseth and Wollbrant, 2015). In fact, when examined locally, there appears to be a positive association between response times and cooperation, among decisions made within 4 s<sup>4</sup>. A negative pattern emerges for slower decisions. The data from Rand et al. (2012, 2014) thus fail to provide meaningful evidence for the hypothesis that cooperation is intuitive rather than deliberative.

More generally, we would call for greater caution in the interpretation of response time data. Although often fast, intuition can also be slow, and, conversely for deliberation—although often slow, it can also be fast (within limits). It is therefore not straightforward to rely on response times—or on experimental time pressure treatments—to disentangle intuition from deliberation in economic decision making.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct, and intellectual contribution to the work, and approved it for publication.

## ACKNOWLEDGMENTS

We are grateful to the reviewers and the editor for helpful comments.

<sup>4</sup>This discussion refers to an analysis that pools the data from all studies in Rand et al. (2014). The pattern is largely consistent when only one-shot public good games are considered. For further details, see Myrseth and Wollbrant (2015).

Recalde, M. P., Riedl, A., and Vesterlund, L. (2015). *Error Prone Inference from Response Time: The Case of Intuitive Generosity in Public-Good Games*. University of Pittsburgh, Department of Economics. Working Paper Series 15/004.

Rubinstein, A. (2007). Instinctive and cognitive reasoning: a study of response times. *Econ. J.* 117, 1243–1259. doi: 10.1111/j.1468-0297.2007.02081.x

Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. detection, search, and attention. *Psychol. Rev.* 84, 1–66. doi: 10.1037/0033-295X.84.1.1

Tinghög, G., Andersson, D., Bonn, C., Böttiger, H., Josephson, C., Lundgren, G., et al. (2013). Intuition and cooperation reconsidered. *Nature* 498, E1–E2. doi: 10.1038/nature12194

Verkoeijen, P. P. J. L., and Bouwmeester, S. (2014). “Does intuition cause cooperation?” *PLoS ONE* 9:e96654. doi: 10.1371/journal.pone.0096654

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Myrseth and Wollbrant. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

