



Contents lists available at ScienceDirect

## Journal of Research in Personality

journal homepage: [www.elsevier.com/locate/jrp](http://www.elsevier.com/locate/jrp)

## Individual differences, judgment biases, and theory-of-mind: Deconstructing the intentional action side effect asymmetry

Edward T. Cokely<sup>a,\*</sup>, Adam Feltz<sup>b</sup><sup>a</sup>Max Planck Institute for Human Development, Center for Adaptive Behavior and Cognition, Königin - Luise Str. 5, 14195 Berlin, Germany<sup>b</sup>Department of Philosophy and Interdisciplinary Studies, Schreiner University, 2100 Memorial Blvd, Kerrville, TX 78028-5697, USA

## ARTICLE INFO

## Keywords:

Judgment

Decision making

Theory-of-mind

Knobe effect

Individual differences

Extraversion

Introversion

Personality

Biases

Working memory

Adaptive processes

Moral

Action theory

Experimental philosophy

## ABSTRACT

When the side effect of an action involves moral considerations (e.g. when a chairman's pursuit of profits harms the environment) it tends to influence theory-of-mind judgments. On average, bad side effects are judged intentional whereas good side effects are judged unintentional. In a series of two experiments, we examined the largely uninvestigated roles of individual differences in this judgment asymmetry. Experiment 1 indicated that extraversion accounted for variations in intentionality judgments, controlling for a range of other general individual differences (e.g. working memory, self-control). Experiment 2 indicated that extraversion's influence was partially mediated by more specific variations in intentional action concepts. A priming manipulation also provided causal evidence of judgment instability and bias. Results suggest that the intentional action judgment asymmetry is multiply determined, reflecting the interplay of individual differences and judgment biases. Implications and the roles of individual differences in judgment and decision-making research are discussed.

© 2008 Elsevier Inc. All rights reserved.

Authorship is equal. We thank Ainsley Mitchum, Alfred Mele, Joshua Knobe, Fiery Cushman, Shaun Nichols, Adrien Barton, and Nadine Fleischhut for their comments and suggestions. We also thank Colleen Kelley, Anders Ericsson, and the ABC Research Group for use of laboratory equipment and resources.

## 1. Introduction

Theory-of-mind is by many accounts a uniquely human capacity that is essential to modern social functioning and cultural development (Gallagher & Frith, 2003; Tomasello, Kruger, & Ratner, 1993). One fundamental and ubiquitous aspect of theory-of-mind involves the assessment of the intentionality of others' actions. Recent research on intentional action judgments has revealed that these assessments can be influenced by the moral valence of an event (Knobe, 2003a; Knobe & Burra, 2006; Leslie, Knobe, & Cohen, 2006; Nichols, 2004). Theoretically, a number of candidate mechanisms such as affective biases and multistage search processes have been proposed to account for this effect (Knobe, 2006). However, research has largely neglected potential contributions of individual differences—a tendency that is not uncommon in the broader context of psychology science (Cronbach, 1957; Funder, 1991, 1995, 2001; Rev-

elle, 1987). Indeed, several interesting variations such as cultural differences are known to affect general theory-of-mind processes (Lillard, 1997, 1998). What can individual differences reveal about the mechanisms of intentional action judgments?

## 1.1. The intentional action side effect asymmetry

Knobe (2003a, 2006) has demonstrated that some people's theory-of-mind judgments are influenced by the "goodness" or "badness" of the side effects of their actions. Side effects of actions are typically taken to be consequences of actions that are foreseen but not intended. People tend to make asymmetric judgments wherein bad side effects are judged as being brought about intentionally yet good side effects are judged as being brought about unintentionally. To illustrate, consider this variant of the paradigmatic cases that evoke the intentional action side effect asymmetry (Knobe, 2003a)<sup>1</sup>:

The vice-president of a company went to the chairman of the board and said, "We are thinking of starting a new program. It will help us increase profits for this year's balance sheet, but in 10 years it will start to (harm/help) the environment." The chairman answered, "I don't care at all about (harming/helping) the environ-

<sup>1</sup> These are the exact scenarios used in Experiments 1 and 2. Participants were asked to what degree they agree on a 7 point Likert scale (1 = disagree, 4 = neutral, 7 = agree) with one of the two following relevant statements: "The chairman intentionally harmed (or helped) the environment."

\* Corresponding author. Fax: +49 30 82 406 394.

E-mail address: [cokely@mpib-berlin.mpg.de](mailto:cokely@mpib-berlin.mpg.de) (E.T. Cokely).

ment. I just want to make as much profit for this year's balance sheet as I can. Let's start the new program." They started the new program. Sure enough, ten years later, the environment started to be (harmed/helped) (Cushman & Mele, 2008).

The only difference is whether the side effect of the chairman's action is 'good' or 'bad'. However, when participants are asked to indicate how much they agree with the statement "the chairman intentionally harmed/helped the environment," their judgments for help versus harm cases are qualitatively different (help is unintentional, harm is intentional). This variation has proven robust and is present in different cultures (Knobe & Burra, 2006), ages (Leslie, et al., 2006), and side effect scenarios (Cushman & Mele, 2008; Knobe, 2003a, 2003b, 2004a, 2004b, 2006; Mele & Cushman, 2007; Nichols & Ulatowski, 2007).

There are several theoretical accounts of the mechanisms of the side effect asymmetry (for a recent review see Feltz, 2007). For example, Knobe (2006) proposes a model with two sub-processes where the identification of a harmful or beneficial side effect triggers a selective search for features that are sufficient to judge the side effect as being brought about intentionally. Other evidence and theory suggests that it is not simply the harmfulness of the side effect that influences judgments but that social norms may also play a role. Scenarios that describe *regretfully* bringing about a harmful side effect can dramatically reduce the asymmetry (Phelan & Sarkissian, 2008). Alternatively, Nadelhoffer (2004) and Malle and Nelson (2003) argue that the asymmetry primarily results from an affective bias. Malle and Nelson suggest that because one has a negative affective reaction to the chairman in the harm scenario, one is more likely to think that the harm is brought about intentionally. Similarly, Nadelhoffer contends that the harm scenario is seen as intentional because it allows participants to blame the chairman more. However, Nadelhoffer further suggests that help judgments are also biased as participants do not want to praise the chairman for his help.

In addition to theoretical accounts that emphasize search and bias mechanisms, emerging research suggests that specific individual differences in intentional action intuitions may also play an influential role. For example, Nichols and Ulatowski (2007) suggest that the asymmetry may be the result of specific individual differences in interpreting the word 'intentionally'. Similarly, Cushman and Mele (2008) have found evidence for multiple folk concepts of intentional action.<sup>2</sup> One concept treats an action performed with desire as necessary for being done intentionally (i.e. desire-concept) and another concept treats an action performed with belief as sufficient for being done intentionally (belief-concept).

Given the number of theoretical interpretations and findings, we hypothesized that the side effect asymmetry was likely multiply determined, resulting from both individual differences and biases. In order to test this hypothesis, we conducted a series of two experiments. In the first experiment, we examined the relationship between the judgment asymmetry and a personality trait (i.e. extraversion) which is known to be associated with increased social sensitivity and emotional expressiveness. In the second experiment, we directly measured and manipulated specific individual differences in folk intuitions (i.e. naïve intuitions of non-specialists) related to intentional action concepts (Cushman & Mele, 2008). For example, we assessed the extent to which the folk tend to act as if *belief-is-sufficient* to judge that a side effect with little or no affective valence is intentional (e.g. does a marksman intentionally heat up the barrel of his gun as a result of shooting at a target). This allowed for a more direct assessment of the inde-

pendent contribution(s) of both global and specific sources of individual variation, and also provided a unique experimental test of the affective bias account.

## 2. Experiment 1

We sought to use an individual differences approach to provide an initial test of the affective bias explanation of the judgment asymmetry given that emerging research has demonstrated a number of theoretically interesting relationships between individual differences and judgments in the related domains of ethics, free will, and epistemology (Feltz & Cokely, 2008a, 2008b; Feltz, Cokely, & Nadelhoffer, 2009; Haidt, 2007). We were primarily interested in the influence of the general trait extraversion, which is often stable across one's lifespan and is known to exert influence in a number of ways. For example, many personality traits may shape information processing by influencing (i) one's ability to detect cues, (ii) one's perception of the saliency of cues, and (iii) one's general motivation to process information (Funder, 1995, 1991; McCrae & Costa, 1990). We reasoned that if the intentional action side effect asymmetry is the result of an affective bias the personality trait extraversion might be associated with variability in those judgments. That is, research indicates extraversion is associated with looser regulation of affective reactions and greater sensitivities to social dynamics, two factors that theoretically may influence the judgment asymmetry.

Specifically, extraversion is a member of the Five Factor model of personality (John, 1999), and is in some way represented in almost all major personality models (Lucas, Diener, Grob, Eunkook, & Liang, 2000). Within the Big Five model, an extravert is defined as one who is a "communicative, sociable, energetic person who thrives on social contact and who does not regulate tightly his/her emotional reactions" (Akert & Panter, 1988, p. 966). Extraverts enjoy social interaction, find it rewarding, and actively seek out opportunities to be socially and emotionally engaged (Ashton, Lee, & Paunonen, 2002; Lucas & Fujita, 2000). Indeed, extraversion is correlated with a variety of unique emotionally expressive behaviors, socially focused judgments, and memory retrieval processes (Akert & Panter, 1988; Chamorro-Premuzic, Furnham, & Ackerman, 2006; Lucas & Fujita, 2000; Zelenski & Larsen, 2002). For these reasons, we expected the socially-minded and emotionally expressive extraverts may be more likely to exhibit an affective bias in the chairman case. Hence, we formed and tested Hypothesis 1: The judgment asymmetry is positively related to extraversion, controlling for other general individual differences.

In this experiment, we were principally concerned with the influence of extraversion on the judgment asymmetry; however, it was also important to assess whether, and the extent to which, extraversion uniquely predicted the judgment asymmetry. Indeed, other individual differences such as intelligence, cognitive impulsivity, and expertise are known to influence judgment in theoretically interesting ways (Cokely, 2007; Ericsson, Prietula, & Cokely, 2007; Frederick, 2005; Shanteau, 1992; Stanovich & West, 2000). Individual differences in working memory capacity have also been shown to be associated with variations in metacognition (Cokely, Kelley, & Gilchrist, 2006). Therefore, we controlled for these and other individual differences by including: (1) the cognitive reflection task (CRT) which measures cognitive impulsivity or one's reliance on more intuitive (e.g. automatic) versus deliberative (e.g. effortful and subjectively controlled) cognitive processing (Frederick, 2005); (2) working memory capacity (as measured by complex span), i.e. the ability to simultaneously process and store information which significantly mediates the relationship between intelligence and cognitive performance (Turner & Engle, 1989); (3) the brief self-control instrument which measures one's habit of exercising self restraint and one's trait of self control; (4) self reported scho-

<sup>2</sup> It should be noted that it is controversial whether one's intuitions reflect one's concept (see Machery, 2008, for a discussion). If one finds this worry compelling, one could interpret the results that people respond "as if" they have the relevant concept that is used to generate their intuitions.

lastic aptitude test (SAT) scores which are known to correlate with general intelligence (Frederick, 2005); and (5) the Big Five personality factors including one's agreeableness, openness to experience, emotional stability, conscientiousness, and extraversion.

## 2.1. Methods

### 2.1.1. Participants

Participants included 95 Florida State University students who participated for partial fulfillment of course requirements.

### 2.1.2. Procedure and materials

All participants were individually tested by a single experimenter in a quiet and private room for between 1 and 2 h. At the start of the experiment the participant reported their sex, age, and SAT scores. Next, a computerized experimenter assisted working memory operation-span measure was administered (OSPAN, Turner & Engle, 1989) followed by the three item CRT task (CRT, Frederick, 2005). Next, participants completed the brief Big Five personality instrument (Gosling, Rentfrow, & Swann, 2003) and the brief self-control measure (Tangney, Baumeister, & Boone, 2004). Lastly, participants completed the modified Knobe (2003a) survey described in the introduction of this paper.<sup>1</sup> The order of the scenarios was systematically counterbalanced (Help-Harm, Harm-Help).

## 2.2. Results

A mixed-model analysis of variance (ANOVA) with order (Help first, Harm second) as a between-subjects variable and side effect (Harm, Help) as a repeated measure revealed a large difference in side effect intentionality judgments (i.e. asymmetry),  $F(1,93) = 148.24$ ,  $p = .001$ ,  $\eta_p^2 = .61$ . Harmful side effects were judged intentional ( $M = 5.0$ ,  $SD = 1.9$ ) whereas helpful side effects were judged unintentional ( $M = 2.1$ ,  $SD = 1.5$ ).<sup>3</sup> Next, we conducted a stepwise multiple linear regression with the side effect asymmetry as the dependent variable and the Big Five, brief self-control, OSPAN, SAT, CRT, and sex as independent variables. The analysis revealed that only extraversion was reliably related to the judgment asymmetry,  $\beta = .29$ ,  $t = 2.46$ ,  $p = .02$ ,  $R^2 = .08$ . All other individual differences were unreliable ( $F_s < 1$ ). A planned hierarchical linear regression was next constructed again using the judgment asymmetry as a dependent variable and with independent variables including (1) Help-Harm order (to control for the observed order effect)<sup>3</sup> and (2) extraversion. The full model was a significant predictor of the asymmetry,  $F(1,89) = 7.71$ ,  $p = .001$ ,  $R^2 = .15$ . After controlling for the order effect, extraversion continued to account for unique reliable variance,  $\beta = .27$ ,  $t = 2.68$ ,  $p = .01$ ,  $R^2_{\text{change}} = .07$ .

Lastly, planned analyses split extraversion scores into rough top (extravert) and bottom (introvert) quartiles.<sup>4</sup> ANOVA revealed a

large significant extraversion (extravert, introvert) by asymmetry interaction,  $F(1,38) = 8.11$ ,  $p = .01$ ,  $d = .9$ . The side effect asymmetry was much larger for extraverts (Harm  $M = 6.2$ ,  $SD = 1.1$ ; Help  $M = 2.1$ ,  $SD = 1.8$ ) than for introverts (Harm  $M = 4.7$ ,  $SD = 2.1$ ; Help  $M = 2.7$ ,  $SD = 1.6$ ) (Fig. 1). Neither sex nor order interacted with the asymmetry ( $F_s < 1$ ).

## 2.3. Discussion

Consistent with Hypothesis 1, extraversion was found to be uniquely associated with the side effect asymmetry, controlling for other individual differences. Results indicated that extraversion reliably accounted for moderate to large amounts of unique variance in the side effect asymmetry. Individuals low in extraversion often did not agree that the chairman intentionally harmed the environment while more extraverted participants showed firm agreement. Thus, we seem to have identified a personality trait that largely accounts for the judgment asymmetry. Considering the host of other individual differences that were assessed, extraversion appears to be the single most influential general personality factor. Broadly, the results of Experiment 1 are consistent with an affective bias account of the judgment asymmetry. However, because more specific individual differences in folk concepts may also play a role (Cushman & Mele, 2008) a second experiment was conducted.

## 3. Experiment 2

The results of Experiment 1 were suggestive although the exact relationships between extraversion, specific folk concepts, and judgment biases remained unclear. Extraversion could have an effect because it (a) was mediated by more specific folk concepts; (b) reflected an affective bias; or (c) both a and b. Following others (Cushman & Mele, 2008; Nadelhoffer, 2006), we sought to assess specific folk concepts using scenarios that have little or no affective content. Furthermore, in measuring folk concepts we also causally tested the affective bias account with a manipulation of presentation orders. We reasoned that when participants made judgments about non-affective side effect cases *before* they made judgments about side effect cases with affective content these measures could prime participants' core concepts thereby reducing and revealing affective reactivity and bias. Thus, Experiment 2 was designed to provide a test of the affective bias account while controlling for

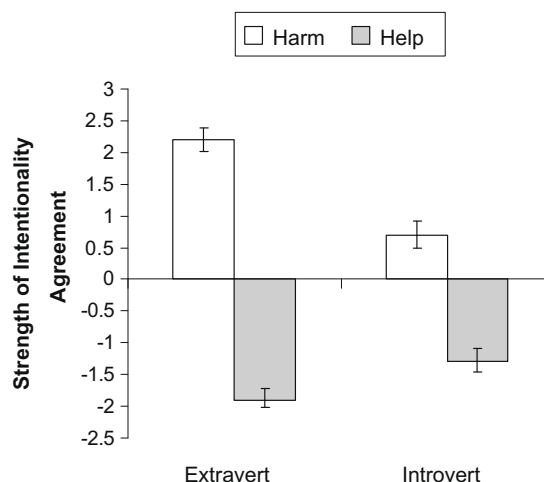


Fig. 1. Introverts and Extraverts (bottom versus top quartile<sup>3</sup>) by side effect intentionality ratings (harm, help). Positive numbers indicate agreement, negative numbers disagreement, and error bars represent the standard error of the mean.

<sup>3</sup> This finding was qualified by the replication of another known effect (Feltz & Cokely, 2007). ANOVA revealed an order effect (Harm first, Harm second) by asymmetry interaction,  $F(1,93) = 7.71$ ,  $p = .01$ ,  $\eta_p^2 = .08$ . When Help was presented first, the Help ( $M = 2$ ) and Harm ( $M = 4.2$ ) asymmetry was significantly smaller as compared to the Help ( $M = 2$ ) and Harm ( $M = 5.8$ ) ratings when Help was presented second. An identifiable sub-group, namely women, were primarily responsible for the order effect. For a more detailed discussion, see Feltz & Cokely (2007).

<sup>4</sup> Rough quartiles were constructed following visual inspection of the data for possible outliers. Next, exact quartiles were computed to identify the cut-off extraversion values for quartile groups (e.g. 25% of the sample scored 7 or less). Finally, rough quartiles were constructed such that each group came as close to the exact number of participants in the quartile without arbitrarily excluding others of the same extraversion score (e.g. if 4 participants had an extraversion score of 7, and 7 was the cut off score, all 4 participants were included whether or not this made the "rough" quartile slightly larger than the exact quartile). In Experiment 1, introverts scored 7 or lower (low) and extraverts scored 13 and higher (high). In Experiment 2, within the belief-is-insufficient group, introverts scored 7 or lower (low) and extraverts scored 11 and higher (high).

and assessing relationships between specific (i.e. concepts) and general (i.e. extraversion) individual differences.

In order to measure specific concepts, Experiment 2 provided participants with two new cases. In the first new case (Deer), we asked participants for intentionality judgments in a scenario where the person *neither desires nor believes* that he will bring about the side effect.<sup>5</sup> This scenario was included to increase attention to specific aspects of the second new scenario (i.e. the role of belief). In the second case (Eagle), we asked participants to make intentionality judgments in a non-moral case where the person *has belief* that he will bring about the side effect *but not the desire* to bring it about.<sup>6</sup> Thus, for subsequent analysis we identified two specific concepts. First, a *belief-is-sufficient* concept was operationalized as the judgment that a side effect in Eagle was brought about intentionally. Alternatively, a *belief-is-insufficient* concept was operationalized as the judgment that the side effect in the Eagle case was not brought about intentionally.

Experiment 2 allowed for a test of three hypotheses. We hypothesized that both specific concepts and general personality differences (i.e. extraversion) would be related to judgment in systematic ways. We predicted that the belief-is-sufficient group would exhibit higher overall and perhaps qualitatively different intentionality ratings as compared to the belief-is-insufficient group. After all, if one applies the belief-is-sufficient concept then the chairman should be seen as acting intentionally because he clearly believes the environment will be affected. However, we also predicted that extraversion would continue to account for unique variance, particularly in harm judgments, as we hypothesized that extraversion was related to the expression of affective bias. Lastly, we tested the affective bias mechanism directly. We hypothesized that when a folk concept was primed (i.e. making intentionality judgments in non-affective cases before making side effect judgments) the priming would attenuate the judgment asymmetry. To summarize, the second experiment tested the following hypotheses (2–4):

2. Priming the belief condition will result in an overall reduction of the intentional action side effect asymmetry.
3. Those who are identified as having a belief-is-sufficient concept will judge both side effects as more intentional, as compared to those who have a belief-is-insufficient concept.
4. Extraversion will continue to account for unique judgment variance in the intentional action side effect asymmetry.

<sup>5</sup> The case that lacked belief and desire was the following. Deer: imagine that there is a man out in the woods who is participating in a hunting competition. After spending hours waiting for a deer to cross his path, the hunter suddenly sees the largest deer he has ever seen. If he can only kill this deer, he will surely win the competition. So, the hunter gets the deer in his sights and pulls the trigger—thereby killing the deer. Unfortunately, the bullet exited the deer's body and struck a hunter who was hiding nearby (Nadelhoffer, 2006). Participants were asked to what degree they agree on a 7 point Likert scale (1 = agree, 4 = neutral, 7 = disagree) with the following statement: "The man intentionally shot the hunter."

<sup>6</sup> The non-moral case with belief and without desire was the following. Eagle: imagine that there is a man out in the woods who is participating in a hunting competition. After spending hours waiting for a deer to cross his path, the hunter suddenly sees the largest deer he has ever seen. If he can only kill this deer, he will surely win the competition. So, the hunter gets the deer in his sights—but at the last second, he notices that there is a beautiful eagle perched in a tree nearby. The hunter realizes that if he shoots the deer, the sound of the gunfire will definitely cause the eagle to fly away. But he does not care at all about the eagle—he just wants to win the competition. So, he shoots and kills the deer. And as expected, the sound of the gunfire causes the eagle to fly away (Nadelhoffer, 2006). Participants were asked to what degree they agree on a 7 point Likert scale (1 = agree, 4 = neutral, 7 = disagree) with the following statement: "The hunter intentionally scared away the eagle."

### 3.1. Methods

#### 3.1.1. Participants

Participants included 133 Florida State University students who participated in exchange for course extra-credit.

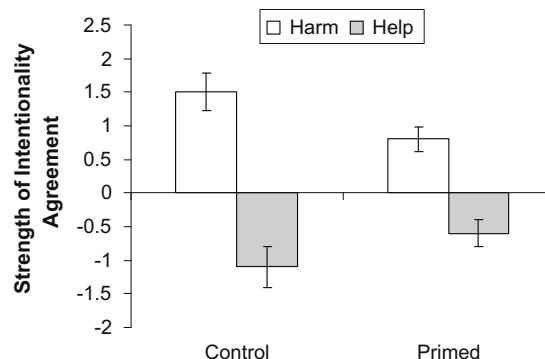
#### 3.1.2. Procedure and materials

Participants were tested in groups. After reporting their sex, participants completed the brief Big Five personality inventory (see Experiment 1) and made judgments about each of the four scenarios (i.e. harm, help, deer, eagle). The order of scenarios was systematically manipulated. The priming cases were presented either before harm and help judgments (i.e. deer, eagle, harm, help) or following harm and help judgments (harm, help, deer, eagle). Harm was always presented before help in order to eliminate the known order effect, simplifying analyses.

### 3.2. Results

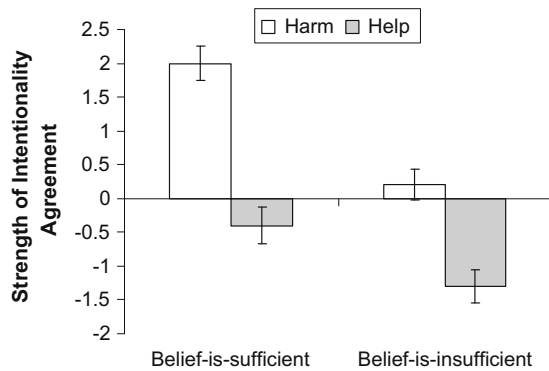
A mixed-model ANOVA with priming (primed, non-primed) and concepts (belief-is-sufficient, belief-is-insufficient) as between-subjects variables and side effect (Harm, Help) as a repeated measure revealed the typical side effect judgment asymmetry,  $F(1, 129) = 81.50, p = .001, \eta_p^2 = .38$ . On average, the harmful side effect was judged intentional ( $M = 5.0, SD = 2.0$ ) whereas the helpful side effect was unintentional ( $M = 3.2, SD = 2.0$ ). However, the asymmetry was qualified by two higher-order interactions. ANOVA revealed a priming (primed, non-primed) by side effect (Harm, Help) interaction,  $F(1, 129) = 6.63, p = .01, \eta_p^2 = .05$ , indicating that the asymmetry was significantly reduced when concepts were primed before side effect judgments (Harm  $M = 4.8, SD = 2.0$ ; Help  $M = 3.4, SD = 2.0$ ), rather than after side effect judgments (Harm  $M = 5.5, SD = 1.9$ ; Help  $M = 3.0, SD = 2.0$ ) (Fig. 2.). Consistent with Hypothesis 1, this interaction indicated the presence of judgment instability and bias. An ANOVA also revealed a concept (belief-is-sufficient, belief-is-insufficient) by side effect (Harm, Help) interaction,  $F(1, 129) = 5.58, p = .02, \eta_p^2 = .04$ . The belief-is-sufficient group showed a larger, qualitatively different asymmetry when judging side effects (Harm  $M = 6.0, SD = 1.6$ ; Help  $M = 3.6, SD = 2.2$ ) as compared to the belief-is-insufficient group (Harm  $M = 4.2, SD = 2.0$ ; Help  $M = 2.8, SD = 1.7$ ) (Fig. 3). Consistent with Hypothesis 2, the belief-is-sufficient group showed higher overall intentionality ratings compared to the belief-is-insufficient group,  $F(1, 129) = 23.44, p = .001, d = .08$ . The three way interaction was not significant ( $F < 1$ ) (Table 1).

Next, a series of planned hierarchical linear regression models was constructed to evaluate relationships among key variables.



**Fig. 2.** The change in judgment asymmetry by priming (primed, control) indicating the presence of judgment bias and instability. Positive numbers indicate agreement, negative numbers indicate disagreement, and error bars represent the standard error of the mean.





**Fig. 3.** Illustrates that large qualitative differences in the judgment asymmetry are predicted by individual differences in specific concepts (belief-is-sufficient, belief-is-insufficient). Positive numbers indicate agreement, negative numbers indicate disagreement, and error bars represent the standard error of the mean.

**Table 1**

Intercorrelations for main variables.

	1	2	3	4
1. Extraversion				
2. Belief-is-sufficient	.19*			
3. Help	-.03	.29**		
4. Harm	.18*	.54**	.28*	
5. Side effect asymmetry	.17*	.21*	-.60**	.60**

\*  $p < .05$ .

\*\*  $p < .01$ .

The final model had three independent variables including (1) priming, (2) extraversion, and (3) concepts as predictors of harm judgments (Table 2). This model accounted for a large amount of variance in judgment,  $F(3, 130) = 21.37$ ,  $p = .001$ ,  $R^2 = .33$ . After controlling for priming, extraversion remained a reliable predictor of the harm judgments. However, extraversion's effect was also mediated by the large effect of concepts (Table 2). Thus, part of the predictive power of extraversion appears to result from extraversion's positive association with the belief-is-sufficient concept.

Lastly planned hierarchical regression models examined extraversion within each of the two concept groups. Regression analysis indicated that for the belief-is-sufficient group, extraversion was unrelated to judgment asymmetry ( $F < 1$ ). Regression analysis next assessed the belief-is-insufficient group using (1) priming and (2) extraversion as predictors. This model was a reliable predictor of the judgment asymmetry,  $F(1, 67) = 3.29$ ,  $p = .04$ ,  $R^2 = .09$ . Consistent with Hypothesis 3, after controlling for the effect of priming, extraversion continued to account for unique variance,

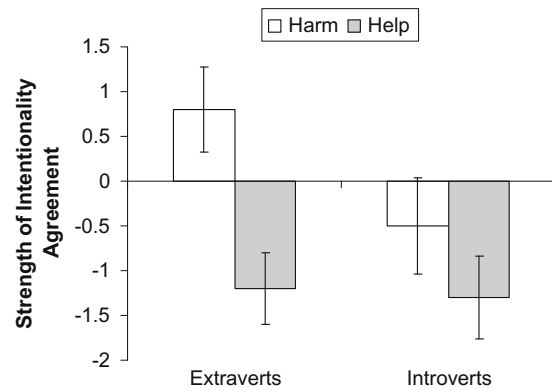
**Table 2**

Hierarchical linear regression analysis explaining intentional action judgments in the Harm condition.

Steps and variables	$\beta$	$R$	$R^2$	$\Delta R^2$	$F$
Model 1					
Order-effect	.13	.13	.02	.02	2.43
Model 2					
Order-effect	.12				
Extraversion	.17*	.22	.05	.03	3.18*
Model 3					
Order-effect	.18*				
Extraversion	.06				
Belief-is-sufficient	.54**	.58	.33	.28	21.37**

\*  $p < .05$ .

\*\*  $p < .01$ .



**Fig. 4.** The side effect judgment asymmetry (Help versus Harm) for belief-is-insufficient, which is predicted by extraversion (Introvert versus Extravert<sup>3</sup>). Positive numbers indicate agreement, negative numbers disagreement, and error bars represent the standard error of the mean.

$F(1, 67) = 4.74$ ,  $p = .03$ ,  $R^2_{change} = .06$ , for those individuals with the belief-is-insufficient concept. As predicted, the unique effect of extraversion primarily reflected a relationship with harm judgments,  $r(68) = .24$ ,  $p = .05$ , (Help n.s.) (Fig. 4).<sup>3</sup>

### 3.3. Discussion

The results of Experiment 2 provide support for Hypotheses 2–4. As predicted by Hypothesis 2, the judgment asymmetry was smaller following priming. This provides causal evidence of the presence of judgment bias in both harm and help cases (see also Nadelhoffer, 2004). However, the priming effect was relatively small compared to the large, qualitative judgment differences associated with different concepts. Consistent with Hypothesis 3, we observed that people with a belief-is-sufficient concept tended to judge actions as more intentional whereas people with a belief-is-insufficient tended to judge that actions as unintentional (Fig. 3). Although it was not anticipated, extraversion was also found to be a significant predictor of differences in concepts. Extraverts were more likely to belong to the belief-is-sufficient group whereas introverts were more likely to belong to the belief-is-insufficient group. The current data are silent on the mechanisms of this relationship. We speculate that holding a belief-is-sufficient concept may be an adaptive or ecologically rational mechanism that helps regulate certain social interactions according to goals (Chase, Hertwig, & Gigerenzer, 1998; Gigerenzer & Hug, 1992; Gigerenzer, Todd, & the ABC Research Group, 1999). That is, when interacting with larger and expressive social groups perhaps an assumption that belief-is-sufficient for intentional action ascription may tend to enable interactions and eliminate obstacles to emotional expressions that are prized by group members (i.e. extraverts). As an example, one might imagine that excessive alcohol consumption (i.e. intent to become inebriated) might be associated with unintended side effects (e.g. infidelity, aggression) that would readily disrupt group dynamics (e.g. relationship turmoil, physical altercations). In this case, a belief-is-sufficient concept may serve a social regulatory function reducing discord by facilitating the assignment of blame for undesirable side effects (Adams & Steadman, 2004).

Consistent with Hypothesis 4, extraversion accounted for unique variance in the judgment asymmetry for people with the belief-is-insufficient concept. This result is consistent with and clarifies Cushman and Mele's (2008) finding that some people who generally have a belief-is-insufficient concept (i.e. desire-concept) tend to make different judgments for moral cases (i.e. when affect is high they behave as if they have a belief-is-sufficient concept). This suggests that affective valence is a critical feature that

selectively triggers the use of specific judgment processes, such as heuristics, for extraverts. This also provides additional converging support for an affective judgment bias account. Furthermore, we speculate that this selective effect of affect may also reflect an adaptive process for extraverts enabling social regulation as previously described. Indeed, it seems that the rapid assignment of blame may be most crucial and valuable for social interactions that involve strong affect and emotion. In summary, Experiment 2 replicated and refined Experiment 1, demonstrating that the intentional action asymmetry is multiply determined, resulting from both affective biases and individual differences in personality and folk concepts.

#### 4. General discussion

The current set of experiments suggest that a variety of factors play theoretically important roles in intentional action judgments. The intentional action side effect asymmetry does not appear to result from a single mechanism but instead reflects the influence of both individual differences and judgment processes (e.g. biases). Evidence of individual differences in bias came from Experiment 1 which revealed that extraversion was uniquely related to the side effect asymmetry, controlling for other potentially influential general individual differences. Experiment 2 provided additional converging evidence of a judgment bias demonstrating that priming intentional action concepts causally reduced the judgment asymmetry. Experiment 2 also replicated the relationship between extraversion and the judgment asymmetry; however, extraversion's effect was shown to be mediated, in part, by its association with specific concepts. Extraverts tended toward a belief-is-sufficient concept whereas introverts tended toward a belief-is-insufficient concept. These specific individual differences in concepts were in turn associated with large, qualitative and theoretically significant differences in judgment (Fig. 3). Those who held the belief-is-sufficient concept tended to judge that all side effects were more intentional whereas those who held a belief-is-insufficient concept tended to judge that all side effects were less intentional. Finally, when individual differences in concepts were taken into account, extraversion continued to predict unique variance in harm judgments for some people (i.e. the belief-is-insufficient group).

##### 4.1. Theoretical implications

Broadly, the current results yield at least three theoretically important findings. First, our findings challenge previous interpretations of the judgment asymmetry (see also Cushman & Mele, 2008). There is not necessarily any intentional action judgment asymmetry as there may not necessarily or typically be any complete judgment reversal (Knobe, 2003a). Rather than any single general 'bias' or judgment process that causes participants to switch from intentional (harm) to unintentional (help) judgments, the asymmetry in judgment appears to be more modest, changing for example from neutral judgments to intentional (belief-is-sufficient) or from unintentional to neutral (belief-is-insufficient) (Fig. 3). The only identifiable and somewhat complete judgment reversal primarily involved the group of extraverts who behaved as if they held a belief-is-insufficient concept (Fig. 4). In these ways and others, the observed judgment asymmetry, which seems to be the product of the interplay of several distinct mechanisms, may be more accurately characterized as a collection of intentional action phenomena or multiple judgment asymmetries.

Second, contrary to what appears to be philosophical orthodoxy, results reveal that there is no *the folk* concept of intentional action. In line with other recent results (Cushman & Mele, 2008;

Nichols & Ulatowski, 2007), we find evidence that participants behave as if there are at least two folk concepts of intentional action: (a) one that treats belief as sufficient for acting intentionally and (b) one that treats belief as insufficient for acting intentionally. The finding that the judgments of the folk are diverse and systematically fragmented mirrors other emerging evidence indicating that stable individual differences play roles in a variety of domains such as ethics, free will, and epistemology (Feltz & Cokely, 2008a, 2008b; Feltz et al., 2009; Haidt, 2007).

Lastly, and more broadly, we think these results provide a somewhat unique demonstration of the too often neglected role of individual differences in human judgment and cognition (Cronbach, 1957; Funder, 2001; Humphreys & Revelle, 1984). An individual differences approach can provide a valuable tool allowing for a higher-fidelity identification of proximal judgment and performance mechanisms, particularly when used in combination with process tracing or experimental manipulations (Broder, 2003; Ericsson & Simon, 1980; Fasolo, Misuraca, & McClelland, 2003; Gaissmaier, Schooler, & Rieskamp, 2006; Mata, Schooler, & Rieskamp, 2007; Payne, 1976; Rhodes & Kelley, 2005; Ricks, Turley-Ames, & Wiley, 2007; Roring & Charness, 2007). Higher fidelity measurement may suggest new research questions and can improve the accuracy of theoretical models and their applications. To illustrate, individual variation in judgment and performance often reflects adaptive, environmentally tuned goals, processes, and representations (Cokely et al., 2006; Ericsson & Kintsch, 1995; Funder, 1995). By reverse engineering variations in goals and strategies (e.g. by identify specific processes and practices that mediate individual differences) we can better understand the sources of individual differences (Cokely, 2007; Ericsson, Roring, & Nandagopal, 2007; Ericsson, Krampe, & Teshromer, 1993; Guida, Tardieu, & Nicolas, 2008; Shanteau, 1992) and design more effective training systems, interventions, and support technologies (Gaeth & Shanteau, 1984; Gigerenzer et al., 1999; Todd & Gigerenzer, 2007). In the case of intentional action judgments, in addition to improving our theoretical understanding of theory-of-mind, understanding individual differences informs legal and public policy issues including legal concepts, jury selection, and jury instructions (Malle & Nelson, 2003; Nadelhoffer, 2006).

In conclusion, the current research demonstrates that there is no single intentional action side effect asymmetry. Instead, the observed judgment asymmetry appears to be multiply determined resulting from the interplay of judgment biases and individual differences. The influence of both general personality and specific conceptual individual differences demonstrates that there is not necessarily a homogenous "*the folk*" whose judgment can be assessed at the level of group means. In contrast, as different people have different goals, sensitivities, and experiences, their judgments and biases can be expected to vary in systematic, adaptive, and theoretically interesting ways. We think that the current results more realistically reflect the depth and complexity of processes that are involved in even the simplest theory-of-mind judgments. On the whole, these results provide converging evidence that the assessment of both global and specific individual differences can be an essential tool for understanding folk intuitions and judgment processes.

#### References

- Adams, F., & Steadman, A. (2004). Intentional action in ordinary language: Core concept or pragmatic understanding. *Analysis*, 74, 173–181.
- Akert, R., & Panter, A. (1988). Extraversion and the ability to decode nonverbal communication. *Personality and Individual Differences*, 9, 965–972.
- Ashton, M., Lee, K., & Paunonen, S. (2002). What is the central feature of extraversion? Social attention versus reward sensitivity. *Journal of Personality and Social Psychology*, 83, 245–252.

- Broder, A. (2003). Decision making with the "adaptive toolbox": Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology-Learning Memory and Cognition*, 29, 611–625.
- Chamorro-Premuzic, T., Furnham, A., & Ackerman, P. (2006). Ability and personality correlates of general knowledge. *Personality and Individual Differences*, 41, 419–429.
- Chase, V. M., Hertwig, R., & Gigerenzer, G. (1998). Visions of rationality. *Trends in Cognitive Sciences*, 2, 206–214.
- Cokely, E.T. (2007). Strategies, heuristics, and adaptive decision-making: New evidence from an individual differences and process tracing approach. In *Proceedings of the 29th annual meeting of the Cognitive Science Society* (p.1726). Mahwah, NJ: Erlbaum.
- Cokely, E. T., Kelley, C. M., & Gilchrist, A. L. (2006). Sources of individual differences in working memory: Contributions of strategy to capacity. *Psychonomic Bulletin & Review*, 13, 991–997.
- Cronbach, L. J. (1957). The 2 disciplines of scientific psychology. *American Psychologist*, 12, 671–684.
- Cushman, F., & Mele, A. (2008). Intentional action: Two-and-a-half folk concepts? In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy* (pp. 171–188). Oxford: Oxford University Press.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211–245.
- Ericsson, K. A., Krampe, R. Th., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363–406.
- Ericsson, K. A., Prietula, M. J., & Cokely, E. T. (2007). The making of an expert. *Harvard Business Review*, 85, 114–121.
- Ericsson, K. A., Roring, R. W., & Nandagopal, K. (2007). Giftedness and evidence for reproducibly superior performance. an account based on the expert performance framework. *High Ability Studies*, 18, 3–56.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87, 215–251.
- Fasolo, B., Misuraca, R., & McClelland, G. H. (2003). Individual differences in adaptive choice strategies. *Research in Economics*, 57, 219–233.
- Feltz, A. (2007). The Knobe effect: A brief overview. *Journal of Mind and Behavior*, 28, 265–277.
- Feltz, A., & Cokely, E.T. (2007). An anomaly in intentional action ascription: More evidence of folk diversity. In *Proceedings of the 29th annual meeting of the Cognitive Science Society* (p. 1748). Mahwah, NJ: Erlbaum.
- Feltz, A., & Cokely, E. T. (2008a). Do judgments about freedom and responsibility depend on who you are?: Personality differences in intuitions about compatibilism and incompatibilism. *Consciousness and Cognition*. doi:10.1016/j.concog.2008.08.001.
- Feltz, A., & Cokely, E. T. (2008b). The fragmented folk: More evidence of stable individual differences in moral judgments and folk intuitions. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 1771–1776). Austin, TX: Cognitive Science Society.
- Feltz, A., Cokely, E. T., & Nadelhoffer, T. (2009). Natural compatibilism versus natural incompatibilism: Back to the drawing board. *Mind and Language*, 24, 1–23.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspective*, 19, 25–42.
- Funder, D. (1991). Global traits: A neo-Allportian approach to personality. *Psychological Science*, 2, 31–39.
- Funder, D. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102.
- Funder, D. (2001). Personality. *Annual Review of Psychology*, 52, 197–221.
- Gaeth, G. J., & Shanteau, J. (1984). Reducing the influence of irrelevant information on experiences decision makers. *Organizational Behavior and Human Performance*, 33(2), 263–282.
- Gaissmaier, W., Schooler, L. J., & Rieskamp, J. (2006). Simple predictions fueled by capacity limitations: When are they successful? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 32, 966–982.
- Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, 7, 77–83.
- Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning - social contracts, cheating, and perspective change. *Cognition*, 43, 127–171.
- Gigerenzer, G., Todd, P. M., & The ABC Research Group (1999). *Simple heuristics that make us smart*. Oxford [u.a.]: Oxford University Press.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37, 504–528.
- Guida, A., Tardieu, H., & Nicolas, S. (2008). The personalisation method applied to a working memory task: Evidence of long-term working memory effects. *European Journal of Cognitive Psychology*. doi:10.1080/09541440802236369.
- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316, 998–1002.
- Humphreys, M. S., & Revelle, W. (1984). Personality, motivation, and performance—a theory of the relationship between individual-differences and information-processing. *Psychological Review*, 91(2), 153–184.
- John, O. (1999). The Big-five trait taxonomy: History, measurement, and theoretical perspectives. In L. Pervin & O. P. John (Eds.), *Handbook of Personality: Theory and Research* (2nd ed (pp. 66–100). New York: Guilford.
- Knobe, J. (2003a). Intentional action and side-effects in ordinary language. *Analysis*, 63, 190–194.
- Knobe, J. (2003b). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology*, 16, 309–323.
- Knobe, J. (2004a). Folk psychology and folk morality: Response to critics. *Journal of Theoretical and Philosophical Psychology*, 24, 270–279.
- Knobe, J. (2004b). Intention, intentional action, and moral considerations. *Analysis*, 64, 181–187.
- Knobe, J. (2006). The concept of intentional action: A case study in uses of folk psychology. *Philosophical Studies*, 130, 203–231.
- Knobe, J., & Burra, A. (2006). Experimental philosophy and folk concepts: Methodological considerations. *Journal of Cognition and Culture*, 6, 331–342.
- Leslie, A., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect: Theory of mind and moral judgment. *Psychological Science*, 17, 421–427.
- Lillard, A. S. (1997). Others folks' theories of mind and behavior. *Psychological Science*, 8, 268–274.
- Lillard, A. (1998). Ethnopsychologies: Cultural variations in theories of mind. *Psychological Bulletin*, 123, 3–32.
- Lucas, R., Diener, E., Grob, A., Eunkook, S., & Liang, S. (2000). Cross-cultural evidence for the fundamental features of extraversion. *Journal of Personality and Social Psychology*, 79, 452–468.
- Lucas, R., & Fujita, F. (2000). Factors influencing the relation between extraversion and pleasant affect. *Journal of Personality and Social Psychology*, 79, 1039–1056.
- Machery, E. (2008). The folk concept of intentional action: Philosophical and experimental issues. *Mind and Language*, 23, 165–189.
- Malle, B. F., & Nelson, S. E. (2003). Judging *Mens Rea*: The tension between folk concepts and legal concepts of intentionality. *Behavioral Sciences and the Law*, 21, 563–580.
- Mata, R., Schooler, L. J., & Rieskamp, J. (2007). The aging decision maker: Cognitive aging and the adaptive selection of decision strategies. *Psychology and Aging*, 22, 796–810.
- McCrae, R., & Costa, P. (1990). *Personality in Adulthood*. New York: Guilford Press.
- Mele, A., & Cushman, F. (2007). Intentional action, folk judgments, and stories: Sorting things out. *Midwest Studies in Philosophy*, 31, 184–201.
- Nadelhoffer, T. (2004). On praise, side effects, and folk ascriptions of intentional action. *Journal of Theoretical and Philosophical Psychology*, 24, 196–213.
- Nadelhoffer, T. (2006). Desire, foresight, intentions, and intentional action: Probing folk intuitions. *Journal of Cognition and Culture*, 6, 133–157.
- Nichols, S. (2004). Folk concepts and intuitions: From philosophy to cognitive science. *Trends in Cognitive Science*, 8, 514–518.
- Nichols, S., & Ulatowski, J. (2007). Intuitions and individual differences: The Knobe effect revisited. *Mind & Language*, 22, 346–365.
- Payne, J. W. (1976). Task complexity and contingent processing in decision-making—information search and protocol analysis. *Organizational Behavior and Human Performance*, 16, 366–387.
- Phelan, M., & Sarkissian, H. (2008). The folk strike back: Or, why you didn't do it intentionally, though it was bad and you knew it. *Philosophical Studies*, 138, 281–298.
- Revelle, W. (1987). Personality and motivation—sources of inefficiency in cognitive performance. *Journal of Research in Personality*, 21(4), 436–452.
- Rhodes, M. G., & Kelley, C. M. (2005). Executive processes, memory accuracy, and memory monitoring: An aging and individual difference analysis. *Journal of Memory and Language*, 52, 578–594.
- Ricks, T. R., Turley-Ames, K. J., & Wiley, J. (2007). Effects of working memory capacity on mental set due to domain knowledge. *Memory & Cognition*, 35, 1456–1462.
- Roring, R. W., & Charness, N. (2007). A multilevel model analysis of expertise in chess across the life span. *Psychology and Aging*, 22, 291–299.
- Shanteau, J. (1992). How much information does an expert use—is it relevant. *Acta Psychologica*, 81(1), 75–86.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, 23, 645–665.
- Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72, 271–322.
- Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science*, 16, 167–171.
- Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences*, 16, 495–511.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language*, 28, 127–154.
- Zelenski, J., & Larsen, R. (2002). Predicting the future: How affect-related personality traits influence likelihood judgments of future events. *Personality and Social Psychology Bulletin*, 28, 1000–1010.