American Economic Review 2024, 114(4): 961–987
https://doi.org/10.1257/aer.20191849

Motivated Errors†

By Christine L. Exley and Judd B. Kessler*

Myriad environments allow for the possibility of confusion. Agents may appeal to such confusion—or the possibility of making an honest mistake—to justify their behavior. In three sets of experiments involving thousands of subjects, we document evidence of such motivated errors. We document this evidence in a simple environment in which the scope for errors is small and in more complex environments in which subjects display correlation neglect and an anchoring bias. (*JEL* C91, D12, D64, D81, D83, D91)

Individuals are keen to view themselves (or be viewed by others) in a more favorable light.[1] To achieve this goal, they might appeal to the possibility of being confused or having made a mistake to help justify their behavior.[2] Indeed, because the possibility of being confused or making a mistake is almost always present, they may make errors to justify their behavior—what we call *motivated errors*—in a wide variety of contexts.

When filing taxes, individuals who want to keep more money for themselves while maintaining the belief that they reported honestly could conveniently miscalculate last year's income so it ends up lower and miscalculate deductions so they end up higher. To rationalize favorable views of a preferred political candidate, they could display more correlation neglect when correlated reports about the candidate are positive but less correlation neglect when reports about the candidate are negative. When they want an excuse to drive rather than walk, they could display more of an anchoring bias when it results in them calculating a longer distance to their destination. To justify an indulgent purchase to themselves, they could miscalculate

*Exley: University of Michigan (email: exley@umich.edu); Kessler: The Wharton School, University of Pennsylvania (email: judd.kessler@wharton.upenn.edu). Pietro Ortoleva and Roland Bénabou were the coeditors for this article. We are grateful to the many colleagues and seminar participants who have provided useful comments on this paper. We are particularly grateful to the editors and four anonymous referees at the *American Economic Review* for suggestions that greatly improved the paper. This study was reviewed by the Institutional Review Board at Harvard Business School.

†Go to https://doi.org/10.1257/aer.20191849 to visit the article page for additional materials and author disclosure statement(s).

[1] Indeed, agents may have self image concerns or social image concerns (Rabin 1995; Bodner and Prelec 2003; Bénabou and Tirole 2004, 2006, 2011; Mijović-Prelec and Prelec 2010; Grossman 2015; Grossman and van der Weele 2017; Bénabou, Falk, and Tirole 2018; Foerster and van der Weele 2021).

[2] To fix ideas about what we mean by agents appealing to the possibility that they made a mistake, akin to the model in Bénabou and Tirole (2006), imagine agents in a decision environment have some small probability of taking a selfish action because they made a mistake due to confusion. This implies that choosing the selfish action cannot be viewed as a definitive desire to be selfish since there is always some probability mass on the selfish action being chosen due to confusion. An agent who wants to avoid image costs of being selfish can then mimic "confused" types when doing so serves as a convenient excuse for selfish behavior. This dynamic will result in more evidence of supposed confusion when acting confused can facilitate selfish behavior.

the post-tax price even though they calculate tax correctly for goods they are less eager to buy. After a purchase, they could conveniently think that they received the correct amount of change if the cashier provides too much but identify a shortfall if the cashier provides too little. They could justify selfishness by making an error when multiplying a restaurant bill by a particular fraction, leading them to leave a smaller tip for the staff.

We present evidence of motivated errors across three studies. Specifically, we investigate evidence for motivated errors in a very simple setting (our Adding Study) and in more complicated settings in which well-established behavioral biases are present (our Correlation Neglect and Anchoring Studies). While the decision environment in our Adding Study is novel and intended to investigate motivated errors when the possibility of making a mistake is likely minimal, our latter studies purposefully investigate evidence for motivated errors when mistakes are likely since correlation neglect (Enke and Zimmermann 2019) and anchoring (Enke et al. 2021) are well-established behavioral biases in environments absent selfish motives.[3]

In the Adding Study, subjects make a series of decisions, choosing between receiving a fixed amount of money for themselves and a sum of amounts for charity. When a zero is added to the sum for charity, subjects become less likely to choose the sum for charity. However, when selfish motives are removed—when subjects instead choose between two payoffs that both benefit charity—decisions are no longer influenced by the addition of a zero. Subjects only act as if they cannot add a zero when doing so can help justify selfish decisions.[4]

Further treatments of our Adding Study show that decreasing the scope for confusion—by providing subjects with more information about the sum going to charity—reduces motivated errors. Motivated errors are cut in half when subjects can click a box to reveal the sum of payoffs or when they are shown the sum by default. Motivated errors are fully eliminated when the sum is shown by default *and* subjects must correctly report it back before making their choice.

In the Correlation Neglect Study, subjects are asked to make predictions about the average of correlated information. When they face selfish motives to provide certain predictions, they may justify selfishness by appealing to the possibility of being confused when making these predictions. We observe results consistent with such motivated errors. Indeed, we find that the opportunity to make motivated errors can both exacerbate and mitigate evidence for correlation neglect.

In the Anchoring Studies, subjects are provided with an uninformative anchor and asked knowledge-based questions. Akin to what we find in the Correlation Neglect Study, we observe results consistent with motivated errors that can both exacerbate and mitigate evidence for an anchoring bias.

[3] By eliciting beliefs about unknown states, our Correlation Neglect and Anchoring Studies are also related to prior work on motivated beliefs about ability (Eil and Rao 2011; Ertac 2011; Grossman and Owens 2012; Mobius et al. 2022; Buser, Gerhards, and van der Weele 2018; Coutts 2018; Heger and Papageorge 2018; Schwardmann and van der Weele 2019; Chew, Huang, and Zhao 2020; Zimmermann 2020), politics (Thaler 2019), beauty (Eil and Rao 2011), financial decisions (Kuhnen 2015), and non-ego-relevant but desirable events (Gotthard-Real 2017). See Bénabou and Tirole (2016) for a review; Bénabou and Tirole (2002) and Köszegi (2006) for related theoretical work; and Schwardmann, Tripodi, and van der Weele (2022) for evidence from the field.

[4] When subjects are directly asked to calculate these sums, they do so correctly in approximately 98 percent of cases, and their ability to do so does not depend on whether a zero is added to the sum.

The evidence for motivated errors in this paper—arising when individuals may appeal to uncertainty about whether they are confused or made a mistake to justify selfishness—contributes to a rich prior literature that has focused on other ways in which individuals may appeal to uncertainty to rationalize behavior that could otherwise be viewed as undesirable. Specifically, we note two broad ways that have been the focus of prior literature.

The first way involves appealing to uncertainty in how decisions map to outcomes. The intuition is as follows. Choosing to benefit oneself to the detriment of others is undesirable, but choosing to benefit oneself when the action may not harm others (e.g., because of uncertainty in how decisions map to outcomes) may be less undesirable. Engaging in this type of motivated reasoning is often facilitated by information avoidance. In the canonical Dana, Weber, and Kuang (2007), selfishness increases when agents can stay uninformed about whether a selfish action harms another subject. Evidence for this type of motivated reasoning is also evident when an individual acts more selfishly when observers are uncertain about the mapping between the individual's decision and outcomes (Dana, Cain, and Dawes 2006; Broberg, Ellingsen, and Johannesson 2007; Andreoni and Bernheim 2009; Linardi and McConnell 2011; Shaw et al. 2014).

The second way involves appealing to uncertainty about whether undesirable decisions are attributable to more innocuous preferences or beliefs. The intuition is as follows. It is bad to be selfish, but there is nothing wrong with holding particular preferences (over, say, risk) or holding particular beliefs (over, say, what payoffs are likely to arise), even if those preferences or beliefs lead you to make selfish decisions. In one of the earliest examples, Snyder et al. (1979) shows that subjects, when deciding whether or not to watch a movie with an individual who has a disability, are more likely to avoid this individual when their avoidance "could masquerade as a movie preference" rather than a dislike for an individual who has a disability. Empirical work documents that individuals appear to use their fairness preferences (Konow 2000), ambiguity preferences (Haisley and Weber 2010), risk preferences (Exley 2016), and beliefs about factors that influence payoffs—such as how others behave (Di Tella et al. 2015)—to rationalize decisions that could otherwise be attributed to selfishness.[5]

Future work may investigate how to mitigate the various ways in which individuals may appeal to uncertainty—including uncertainty about being confused or making a mistake—to justify their behavior. This question is interesting and potentially important for a broad range of contexts because the possibility of being confused exists in many contexts, as highlighted by the examples at the beginning of our introduction, the results in our Correlation Neglect and Anchoring Studies, and

[5] Other examples include cases where decisions may be rationalized by condo preferences (Hsee 1996), dispute-related fairness preferences (Babcock et al. 1995), honesty preferences (Danilov and Saccardo 2016), preferences about charity metrics (Gneezy, Keenan, and Gneezy 2014; Exley 2020; Palma and Xu 2019), and beliefs about competence (Liu and Lin 2018). In addition to early work (Kunda 1990; Batson et al. 1997) and review articles (Chance and Norton 2015; Gino, Norton, and Weber 2016) that stress how ambiguity and uncertainty contribute to motivated reasoning, ambiguity and risk preferences may also be relevant when payoffs explicitly depend on ambiguity or risk (Dana, Weber, and Kuang 2007; Oberholzer-Gee and Eichenberger 2008; Gneezy et al. 2020; Garcia, Massoni, and Villeval 2020; Regner 2018; Olschewski, Dietsch, and Ludvig 2019) or when payoffs are influenced by the (unknown) behavior of others (Falk and Szech 2013; Bartling and Özdemir 2023; Falk and Szech 2020; Gneezy, Saccardo, and van Veldhuizen 2018).

the discussion in Bénabou and Tirole (2016).[6] This question may also be difficult because individuals may exploit even a minor level of uncertainty about being confused. Indeed, subjects in our Adding Study are less likely to choose the sum for charity when a zero is added to it, even though adding a zero to a sum has *no impact* on their choices when selfish motives are removed. Future work may also investigate how different types of excuses operate and interact with one another.

The rest of the paper proceeds as follows. Section I describes the design and results of the Adding Study. Section II describes the design and results of the Correlation Neglect Study. Section III describes the design and results of the Anchoring Studies. Section IV concludes.[7] All our data are available in the repository Exley and Kessler (2024).

I. The Adding Study

In the Adding Study, we examine how subjects respond to a zero being added to a payoff. We find that subjects are more likely to choose a payoff for themselves over a payoff for charity when a zero is added to the payoff for charity. But, when choosing between two payoffs for charity, subjects do not respond to the addition of a zero. Subjects make *motivated errors*: subjects only act as if they make simple addition errors when doing so can justify selfish decisions. In this section, we present the design and results of the Adding Study, documenting motivated errors and showing how they can be mitigated.

A. *The Adding Study*, *Design*

The Adding Study included 1,769 subjects from Amazon Mechanical Turk ("MTurk") who participated in 1 of 8 treatments.[8] Below, we describe the design of our main treatments: the *Charity*/*Charity* treatment and the *Self*/*Charity* treatment. We describe additional treatments as they become relevant in our discussion of results.

In both the *Charity*/*Charity* and *Self*/*Charity* treatments, we examine how adding a zero to a sum of payoffs affects choices. The main variation across treatments is that selfish motives are relevant in the *Self*/*Charity* treatment but not relevant in the *Charity*/*Charity* treatment. We can thus examine choices when subjects may make motivated errors—because appealing to the possibility of simple addition errors may justify selfishness in the *Self*/*Charity* treatment—and when selfishness motives are not relevant in the *Charity*/*Charity* treatment.

[6] See also other recent work inspired by Bénabou and Tirole (2002) on false or selective memory (Chew, Huang, and Zhao 2020; Zimmermann 2020; Saucet and Villeval 2019).

[7] Online Appendixes A, B, and C provide additional design details for the Adding Study, additional results from all the studies, and additional results on information avoidance from the Adding Study. Online Appendix D provides screenshots of the experimental interface across all of our studies. Online Appendix E provides information on prior drafts of this paper.

[8] See the discussion around online Appendix Table A.1 for more details. In all study treatments, each subject received $4 for completing the 25-minute study. In addition, one randomly selected decision for each subject was implemented for bonus payment and resulted in an additional payment for the subject or a donation to charity. Full instructions and additional details for the Adding Study can be found in online Appendix D.D1.

Subjects make 48 binary decisions. In each decision of the *Charity*/*Charity* treatment, one option is for 150 cents to go to the Make-A-Wish Foundation. We call this their "outside option," and it is constant across decisions. The other option is for the sum of four or five summands to go to the Make-A-Wish Foundation. We call this sum a "bundle," and it changes across decisions.

Each bundle includes four or five summands that are either zero or the same nonzero number, d, which is constant within a bundle. Consequently, the sum in the bundle can always be calculated as $n \times d$ (where n is the number of times the positive number d appears, with all remaining summands being zero). The first summand in a bundle is revealed by default, and subjects must click to reveal all remaining summands before making their decision. We present the bundles in this interactive manner (see online Appendix Figure A.1 for an example) to help ensure subjects view all summands in a bundle and to facilitate results relating to how subjects may avoid information as an excuse (see Section IE).

Our main analysis involves 36 bundles (12 bundles with 4 summands and 24 bundles with 5 summands) with $n \in \{2,3,4,5\}$ and $d \in \{51,52,53,54,55,56,57,58,59\}$. These values of d imply that $n \times d$ is substantially below 150 cents for the bundles with $n = 2$, slightly above 150 cents for bundles with $n = 3$, and substantially above 150 cents for the bundles with $n = 4$ and $n = 5$.

To construct these 36 bundles, we start with 12 baseline bundles of 4 summands each and construct 12 five-summand bundles by adding a fifth summand that is 0 to each baseline bundle and 12 five-summand bundles by adding a fifth amount that is d to each baseline bundle (see online Appendix Table A.2 for additional details). Our main identification strategy involves comparing the rate at which subjects choose a four-summand bundle to the rate at which they choose the corresponding five-summand bundle constructed by adding a zero.[9] The order in which subjects make their 48 binary decisions varies by subject.[10]

In the *Self*/*Charity* treatment, everything is identical to the *Charity*/*Charity* treatment except that the "outside option" in each binary decision is an amount of money that goes to the subject (rather than to the Make-A-Wish Foundation).

To make things as similar as possible across these treatments, the specific level of the outside option is determined at the individual level through a multiple price list that precedes the 48 binary decisions described above. The multiple price list elicits an X value that makes the subject indifferent between X cents for themselves and 150 cents for the Make-A-Wish Foundation.[11] Once we identify this X value, we set each subject's outside option to X cents for themselves if they are randomized to the

[9] Along with our 36 main bundles, we have 12 additional bundles with 4 summands each (see online Appendix Table A.3 for details). We included these bundles to balance the number of bundles of each size (i.e., to have 24 bundles with 4 summands and 24 with 5 summands) and to provide additional data to perform secondary analyses.

[10] Half of subjects make their 24 decisions involving 4-summand bundles first, and the other half make their 24 decisions involving 5-summand bundles first. In addition, within each block of 24 decisions, the order in which each bundle is shown randomly varies for each subject.

[11] In particular, the multiple price list generates an indifference range for X. We assign subjects an X value equal to the lower bound of their indifference range unless the lower bound of the indifference range is 0, in which case we assign $X = 5$ cents. The distribution of X values are displayed in panel A of online Appendix Figure B.1. The multiple price list and additional details describing it are shown in online Appendix Figure D.3. To obtain more precise estimates of X cents, one could employ a treatment of the DOSE approach in Wang, Filiba, and Camerer (2010).

Self/*Charity* treatment (and to 150 cents for the Make-A-Wish Foundation if they are randomized to the *Charity*/*Charity* treatment).

By calibrating the outside option in the *Self*/*Charity* treatment to be equivalent to 150 cents for charity, we can ensure that—as in the *Charity*/*Charity* treatment—each subject is close to indifferent between the outside option and the bundle for the $n = 3$ decisions (and further from indifferent when $n = 2$ and $n \geq 4$), so we have a well-controlled measure of how likely the subject is to select the bundle.[12] If we had not calibrated the outside option and instead set it directly (e.g., to 150 cents for the subject to keep the nominal amount the same), we might have ended up with subjects always choosing the outside option in the *Self*/*Charity* treatment. But, while a bad calibration—that set the X too high or too low relative to the value of 150 cents for charity—could cause a level shift in how often subjects choose the outside option across conditions, we stress that a bad calibration *cannot explain* why adding a zero would affect behavior. That is, the identification strategy in the Adding Study relies on comparing the rates at which the bundles are chosen with and without an additional zero, and subjects should not respond to the addition of the zero, regardless of the level of the outside option.

B. *The Adding Study*, *Main Results*

What do subjects choose in their binary decisions? Figure 1 shows results from the *Charity*/*Charity* treatment (panel A) and the *Self*/*Charity* treatment (panel B). The figure groups bundles by the number of nonzero summands, $n \in \{2, 3, 4, 5\}$, divided by the total number of summands (i.e., 4 or 5). We can see how behavior changes when a zero is added to a bundle by comparing the fraction choosing the bundle across $4/4$ and $4/5$, across $3/4$ and $3/5$, and across $2/4$ and $2/5$.[13]

Panel A of Figure 1 shows that absent selfish motives, subjects do not respond to the addition of a zero. In the *Charity*/*Charity* treatment, whether or not $n \geq 3$, and thus whether the sum of the bundle is more than 150 cents, is the key determinant in whether the bundle is selected. Comparing across bars with the same color (i.e., the same number of nonzero summands), we see that adding a zero to a bundle does not influence whether the bundle is selected.

When subjects may make motivated errors to justify selfish decisions, however, the results look different. Panel B of Figure 1 shows that subjects in the *Self*/*Charity* treatment are less likely to choose $4/5$ bundles than $4/4$ bundles, less likely to choose $3/5$ bundles than $3/4$ bundles, and less likely to choose $2/5$ bundles than $2/4$ bundles. Subjects are systematically less likely to choose a bundle when a zero is added to it. Subjects act as if they make simple addition errors when doing so can justify selfish decisions—that is, when doing so can justify choosing the outside option that benefits themselves rather than the bundle that benefits the charity.

The regressions in Table 1 confirm the results from Figure 1. Subjects are unresponsive to the addition of a zero in the *Charity*/*Charity* treatment (panel A) but

[12] This is a methodological contribution that we have also used in our other work (Exley 2016, 2020). See a discussion of the advantages of this procedure in Gauriot, Heger, and Slonim (2020).

[13] A response to the addition of a zero could relate to the affect heuristic (Gilovich, Griffin, and Kahneman 2002) and errors arising from irrelevant attributes (Chadd, Filiz-Ozbay, and Ozbay 2021).

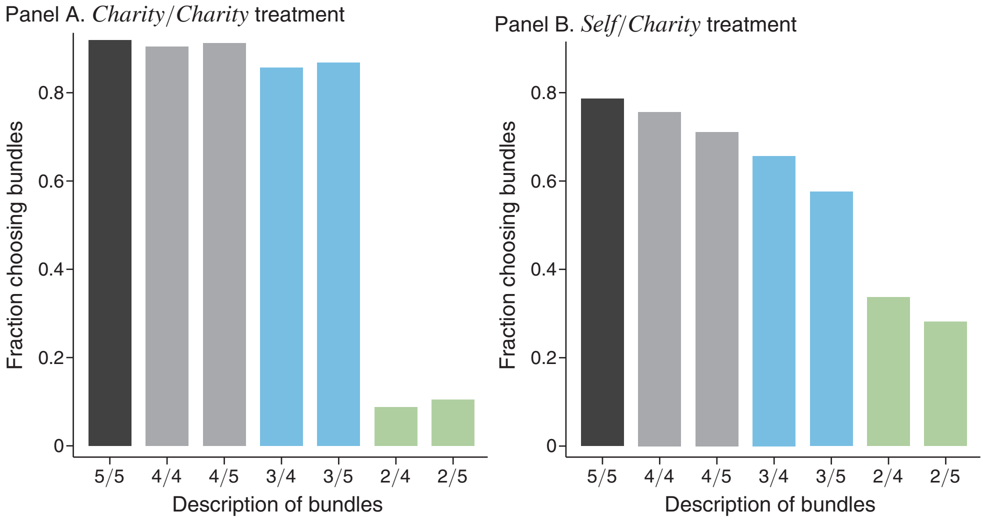

FIGURE 1. IN THE *CHARITY/CHARITY* AND *SELF/CHARITY* TREATMENTS OF THE ADDING STUDY, FRACTION CHOOSING A MAIN BUNDLE

Note: Data include all subjects' decisions in all main bundles: in the *Charity/Charity* treatment of the Adding Study in panel A and in the *Self/Charity* treatment of the Adding Study in panel B.

respond to the addition of a zero in the *Self/Charity* treatment when doing so may help them to justify selfishness (panel B). These regressions include a dummy for each baseline bundle and estimate the impact of adding a zero to the bundle, the coefficient on $(^{+}0)$, and of improving the bundle by adding a nonzero amount to it, the coefficient on $(^{+}1)$. In particular, the regression specification is

$$(1) \qquad \Pr(\mathit{choose\ bundle}) = \beta_1(^{+}0) + \beta_2(^{+}1) + \sum_{n=2}^{4}\sum_{d=51}^{59} k_n \times l_d + \epsilon,$$

where k_n are dummies for the number of nonzero amounts in the baseline bundle and l_d are dummies for the value of the nonzero amounts in the bundle.

The coefficient on $(^{+}0)$ in column 1 of panel A is near zero and insignificant, suggesting that subjects do not respond to the addition of a zero in the *Charity/Charity* treatment when selfish motives are not relevant. This coefficient remains unchanged regardless of whether the baseline bundle has no zeros (column 2) or has at least one (column 3) and whether we exclude subjects who always choose the outside option or always choose the bundle (column 4) or exclude subjects for whom we assigned an outside option of 5 cents because the lower bound of their indifference range was 0 cents (column 5).

By contrast, when subjects may make simple addition errors to justify selfish decisions, the coefficient on $(^{+}0)$ in column 1 of panel B is negative and significant, showing that subjects do respond to the addition of a zero in the *Self/Charity* treatment. Specifically, subjects in the *Self/Charity* treatment are 6 percentage points less likely to choose the bundle for charity—or 6 percentage points more likely to instead choose money for themselves—when a zero is added to the bundle, even

Table 1—In the *Charity*/*Charity* and *Self*/*Charity* Treatment of the Adding Study, Regression of Choosing a Main Bundle

Sample:	Full			Choice varies	X is lower bound
	Main bundles (1)	If 4/4 baseline (2)	If 2/4 or 3/4 baseline (3)	Main bundles (4)	Main bundles (5)
Panel A. Charity/Charity treatment					
$(^{+}0)$	0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$(^{+}1)$	0.28	0.02	0.42	0.29	0.28
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Observations	7,164	2,388	4,776	7,092	5,832
Panel B. Self/Charity treatment					
$(^{+}0)$	−0.06	−0.04	−0.07	−0.08	−0.07
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
$(^{+}1)$	0.11	0.03	0.15	0.14	0.12
	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Observations	7,128	2,376	4,752	5,616	6,048
$k_n \times l_d$ fixed effects	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are clustered at the subject level and shown in parentheses. The results are from a linear probability model of whether a subject chose a main bundle in the *Charity*/*Charity* treatment (panel A) and the *Self*/*Charity* treatment (panel B) of the Adding Study. $(^{+}0)$ is an indicator for a five-summand bundle that is constructed by adding a fifth amount that is equal to zero to a baseline bundle, $(^{+}1)$ is an indicator for a five-summand bundle that is constructed by adding a fifth amount that is nonzero to a baseline bundle, and $k_n \times l_d$ fixed effects include all possible interactions of dummies for the number of nonzero amounts within the baseline bundle (see online Appendix Table A.2) and dummies for the value of the nonzero amount d in the bundle to fully control for the sum of the amounts in the baseline bundle. Columns 1–3 analyze all subjects' decisions: in all main bundles in column 1; involving the baseline bundles with four nonzero amounts in column 2; and involving the baseline bundles with two or three nonzero amounts, which already have at least one zero, in column 3. Column 4 analyzes all main bundles but among a restricted sample of subjects who choose the bundle at least once and choose their outside option at least once. Column 5 analyzes all main bundles but among a restricted sample of subjects with outside option X set to the lower bound of their indifference range (thus excluding subjects with a zero lower bound).

though the sum has not changed. This effect is large. It is 10 percent of the likelihood of choosing a baseline bundle, which is 0.58. It is more than half the magnitude of the 11 percentage point increase observed from adding a nonzero amount to a bundle (see the coefficient estimate on $(^{+}1)$), which on average increases the total amount donated to charity by 33 percent.

The other columns of panel B show evidence of motivated errors regardless of whether the baseline bundle has no zeros (column 2) or has at least one (column 3). Evidence of motivated errors is robust to subjects who choose the bundle at least once and choose their outside option at least once (column 4) and robust to excluding subjects for whom we assigned an outside option of 5 cents because the lower bound of their indifference range was 0 cents (column 5). In addition, as shown in online Appendix Table B.4, our results persist regardless of whether subjects evaluate the 5-summand or the 4-summand bundles first and whether we focus on the first 12 or last 12 decisions within each set of 24.[14]

[14] That evidence of motivated errors persists regardless of order is evidence against cognitive dissonance or consistency influencing subjects' decisions to a substantial degree (Cialdini 1984; Bazerman, Loewenstein, and White 1992; Babcock et al. 1995; Konow 2000; Haisley and Weber 2010; Gneezy et al. 2012; Bohnet, van Geen, and

Online Appendix Table B.1 shows that the coefficients on $(^{+}0)$ are statistically significantly different in the *Self*/*Charity* and *Charity*/*Charity* treatments. We also find significant differences in the likelihood of responding to the addition of a zero across treatments at the individual level. First, the fraction of subjects who respond to the addition of the 0 at least once is 50 percent in the *Self*/*Charity* treatment, but it is only 26 percent in the *Charity*/*Charity* treatment ($p < 0.01$). Second, the fraction of subjects who respond to the addition of a 0 in at least two bundles with different sums is 29 percent in the *Self*/*Charity* treatment but only 14 percent in the *Charity*/*Charity* treatment ($p < 0.01$). This latter result shows that subjects' decisions do not simply reflect the use of an additional zero as a way to break indifference between the bundle and the outside option since they could not be indifferent between their outside option and two bundles with different total amounts. Third, we find that the fraction of subjects who display nonmonotonic behavior (with regard to the sum of donations made by a bundle) is 37 percent in the *Self*/*Charity* treatment but is only 25 percent in the *Charity*/*Charity* treatment ($p < 0.01$).[15] As expected, this nonmonotonic classification is significantly correlated with whether subjects respond to the addition of a zero.

When selfish motives are not relevant—as in the *Charity*/*Charity* treatment—subjects do not respond to the addition of a zero. But in the *Self*/*Charity* treatment, when subjects can justify selfishness by appealing to the possibility of a simple addition error, subjects choose a bundle less often when a zero is added to it. That is, subjects make *motivated errors*.

C. *The Adding Study*, *Additional Results Absent Selfish Motives*

In this section, we provide two additional pieces of evidence that—in the absence of selfish motives—subjects do not make errors when a zero is added to a bundle.

First, we highlight that subjects are able to correctly sum the amounts in the bundle—and that their ability to correctly sum the numbers does not depend on adding a zero—lending additional credence to the notion that they are using the additional zero as an excuse to be selfish. In a supplemental study, the *Calculation* study, subjects are shown 12 of our main bundles (6 four-summand bundles and the corresponding 6 five-summand bundles constructed by adding a 0 to them) and are asked to report the sum of each bundle.[16] To incentivize accuracy, subjects are

Bazerman 2016; Falk and Zimmermann 2017, 2018; Golman et al. 2016; Gneezy et al. 2020; Gneezy, Saccardo, and van Veldhuizen 2018). Of course, the number of decisions that subjects make in our study may minimize the role of these mechanisms since subjects may be less constrained by prior decisions if they cannot fully remember those prior decisions (Bénabou and Tirole 2002; Zimmermann 2020).

[15] Our measure of nonmonotonicity uses decisions from four non-main bundles that are denoted as $4^L/4$*-bundles* because all four amounts are nonzero, but each amount is smaller than the amounts in the main bundles. The nonzero amounts in these bundles are randomly selected to be d^L cents, where $d^L \in \{30, 31, 32, 33, 34, 35, 36, 37, 38\}$ (for more details about these bundles, see online Appendix Table A.3). These bundles were constructed so that the sum of each bundle was close to, but lower than, the sum of each $3/4$*-bundle* and each $3/5$*-bundle* (i.e., $3 \times d > 4 \times d^L$ for all d and d^L). Thus, we call a subject nonmonotonic if the subject chooses one or more $4^L/4$*-bundles* and fails to choose all of the $3/4$*-bundles* and $3/5$*-bundles*. While we could construct other measures of nonmonotonic behavior, even among this set of 16 bundles, this measure seems particularly natural since it utilizes bundles designed to be close in sum to our main bundles but with significantly lower individual donation amounts.

[16] In October 2019, we recruited 100 subjects from Amazon's Mechanical Turk to complete the *Calculation* study. Each subject received \$3 for completing the 20-minute study as well as any additional bonus payments from

told that 1 bundle will be randomly selected and the subject will receive a 25 cent bonus if they correctly sum the amounts in that bundle. Nearly all calculations are accurate: 98 percent and 99 percent of the calculations involving the 4-summand and 5-summand bundles, respectively, are answered correctly (the 1 percentage point difference is not statistically significant). Subjects are able to sum the bundles, regardless of whether the bundle includes an additional zero.

Second, we find that in a treatment that removes selfish motives but where subjects' own money is still at stake, subjects also do not respond to the addition of a zero. In the *Self*/*Self* treatment, the subject is the recipient of both the outside option (set to 150 cents for the subject) and the sum of the amounts in the bundle. Subjects who want to maximize earnings in this treatment should choose the bundle whenever its sum is greater than 150 cents. As shown in online Appendix Figure B.2 and online Appendix Table B.2, subjects do not respond to the addition of a zero in the *Self*/*Self* treatment.[17]

D. *The Adding Study*, *Additional Results on Decreasing the Scope for Confusion*

To explore the impact of reducing the plausible scope for being confused or making addition errors, we provide subjects with more information on the bundle's sum in three additional treatments that are built off of the *Self*/*Charity* treatment: the *Self*/*Charity* (*sum optional*), *Self*/*Charity* (*sum shown*), and *Self*/*Charity* (*sum unavoidable*) treatments.[18] Online Appendix Table B.3 presents the corresponding results (column 1 reproduces results from the *Self*/*Charity* treatment for reference).

In the *Self*/*Charity* (*sum optional*) treatment, subjects have the option to click a button to reveal the bundle's sum (see panel B of online Appendix Figure A.2). As shown in column 2 of online Appendix Table B.3, adding a zero to a bundle still statistically significantly decreases subjects' willingness to choose a bundle by 3 percentage points in the *Self*/*Charity* (*sum optional*) treatment. However, this change to the decision environment proves somewhat effective at mitigating motivated errors: this 3 percentage point effect is significantly smaller ($p < 0.1$) than the 6 percentage point effect observed in the *Self*/*Charity* treatment.

Motivated errors may not be fully eliminated in the *Self*/*Charity* (*sum optional*) treatment because subjects simply do not choose to click to reveal the sum and so can still appeal to the possibility of not knowing it. Indeed, subjects in that treatment only clicked to reveal the sum 58 percent of the time. To further reduce this possibility, the *Self*/*Charity* (*sum shown*) treatment displays the sum on each decision screen (see panel C of online Appendix Figure A.2). This change proves just as effective—but not more effective—at reducing motivated errors as the option to

one randomly selected decision. As with our other studies run in 2019, workers must have previously completed at least 100 human intelligence tasks with a 95 percent or better approval rating and must have been working from a United States IP address.

[17] We view a comparison between the *Self*/*Charity* treatment and *Charity*/*Charity* treatment as cleaner than a comparison between the *Self*/*Charity* treatment and *Self*/*Self* treatment because the latter also requires us to vary the recipient of the bundles.

[18] We consider adjustments to the decision environment that are common debiasing strategies. For reviews, see Conlisk (1996); Rabin (1998); DellaVigna (2009); Madrian (2014); and Gabaix (2019). For related examples, see List (2003); Chetty, Looney, and Kroft (2009); Finkelstein (2009); Brocas et al. (2014); Hanna, Mullainathan, and Schwartzstein (2014); Schwartzstein (2014); Taubinsky and Rees-Jones (2018); and Enke (2020).

reveal the sum. As shown in column 3 of online Appendix Table B.3, adding a zero to a bundle still significantly decreases subjects' willingness to choose a bundle by 3 percentage points. But, again, this 3 percentage point effect is significantly smaller ($p < 0.1$) than the 6 percentage point effect observed in the *Self/Charity* treatment.

It is possible that motivated errors are not eliminated in the *Self/Charity* (*sum shown*) treatment because subjects can still appeal to the possibility of not knowing the sum, even though it is displayed on each decision screen (e.g., there is no guarantee the subject saw or attended to the sum, even though it was shown). In the *Self/Charity* (*sum unavoidable*) treatment, we make the sum salient on the decision screen (see panel D of online Appendix Figure A.2) and require subjects to correctly report the bundle's sum before making the associated decision (see online Appendix Figure D.9). This combination proves successful at eliminating motivated errors. In particular, as shown in column 4 of online Appendix Table B.3, adding a zero to a bundle no longer statistically significantly influences subjects' willingness to choose a bundle. The impact of adding a zero is 4 percentage points smaller ($p < 0.05$) in the *Self/Charity* (*sum unavoidable*) treatment than the effect observed in the *Self/Charity* treatment.

The above results reveal that we only succeed at eliminating motivated errors when it is unambiguous to subjects—and to the experimenter—that they know the sum (e.g., because they are required to accurately report it back as part of the study). These results highlight the potential difficulty with constraining individuals' ability to appeal to the possibility of being confused or making mistakes to justify selfishness.

E. *The Adding Study*, *Additional Results on Information Avoidance*

While the results from Section ID consider the impact of reducing the plausible scope for being confused, increasing the plausible scope for being confused begets new questions. For instance, recall that subjects were required to reveal all summands in a bundle before making a decision in the *Self/Charity* treatment. This implies that—while subjects could appeal to being confused and not knowing the sum to justify selfish decisions—subjects could *not* appeal to lacking the information necessary to determine the sum. In this way, while conceptually similar, the evidence for motivated errors in the *Self/Charity* treatment is different than work on information avoidance in which—as in Dana, Weber, and Kuang (2007) and the rich literature that followed it—subjects can avoid information needed to determine the payoff consequences of their decisions.

In the *Self/Charity* (*summands optional*) treatment, subjects could avoid information needed to determine the payoff consequences of their decisions because they were *not* required to reveal all summands in a bundle before making a decision. The results in online Appendix Table C.1 show that subjects in the *Self/Charity* (*summands optional*) treatment choose a bundle less often when a zero is added to it, even after they have chosen to acquire information on each summand in that bundle. That is, while one could speculate that evidence for motivated errors would only arise when subjects would have exploited information avoidance as an excuse if given the opportunity to do so, these results show that this is not the case. See online Appendix C.C1 for more details on the design and results of the *Self/Charity* (*summands optional*) treatment. See online Appendix C.C2 for additional results, including also a discussion of the *Charity/Charity* (*summands optional*) treatment.

II. The Correlation Neglect Study

In the Correlation Neglect Study, subjects are asked to make predictions about the average of correlated signals. When selfish motives are not relevant, there is evidence for correlation neglect: subjects overestimate the truth when a high signal is correlated with their other signals and underestimate the truth when a low signal is correlated with their other signals. When selfish motives are relevant, we observe evidence for motivated errors. In particular, the extent to which we observe correlation neglect aligns with subjects making errors—or appealing to the possibility of being confused when making predictions about the average of correlated signals—to justify predictions that align with their selfish motives.

A. *The Correlation Neglect Study*, *Design*

The Correlation Neglect Study included 1,200 subjects from MTurk who participated in one of three treatments.[19]

In Part 1 of the study, each subject completed the same multiple price list as in the Adding Study to elicit an X value such that the subject is indifferent between X cents for themselves and 150 cents for the Make-A-Wish Foundation.[20] In Part 2, building off of the design in Enke and Zimmermann (2019), subjects answer ten questions in which they are asked to predict a number when information on the number is presented in a correlated way. In particular, subjects are asked to predict a number that equals the average of four estimates: Estimate 1, Estimate 2, Estimate 3, and Estimate 4. They are asked about their prediction using a slider that selects a range of five numbers on the support of 0 to 100.

Prior to making this prediction, subjects are directly informed of Estimate 1 and are informed of the average of Estimates 1 and 2 (as the output of "Channel 1 News"), the average of Estimates 1 and 3 (as the output of "Channel 2 News"), and the average of Estimates 1 and 4 (as the output of "Channel 3 News"). Figure 2 shows an example of how the correlated information is presented to subjects.

For five of the ten questions (i.e., the odd-numbered questions in Table 2), Estimate 1 is the smallest of the four estimates; we call these the *Low Estimate* 1 questions. For the other five questions (i.e., the even-numbered questions in Table 2), Estimate 1 is the largest of the four estimates; we call these the *High Estimate* 1 questions.

To cleanly test for the impact of a low Estimate 1 versus a high Estimate 1, the questions were formulated as pairs. Both questions in a pair (e.g., Q1 and Q2) have

[19] In July 2021, we recruited and randomized 1,200 subjects from MTurk into one of three study treatments. See the notes of online Appendix Table A.1 for details on eligibility conditions, which also applied to Study 2. Full instructions for Study 2 can be found in online Appendix D.D2. Each subject received \$3.50 for completing the 20-minute study. In addition, one randomly selected decision for each subject was implemented for bonus payment and resulted in an additional payment for the subject and/or a donation to charity.

[20] As before, they complete a multiple price list that generates an indifference range for X. We assign subjects an X value equal to the lower bound of their indifference range unless the lower bound of the indifference range is 0, in which case we assign $X = 5$ cents. The distribution of X values are displayed in panel B of online Appendix Figure B.1.

In this question, the number equals the average of four estimates: Estimate 1, Estimate 2, Estimate 3, and Estimate 4. You will be directly informed of Estimate 1. You will also be informed of the reports about the estimates from three news channels. Channel 1 News reports the average of Estimate 1 and Estimate 2. Channel 2 News reports the average of Estimate 1 and Estimate 3. Channel 3 News reports the average of Estimate 1 and Estimate 4. Specifically, note that:

- Estimate 1 is **3**.
- Channel 1 News reports **27.5**.
- Channel 2 News reports **28.5**.
- Channel 3 News reports **51**.

Figure 2. Example of How Correlated Information Is Presented

Table 2—Questions in the Correlation Neglect Study

Q number	Correct (correlated) answer	Estimate (1)	Estimate (2)	Estimate (3)	Estimate (4)	News channel (1)	News channel (2)	News channel (3)
1	50 (29)	8	40	59	93	24	33.5	50.5
2	50 (71.5)	93	8	40	59	50.5	66.5	76
3	47 (24.5)	2	43	50	93	22.5	26	47.5
4	47 (70)	93	2	43	50	47.5	68	71.5
5	54 (29)	4	59	55	98	31.5	29.5	51
6	54 (76)	98	4	59	55	51	78.5	76.5
7	49 (27)	5	51	48	92	28	26.5	48.5
8	49 (70.5)	92	5	51	48	48.5	71.5	70
9	52 (27.5)	3	52	54	99	27.5	28.5	51
10	52 (75.5)	99	3	52	54	51	75.5	76.5

Notes: This table shows the estimates and the reports of the news channels for each of the ten questions in the Correlation Neglect Study. The questions are constructed in pairs with the same correct answer but a different Estimate 1, which is then averaged with the other estimates to generate the news channel reports. The "Correct (correlated) answer" column shows the correct answer first and then the correlated answer in parentheses, the latter of which is the average of Estimate 1 and the reports of the three news channels (i.e., what an agent would guess if they fully ignored the correlated nature of the information and treated each signal as fully independent). Each subject sees these ten questions in a randomized order.

the same set of four estimates, and thus the same correct answer, but they either use the lowest estimate as Estimate 1 (as in Q1), so correlation neglect would pull responses down, or the highest estimate as Estimate 1 (as in Q2), so correlation neglect would pull responses up.

To measure the extent of correlation neglect, we can estimate how far subjects' answers are from the truth in each question. We can also compare responses in the *Low Estimate* 1 questions to responses in the corresponding *High Estimate* 1 questions, which allows us to measure the difference in subjects' responses when the (same) information is presented such that correlation neglect pulls the estimate down rather than up.

Subjects answer these ten questions in one of three treatments. In the *Control* condition, the charity always receives 150 cents, regardless of whether the subject provides an underestimate (i.e., selects a range that is too low), provides an overestimate (i.e., selects a range that is too high), or provides a correct answer (i.e., selects a range that includes the true number). In addition, the subject receives *X* cents if they provide a correct answer, giving them a private incentive to provide a correct answer.

In the *Underestimate* treatment, we replace the payment associated with underestimating to be a bonus of *X* cents for the subject rather than 150 cents for charity. Equivalently, in the *Overestimate* treatment, we replace the payment associated with overestimating to be a bonus of *X* cents for the subject rather than 150 cents for charity.

Since we elicit an *X* value such that the subject is indifferent between *X* cents for themselves and 150 cents for the Make-A-Wish Foundation, one could expect that we observe no differences in predictions across our treatments. Alternatively, one could expect differences to emerge. We elicited *X* via a multiple price list that required subjects to directly choose between money for themselves and money for charity, and thus in an environment in which there may be little room for excuses that could alleviate image costs associated with being selfish. When making predictions about the average of correlated information, however, image costs associated with being selfish could be alleviated by appealing to the possibility of being confused. In particular, subjects may appeal to the possibility of being confused to justify predicting a lower number in the *Underestimate* treatment, when underestimating secures them *X* cents for themselves (rather than 150 cents for charity). Similarly, subjects may appeal to the possibility of being confused to justify predicting a higher number in the *Overestimate* treatment when overestimating secures them *X* cents for themselves (rather than 150 cents for charity).

Thus, the extent to which we observe correlation neglect may align with subjects making such motivated errors. In particular, if subjects make motivated errors, correlation neglect may be exacerbated in some cases (i.e., *Low Estimate* 1 questions in the *Underestimate* treatment and *High Estimate* 1 questions in the *Overestimate* treatment) but instead mitigated in other cases (i.e., *Low Estimate* 1 questions in the *Overestimate* treatment and *High Estimate* 1 questions in the *Underestimate* treatment).

B. *The Correlation Neglect Study*, *Results*

Table 3 presents regression results about the size of the errors subjects make (where error is defined as the midpoint of the subject's range on the slider minus the correct answer for that question, which allows us to analyze data from questions with different correct answers in the same regression). The regressions include fixed effects for each question pair and estimate the impact of being a *Low Estimate* 1 question. Since *High Estimate* 1 questions are the excluded group, the regression estimates the extent to which subjects display correlation neglect by comparing how they answer when correlation neglect pulls their answer down versus when it pulls their answer up.

Column 1 presents results from the *Control* condition. The suppressed fixed effects indicate that subjects significantly overestimate the answer when the question involves a *High Estimate* 1. The coefficient estimates on the indicator for it being a *Low Estimate* 1 question show that subjects provide significantly lower

Table 3—Regression of Errors in the Correlation Neglect Study

Sample:		When selfish motives should	
	Control	Exacerbate bias	Mitigate bias
	(1)	(2)	(3)
Low Estimate 1	−19.50	−24.81	−7.70
	(1.03)	(1.37)	(1.43)
Observations	4,080	3,960	3,960
Question pair fixed effects	Yes	Yes	Yes

Notes: Standard errors are clustered at the subject level and shown in parentheses. The results are from an OLS regression of the true answer to a question minus the answer provided by a subject in that question (i.e., the "error") in the Correlation Neglect Study. *Low Estimate* 1 is an indicator for *Low Estimate* 1 questions (i.e., the odd-numbered questions in Table 2). Column 1 analyzes answers from the *Control* condition. Column 2 analyzes questions in which selfish motives are expected to exacerbate evidence for correlation neglect (i.e., *Low Estimate* 1 questions in the *Underestimate* treatment and *High Estimate* 1 questions in the *Overestimate* treatment). Column 3 analyzes questions in which selfish motives are expected to mitigate evidence for correlation neglect (i.e., *High Estimate* 1 questions in the *Underestimate* treatment and *Low Estimate* 1 questions in the *Overestimate* treatment).

answers—by an average of 19.50—when the question involves a *Low Estimate* 1 rather than a *High Estimate* 1. Consistent with correlation neglect, subjects provide answers that are substantially lower (nearly 20 points lower on a 0–100 support) when Estimate 1 is low rather than high. This pattern is also evident in panel A of Figure 3, which shows the distributions of the errors in the *Control* condition.[21]

Column 2 of Table 3 presents results from the same specification as column 1 but looking at cases when motivated errors could exacerbate correlation neglect. This arises when subjects in the *Overestimate* treatment answer a *High Estimate* 1 question; knowing that they secure money for themselves if they overestimate the answer, they may appeal to the possibility of being confused to justify choosing a higher answer. This also arises when subjects in the *Underestimate* treatment answer a *Low Estimate* 1 question; knowing that they secure money for themselves if they underestimate the answer, they may appeal to the possibility of being confused to justify choosing a lower answer. The coefficient estimate of 24.81 on the indicator for *Low Estimate* 1 shows that, in these cases, evidence for correlation neglect is indeed larger in magnitude. The full distributions of subject responses when motivated errors could exacerbate correlation neglect can be seen in panel B of Figure 3. Column 1 of online Appendix Table B.5 compares this coefficient to the coefficient in the *Control* treatment and finds that they are statistically significantly different ($p < 0.01$).

Column 3 of Table 3 presents results from the same specification as column 1 but looking at cases when motivated errors could instead mitigate correlation neglect. This arises when subjects in the *Overestimate* treatment answer a *Low Estimate* 1

[21] Column 1 of online Appendix Table B.7 provides an alternative specification that reports the extent to which subjects overestimate the truth in *High Estimate* 1 questions, reported as the coefficient on *Bias Pulls Up*, and how much they underestimate the truth in *Low Estimate* 1 questions, reported as the coefficient on *Bias Pulls Down*.

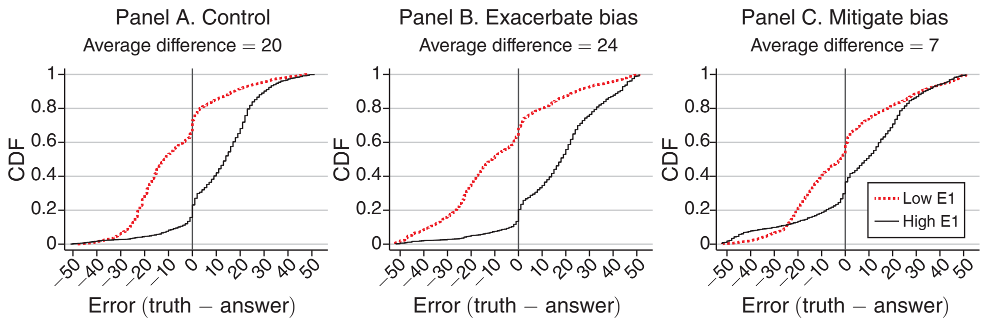

Figure 3. CDFs of Errors in Correlation Neglect Study

Notes: This figure shows CDFs of errors (the true answer to a question minus the answer provided by a subject in that question) in the Correlation Neglect Study. *Low E*1 refers to questions with a low Estimate 1 (i.e., the odd-numbered questions in Table 2). *High E*1 refers to questions with a high Estimate 1 (i.e., the even-numbered questions in Table 2).

question and when subjects in the *Underestimate* treatment answer a *High Estimate* 1 question. The coefficient estimate of 7.70 on the indicator for *Low Estimate* 1 shows that subjects still display correlation neglect but to a lesser extent. This is also evident by comparing panel C to panel A in Figure 3. Column 1 of online Appendix Table B.5 compares this coefficient to the coefficient in the *Control* treatment and finds that they are statistically significantly different ($p < 0.01$).

Online Appendix Table B.5 presents a series of robustness checks and shows that the results prove robust. In each column, the coefficient on *Low Estimate* 1 is negative and significant, demonstrating evidence of correlation neglect in the *Control* condition. In addition, in each of the columns, consistent with subjects making motivated errors, the coefficient on *Low Estimate* 1 × *Exacerbate Bias* is negative and significant, and the coefficient on *Low Estimate* 1 × *Mitigate Bias* is positive and significant. Columns 2–6 show that the results are robust to separately considering each pair of questions. Column 7 shows that the results are robust to excluding subjects with an outside option that could be less than 5 cents or could be more than 150 cents. Columns 8 and 9 show the results are robust to considering only the first five or only the last five questions subjects answer, respectively.

Motivated by whether measurement error in how we estimate subjects' *X* values could generate our results, we conclude with a few observations.[22] First, in all treatments, subjects secure the highest payoff—150 cents for charity *and X* cents for themselves—when they provide the correct answer. All of our treatment differences are driven by varying whether subjects only secure 150 cents for charity *or* only secure *X* cents for themselves when their answer is too low or too high. This design pushes against us finding any treatment differences, regardless of how *X* is measured, since subjects have a clear incentive to provide a correct answer (i.e., given this incentive for accuracy, the fact that responses change in our treatment

[22] We also emphasize that—as detailed in Section IA—our Adding Study documents evidence for motivated errors in a simple paradigm where we do not rely on *X* for identification.

conditions is notable).[23] Second, symmetric measurement error in X should wash out and not drive our treatment effects.[24] Third, if one were to worry about asymmetric measurement error where subjects systematically prefer X for themselves to 150 cents for charity, we note that we pick X to guard against this concern. Specifically, as detailed in footnote 20, we set X, when possible, to the implied lower bound of the indifference range between X cents for the subject and 150 cents for charity.[25] Fourth, if one suspects that subjects *actually* prefer X cents for themselves to 150 cents for charity in the absence of image concerns—and that our results are driven by these image concerns being alleviated when subjects make predictions because they can appeal to being confused to justify selfishness—we note that this is essentially a restatement of our preferred explanation. Given the similarity in design, these same arguments also apply to the Anchoring Studies described next.

III. The Anchoring Studies

In the Anchoring Studies, we examine how subjects' answers to knowledge-based questions are influenced by whether a random anchor is high or low (for seminal examples of anchoring, see Tversky and Kahneman 1974 and Ariely, Loewenstein, and Prelec 2003; for a recent meta-study, see Ioannidis, Offerman, and Sloof 2020). When selfish motives are not relevant, we observe evidence of an anchoring bias (i.e., an assimilation effect): when the random anchor is high, responses are pulled up; when the random anchor is low, responses are pulled down. When selfish motives are relevant, however, evidence emerges for motivated errors. In particular, the extent to which we observe an anchoring bias aligns with subjects making errors—or appealing to the possibility of being confused when answering these questions—to justify answers that align with their selfish motives.

A. *The Anchoring Studies, Design*

A total of 1,195 individuals participated in one of three treatments of Anchoring Study A, and another 1,192 individuals participated in one of the three treatments of Anchoring Study B.[26]

[23] To the extent that subjects respond to our treatment conditions, this likely involves them trading off their perceived probability of getting the answer correct to increase their perceived probability of ending up in a particular region (e.g., receiving X cents for themselves rather than 150 cents for charity) conditional on being incorrect.

[24] Under the null that subjects are not using uncertainty as an excuse for selfishness, if there is not measurement error, so subjects value X cents for themselves as equivalent to 150 cents for charity, subjects have no extra incentive to underestimate or overestimate in any treatment because they always secure X cents for themselves or 150 cents for charity when they provide an answer that is too high or too low. If X is measured with noise around zero, (e.g., if subjects are equally likely to have an estimated X valued at 150 cents $+ \epsilon$ or 150 cents $- \epsilon$ for charity), subjects with X valued at 150 cents $+ \epsilon$ will have an incentive to hedge their guess in the direction where X has replaced 150 cents for charity, while subjects with X valued at 150 cents $- \epsilon$ will have an incentive to hedge their guess in the opposite direction.

[25] Under the null that subjects are not using uncertainty as an excuse for selfishness, setting X at the lower bound pushes against us finding the treatment effects we observe.

[26] Subjects were recruited and randomized into one of the three study treatments for Anchoring Study A in December 2021 and for Anchoring Study B in January 2022. See the notes of online Appendix Table A.1 for details on eligibility criteria, which also applied to the Anchoring Studies. Full instructions for Anchoring Study A and Anchoring Study B can be found in online Appendixes D.D3 and D.D4, respectively. Each subject received $2 for completing the 10-minute study. In addition, one randomly selected decision for each subject was implemented for bonus payment and resulted in an additional payment for the subject and/or a donation to charity.

The Anchoring Studies followed the same structure as the Correlation Neglect Study in a very similar paradigm. In Part 1, all subjects complete the same multiple price list as in the other studies to elicit an X value for each subject. In Part 2, building off of the design in Enke et al. (2021), subjects respond to four questions in which they are asked to guess an answer after being provided with a random, uninformative low or high anchor.[27] As shown in online Appendixes D.D3 and D.D4, we use the same four questions as Enke et al. (2021), which asks trivia questions about (i) how many minutes it takes for light to travel from the sun to Jupiter, (ii) how many days it took to complete the first airplane trip across the continental United States, (iii) the number of millions of people living in Uzbekistan, and (iv) the weight (in hundred of tons) of the Eiffel tower.

We use two different anchoring manipulations in the Anchoring Studies. In Anchoring Study A, subjects are first asked to consider whether the answer is above or below a number. Subjects are told that this number (i.e., the anchor) is randomly selected to be either 20 or 80. Subjects are then asked to select their precise answer to the question on a slider from 0 to 100, and the default answer on the slider is set to the anchor. Subjects have 15 seconds to answer each question and are told that the default answer will be implemented if they do not select a different answer themselves.

In Anchoring Study B, subjects are first asked whether the answer is above or below a number. As in Anchoring Study A, subjects are told that the anchor is randomly selected to be either 20 or 80. In Anchoring Study B, however, subjects must provide an answer (i.e., they must guess whether the answer to the question is above or below the anchor). Then, subjects are asked to select the answer on a slider from 0 to 100 with the default position of the slider set to the midpoint of the support (i.e., to 50 out of 100). Subjects are given as much time as they would like to answer the question, and they are required to answer this question to continue with the study.

To summarize, the main differences between the two anchoring manipulations are as follows. Anchoring Study A provides a visual anchor (since the slider is set to the anchor), and the anchor is an explicit default (since the anchor is implemented in the event that a subject does not answer the question within the designated time). By contrast, Anchoring Study B does not provide a visual anchor or use a default and instead requires subjects to answer the anchoring question before providing a precise answer to the main question that determines payoffs. We provided a time limit in Anchoring Study A to discourage subjects from trying to search online for the correct answer to the question. The time limit required us to have a default answer to implement if the subject did not make a choice within the designated time. We ran Anchoring Study B to more closely follow the procedures in Enke et al. (2021) and to investigate a setting in which the anchor was not the default answer. Despite some mild cause for concern that subjects might search for correct answers online in Anchoring Study B (answers are correct only

[27] Since the meta-study in Ioannidis, Offerman, and Sloof (2020) reveals that uninformative anchors do not always result in an anchoring bias, we closely followed the decision context of Enke et al. (2021), in which an anchoring bias is known to induce an assimilation effect (rather than a contrast effect). For a discussion of the differences between assimilation and contrast effects, see Herr, Sherman, and Fazio (1983) and Bordalo, Gennaioli, and Shleifer (2020).

TABLE 4—REGRESSION OF ERRORS IN THE ANCHORING STUDIES

Sample:		When selfish motives should	
	Control (1)	Exacerbate bias (2)	Mitigate bias (3)
Panel A. Anchoring Study A			
Low Anchor	−19.93	−27.91	−9.11
	(1.47)	(1.61)	(1.79)
Observations	1,592	1,588	1,600
Panel B. Anchoring Study B			
Low Anchor	−14.35	−21.36	−5.01
	(1.52)	(1.96)	(1.94)
Observations	1,584	1,618	1,562
Question fixed effects	Yes	Yes	Yes

Notes: Standard errors are clustered at the subject level and shown in parentheses. The results are from OLS regressions of the true answer to a question minus the answer provided by a subject in that question (i.e., the "error") in Anchoring Study A in panel A and in Anchoring Study B in panel B. *Low Anchor* is an indicator for questions with a low anchor. Column 1 analyzes answers from the *Control* condition. Column 2 analyzes questions in which selfish motives are expected to exacerbate evidence for the anchoring bias (i.e., questions with low anchors in the *Underestimate* treatment and questions with high anchors in the *Overestimate* treatment). Column 3 analyzes questions in which selfish motives are expected to mitigate evidence for the anchoring bias (i.e., questions with high anchors in the *Underestimate* treatment and questions with low anchors in the *Overestimate* treatment).

5 percent of the time in Anchoring Study A but are correct 13 percent of the time in Anchoring Study B), results are very similar and are quite robust across both studies.

Following the design of the Correlation Neglect Study, subjects in the Anchoring Studies are randomly assigned to a *Control* treatment, an *Underestimate* treatment, or an *Overestimate* treatment. Following Enke et al. (2021), answers are considered right if they are no more than two different from the truth. The payoffs from providing an answer that is correct, is too low, or is too high depend on treatment in exactly the same way as detailed in the Correlation Neglect Study (see Section IIA).

B. *The Anchoring Studies*, *Results*

Table 4 presents regression results about the size of the errors subjects make (where error is defined as the midpoint of the subject's range on the slider minus the correct answer for that question, which allows us to analyze data from questions with different correct answers in the same regression). Panel A shows the results from Anchoring Study A, and panel B shows the results from Anchoring Study B. The regressions include fixed effects for each question and estimate the impact of having a *Low Anchor* for that question. Since having a high anchor is the excluded group, the regression estimates the extent of anchoring bias subjects display by comparing how they answer when the anchor pulls their answer down versus when the anchor pulls it up.

Column 1 presents results from the *Control* condition. The negative coefficient estimates on *Low Anchor* reveal that subjects provide significantly lower

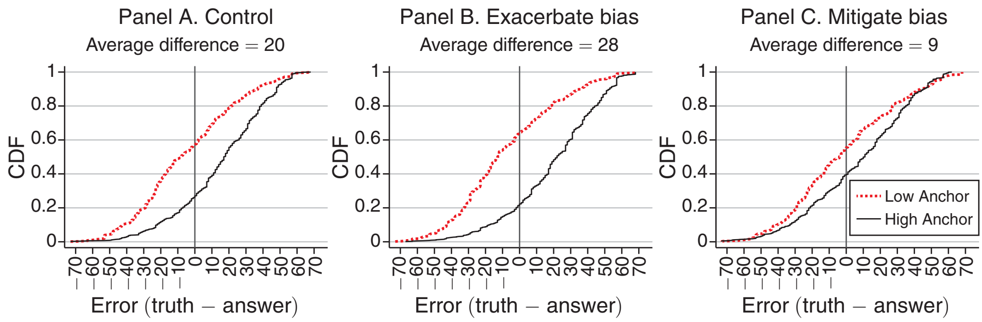

Figure 4. CDFs of Errors in Anchoring Study A

Notes: This figure shows CDFs of errors (the true answer to a question minus the answer provided by a subject in that question) in Anchoring Study A. *Low Anchor* refers to questions with a low anchor. *High Anchor* refers to questions with a high anchor.

answers—by an average of 19.93 in Anchoring Study A and an average of 14.35 in Anchoring Study B—when the question involves a low anchor rather than a high anchor, clear evidence of an anchoring bias. This pattern is also evident in panel A of Figures 4 and 5, which show the distributions of the errors in the *Control* condition for each study.[28]

Following the structure of Table 3, column 2 of Table 4 presents results from the same specification as column 1 but looking at cases when motivated errors could exacerbate an anchoring bias. This arises when subjects in the *Overestimate* treatment face a high anchor and when subjects in the *Underestimate* treatment face a low anchor. The coefficient estimates on *Low Anchor* of 27.91 in Anchoring Study A and 21.36 in Anchoring Study B show that in these cases, subjects display an anchoring bias that is larger in magnitude than in the *Control* treatments. This is also evident by comparing panel B to panel A in Figures 4 and 5. Column 1 of online Appendix Table B.6 compares these coefficients to the coefficients from the *Control* treatments of the two Anchoring Studies and finds that they are statistically significantly larger ($p < 0.01$ for both).

Column 3 of Table 4 presents results from the same specification as column 1 but looking at cases when motivated errors could mitigate an anchoring bias. This arises when subjects in the *Overestimate* treatment face a low anchor and when subjects in the *Underestimate* treatment face a high anchor. The coefficient estimate on *Low Anchor* of 9.11 in Anchoring Study A and 5.01 in Anchoring Study B show that subjects still display an anchoring bias, but one that is smaller in magnitude than in the *Control* treatments. This is also evident by comparing panel C to panel A in Figures 4 and 5. Column 1 of online Appendix Table B.6 compares these coefficients to the coefficients from the *Control* treatments of the two Anchoring Studies and finds that they are statistically significantly smaller ($p < 0.01$ for both).

[28] Online Appendix Table B.7 provides an alternative specification that reports the extent to which subjects overestimate the truth when facing a high anchor, reported as the coefficient on *Bias Pulls Up* in columns 2 and 3 (for Anchoring Studies A and B, respectively). In these studies, subjects still overestimate the truth on average when facing a low anchor (as can be seen from the positive coefficient on *Bias Pulls Down* in columns 2 and 3), although, as can be seen in Table 4, the extent of the overestimate is dramatically reduced when the anchor is low.

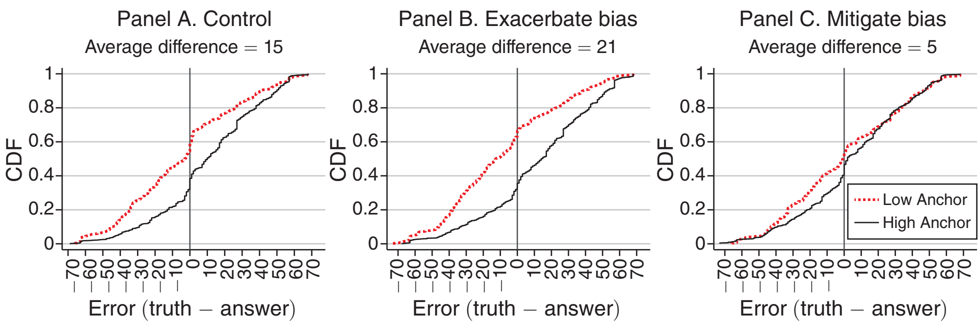

FIGURE 5. CDFS OF ERRORS IN ANCHORING STUDY B

Notes: This figure shows CDFs of errors (the true answer to a question minus the answer provided by a subject in that question) in Anchoring Study B. *Low Anchor* refers to questions with a low anchor. *High Anchor* refers to questions with a high anchor.

Online Appendix Table B.6 presents the same set of robustness checks as those shown in online Appendix Table B.5. Results from Anchoring Study A are shown in panel A, and results from Anchoring Study B are shown in panel B. Our findings prove robust. In each of the columns, the coefficient on *Low Anchor* is negative and significant, demonstrating evidence of anchoring in the *Control* condition. In addition, consistent with the role of motivated errors, the coefficient on *Low Anchor* × *Exacerbate Bias* is negative and significant in nearly all of the regressions, and the coefficient on *Low Anchor* × *Mitigate Bias* is positive and significant in all of the regressions.[29] Columns 2–5 show that the results are robust to separately considering each question. Column 6 shows that the results are robust to excluding subjects with an outside option that could be less than 5 cents or could be more than 150 cents. Columns 7 and 8 show that the results are robust to considering only the first two or only the last two questions subjects answer, respectively.

IV. Conclusion

Across studies involving over 5,000 subjects, we investigate evidence for motivated errors in both a simple environment in which individuals may make addition errors and in more complex environments in which individuals may display well-established behavioral biases.

In the simple environment (in our Adding Study), when choosing between a payoff for charity and a payoff for themselves, individuals are less likely to choose a payoff for charity when a zero is added to it. But when selfish motives are removed, the addition of a zero does not affect their choices. They only act as if they suffer from an addition error when doing so can help justify a selfish decision. This motivated error proves difficult to counter and is only eliminated when the environment dramatically mitigates the possibility of confusion about the sum going to charity.

[29] The three coefficient estimates on *Low Anchor* × *Exacerbate Bias* that are not statistically significant are directionally negative and arise in cases where we are investigating a subset of our data.

In more complex environments, subjects answer questions about the average of correlated information (in our Correlation Neglect Study) or about some fact after observing an anchor (in our Anchoring Studies). When they face selfish motives to provide certain answers, they may justify selfishness by appealing to the possibility of being confused when providing their answers. We observe evidence consistent with such motivated errors.

In light of our results, we make several observations that may be informative for future work. First, since the possibility of being confused or making a mistake is often present, an important avenue for future work may relate to investigating techniques aimed at reducing motivated errors. Related to work on the role of cognitive dissonance (Babcock et al. 1995; Haisley and Weber 2010; Gneezy et al. 2020; Gneezy, Saccardo, and van Veldhuizen 2018), one approach may be to make motivated errors more salient to individuals by highlighting the lack of errors when errors cannot be used as an excuse for undesirable behaviors.

Second, since social image concerns are rather muted in our studies (i.e., subjects are anonymous and do not interact with the experimenter in person), an interesting open question is how motivated errors operate in environments with heightened observability. On one hand, observability may mitigate some types of motivated errors if, for example, it is harder for individuals to rationalize decisions by acting confused when they are observed directly (e.g., if they have to act as if they are confused rather than just making choices consistent with confusion).[30] In addition, in simple environments, individuals may be concerned about observers pointing out the implausibility of confusion as an excuse. On the other hand, observability may exacerbate motivated errors since individuals may have an increased desire to rationalize their decisions to others and so may make motivated errors rather than engaging in obvious selfishness.[31]

Third, we explore three contexts in which individuals make motivated errors, but there are many other contexts in which individuals could make motivated errors. Akin to the call put forth in Bénabou and Tirole (2016), future work may examine whether individuals make motivated errors in other contexts, such as in the presence of well-established behavioral biases and cognitive limitations like base-rate neglect, confirmation bias, the gambler's fallacy, and default effects—the last of which may be particularly promising to explore given that defaults substantially influence how much information individuals acquire in the classic moral wiggle room paradigm (see Grossman 2014).

REFERENCES

Andreoni, James, and B. Douglas Bernheim. 2009. "Social Image and the Fifty–Fifty Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica* 77 (5): 1607–36.

Andreoni, James, Justin M. Rao, and Hannah Trachtman. 2017. "Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving." *Journal of Political Economy* 125 (3): 625–53.

[30] That said, just as individuals often deceive others to achieve more self-serving outcomes (Gneezy 2005; Shalvi et al. 2011; Shalvi, Eldar, and Bereby-Meyer 2012; Gino and Ariely 2012; Gino, Ayal, and Ariely 2013; Pittarello et al. 2015; Bicchieri, Dimant, and Sonderegger 2023), they may succeed at deceiving others about the drivers of their behavior.

[31] For related work, see Foerster and van der Weele (2021).

Ariely, Dan, George Loewenstein, and Drazen Prelec. 2003. "'Coherent arbitrariness': Stable Demand Curves Without Stable Preferences." *Quarterly Journal of Economics* 118 (1): 73–106.

Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer. 1995. "Biased Judgments of Fairness in Bargaining." *American Economic Review* 85 (5): 1337–43.

Bartling, Björn, and Yagiz Özdemir. 2023. "The Limits to Moral Erosion in Markets: Social Norms and the Replacement Excuse." *Games and Economic Behavior* 138: 143–60.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka. 2016. "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition." *American Economic Review* 106 (6): 1437–75.

Batson, C. Daniel, Diane Kobrynowicz, Jessica L. Dinnerstein, Hannah C. Kampf, and Angela D. Wilson. 1997. "In a Very Different Voice: Unmasking Moral Hypocrisy." *Journal of Personality and Social Psychology* 72 (6): 1335–48.

Bazerman, Max H., George F. Loewenstein, and Sally Blount White. 1992. "Reversals of Preference in Allocation Decisions: Judging an Alternative Versus Choosing Among Alternatives." *Administrative Science Quarterly* 37 (2): 220–40.

Bénabou, Roland, and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *Quarterly Journal of Economics* 117 (3): 871–915.

Bénabou, Roland, and Jean Tirole. 2004. "Willpower and Personal Rules." *Journal of Political Economy* 112 (4): 848–86.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–78.

Bénabou, Roland, and Jean Tirole. 2011. "Identity, Morals, and Taboos: Beliefs as Assets." *Quarterly Journal of Economics* 126 (2): 805–55.

Bénabou, Roland, and Jean Tirole. 2016. "Mindful Economics: The Production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives* 30 (3): 141–64.

Bénabou, Roland, Armin Falk, and Jean Tirole. 2018. "Narratives, Imperatives and Moral Reasoning." Unpublished.

Bicchieri, Cristina, Eugen Dimant, and Silvia Sonderegger. 2023. "It's not a lie if you believe the norm does not apply: Conditional norm-following and belief distortion." *Games and Economic Behavior* 138: 321–54.

Bodner, Ronit, and Drazen Prelec. 2003. "Self-Signaling and Diagnostic Utility in Everyday Decision Making." In *Psychology of Economic Decisions,* Vol. 1, edited by Isabelle Brocas and Juan D. Carrillo, 105–26. Oxford: Oxford University Press.

Bohnet, Iris, Alexandra van Geen, and Max Bazerman. 2016. "When Performance Trumps Gender Bias: Joint Versus Separate Evaluation." *Management Science* 62 (5): 1225–34.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2020. "Memory, Attention, and Choice." *Quarterly Journal of Economics* 135 (3): 1399–1442.

Broberg, Tomas, Tore Ellingsen, and Magnus Johannesson. 2007. "Is Generosity Involuntary?" *Economics Letters* 94 (1): 32–37.

Brocas, Isabelle, Juan D. Carrillo, Stephanie W. Wang, and Colin F. Camerer. 2014. "Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games." *Review of Economic Studies* 81 (3): 944–70.

Buser, Thomas, Leonie Gerhards, and Joël J. van der Weele. 2018. "Measuring Responsiveness to Feedback as a Personal Trait." *Journal of Risk and Uncertainty* 56 (2): 165–92.

Chadd, Ian, Emel Filiz-Ozbay, and Erkut Y. Ozbay. 2021. "The Relevance of Irrelevant Information." *Experimental Economics* 24 (3): 985–1018.

Chance, Zoe, and Michael I. Norton. 2015. "The What and Why of Self-Deception." *Current Opinion in Psychology* 6: 104–07.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99 (4): 1145–77.

Chew, Soo Hong, Wei Huang, and Xiaojian Zhao. 2020. "Motivated False Memory." *Journal of Political Economy* 128 (10): 3913–39.

Cialdini, Robert. 1984. *Influence, the Psychology of Persuasion.* New York, NY: Harper Collins.

Conlisk, John. 1996. "Why Bounded Rationality?" *Journal of Economic Literature* 34 (2): 669–700.

Conrads, Julian, and Bernd Irlenbusch. 2013. "Strategic ignorance in ultimatum bargaining." *Journal of Economic Behavior and Organization* 92: 104–15.

Coutts, Alexander. 2018. "Good News and Bad News are Still News: Experimental Evidence on Belief Updating." *Experimental Economics* 22 (2): 369–95.

Dana, Jason, Daylian M. Cain, and Robyn M. Dawes. 2006. "What You Don't Know Won't Hurt Me: Costly (but Quiet) Exit in Dictator Games." *Organizational Behavior and Human Decision Processes* 100 (2): 193–201.

Dana, Jason, Roberto A. Weber, and Jason Xi Kuang. 2007. "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory* 33 (1): 67–80.

Danilov, Anastasia, and Silvia Saccardo. 2016. "Disguised Discrimination." Unpublished.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47 (2): 315–72.

DellaVigna, Stefano, John A. List, and Ulrike Malmendier. 2012. "Testing for Altruism and Social Pressure in Charitable Giving." *Quarterly Journal of Economics* 127 (1): 1–56.

Di Tella, Rafael, Ricardo Perez-Truglia, Andres Babino, and Mariano Sigman. 2015. "Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism." *American Economic Review* 105 (11): 3416–42.

Eil, David, and Justin M. Rao. 2011. "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself." *American Economic Journal: Microeconomics* 3 (2): 114–38.

Enke, Benjamin. 2020. "What You See Is All There Is." *Quarterly Journal of Economics* 135 (3): 1363–98.

Enke, Benjamin, Uri Gneezy, Brian Hall, David C. Martin, Vadim Nelidov, Theo Offerman, and Jeroen van de Ven. 2021. "Cognitive Biases: Mistakes or Missing Stakes?" *Review of Economics and Statistics* 105 (4): 818–32.

Enke, Benjamin, and Florian Zimmermann. 2019. "Correlation Neglect in Belief Formation." *Review of Economic Studies* 88 (1): 313–32.

Ertac, Seda. 2011. "Does Self-Relevance Affect Information Processing? Experimental Evidence on the Response to Performance and Non-Performance Feedback." *Journal of Economic Behavior and Organization* 80 (3): 532–45.

Exley, Christine L. 2016. "Excusing Selfishness in Charitable Giving: The Role of Risk." *Review of Economic Studies* 83 (2): 587–628.

Exley, Christine L. 2020. "Using Charity Performance Metrics as an Excuse Not to Give." *Management Science* 66 (2): 553–63.

Exley, Christine L., and Judd B. Kessler. 2023. "Information Avoidance and Image Concerns." *Economic Journal* 133 (656): 3153–68.

Exley, Christine L., and Judd B. Kessler. 2024. "Replication data for: Motivated Errors." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E195862V1.

Falk, Armin, and Nora Szech. 2013. "Morals and Markets." *Science* 340 (6133): 707–11.

Falk, Armin, and Nora Szech. 2020. "Diffusion of Being Pivotal and Immoral Outcomes." *Review of Economic Studies* 87 (5): 2205–29.

Falk, Armin, and Florian Zimmermann. 2017. "Consistency as a Signal of Skills." *Management Science* 63 (7): 2197–2210.

Falk, Armin, and Florian Zimmermann. 2018. "Information Processing and Commitment." *Economic Journal* 128 (613): 1983–2002.

Feiler, Lauren. 2014. "Testing models of information avoidance with binary choice dictator games." *Journal of Economic Psychology* 45: 253–67.

Finkelstein, Amy. 2009. "E-ztax: Tax Salience and Tax Rates." *Quarterly Journal of Economics* 124 (3): 969–1010.

Foerster, Manuel, and Joël J. van der Weele. 2021. "Casting Doubt: Image Concerns and the Communication of Social Impact." *Economic Journal* 131 (639): 2887–2919.

Freddi, Eleonora. 2021. "Do People Avoid Morally Relevant Information? Evidence from the Refugee Crisis." *Review of Economics and Statistics* 103 (4): 605–20.

Gabaix, Xavier. 2019. "Behavioral Inattention." In *Handbook of Behavioral Economics: Applications and Foundations*, Vol. 2, edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, 261–343. Amsterdam: Elsevier.

Garcia, Thomas, Sébastien Massoni, and Marie Claire Villeval. 2020. "Ambiguity and Excuse-Driven Behavior in Charitable Giving." *European Economic Review* 124: 103412.

Gauriot, Romain, Stephanie A. Heger, and Robert Slonim. 2020. "Altruism or Diminishing Marginal Utility?" *Journal of Economic Behavior and Organization* 180: 24–48.

Gilovich, Thomas, Dale W. Griffin, and Daniel Kahneman. 2002. "The Affect Heuris." In *Heuristics and biases: The Psychology of Intuitive Judgment*, edited by Thomas Gilovich, Dale W. Griffin, and Daniel Kahneman, 397–420. New York, NY: Cambridge University Press.

Gino, Francesca, and Dan Ariely. 2012. "The Dark Side of Creativity: Original Thinkers can be More Dishonest." *Journal of Personality and Social Psychology* 102 (3): 445–59.

Gino, Francesca, Michael I. Norton, and Roberto A. Weber. 2016. "Motivated Bayesians: Feeling Moral While Acting Egoistically." *Journal of Economic Perspectives* 30 (3): 189–212.

Gino, Francesca, Shahar Ayal, and Dan Ariely. 2013. "Self-Serving Altruism? The Lure of Unethical Actions that Benefit Others." *Journal of Economic Behavior and Organization* 93: 285–92.

Gneezy, Ayelet, Alex Imas, Amber Brown, Leif D. Nelson, and Michael I. Norton. 2012. "Paying to be Nice: Consistency and Costly Prosocial Behavior." *Management Science* 58 (1): 179–87.

Gneezy, Uri. 2005. "Deception: The Role of Consequences." *American Economic Review* 95 (1): 384–94.

Gneezy, Uri, Elizabeth A. Keenan, and Ayelet Gneezy. 2014. "Avoiding Overhead Aversion in Charity." *Science* 346 (6209): 632–35.

Gneezy, Uri, Silvia Saccardo, Marta Serra-Garcia, and Roel van Veldhuizen. 2020. "Bribing the Self." *Games and Economic Behavior* 120: 311–24.

Gneezy, Uri, Silvia Saccardo, and Roel van Veldhuizen. 2018. "Bribery: Behavioral Drivers of Distorted Decisions." *Journal of the European Economic Association* 17 (3): 917–46.

Golman, Russell, David Hagmann, and George Loewenstein. 2017. "Information Avoidance." *Journal of Economic Literature* 55 (1): 96–135.

Golman, Russell, George Loewenstein, Karl Ove Moene, and Luca Zarri. 2016. "The Preference for Belief Consonance." *Journal of Economic Perspectives* 30 (3): 165–88.

Gotthard-Real, Alexander. 2017. "Desirability and Information Processing: An Experimental Study." *Economics Letters* 152: 96–99.

Grossman, Zachary. 2014. "Strategic Ignorance and the Robustness of Social Preferences." *Management Science* 60 (11): 2659–65.

Grossman, Zachary. 2015. "Self-Signaling and Social-Signaling in Giving." *Journal of Economic Behavior and Organization* 117: 26–39.

Grossman, Zachary, and David Owens. 2012. "An Unlucky Feeling: Overconfidence and Noisy Feedback." *Journal of Economic Behavior and Organization* 84 (2): 510–24.

Grossman, Zachary, and Joël J. van der Weele. 2017. "Self-Image and Willful Ignorance in Social Decisions." *Journal of the European Economic Association* 15 (1): 173–217.

Haisley, Emily C., and Roberto A. Weber. 2010. "Self-Serving Interpretations of Ambiguity in Other-Regarding Behavior." *Games and Economic Behavior* 68 (2): 614–25.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein. 2014. "Learning Through Noticing: Theory and Evidence from a Field Experiment." *Quarterly Journal of Economics* 129 (3): 1311–53.

Heger, Stephanie A., and Nicholas W. Papageorge. 2018. "We Should Totally Open a Restaurant: How Optimism and Overconfidence Affect Beliefs." *Journal of Economic Psychology* 67: 177–90.

Herr, Paul M., Steven J. Sherman, and Russell H. Fazio. 1983. "On the Consequences of Priming: Assimilation and Contrast Effects." *Journal of Experimental Social Psychology* 19 (4): 323–40.

Hsee, Christopher K. 1996. "Elastic Justification: How Unjustifiable Factors Influence Judgments." *Organizational Behavior and Human Decision Processes* 66 (1): 122–29.

Ioannidis, Konstantinos, Theo Offerman, and Randolph Sloof. 2020. "On the Effect of Anchoring on Valuations when the Anchor is Transparently Uninformative." *Journal of the Economic Science Association* 6 (1): 77–94.

Jacobsen, Karin J., Kari H. Eika, Leif Helland, Jo Thori Lind, and Karine Nyborg. 2011. "Are nurses more altruistic than real estate brokers?" *Journal of Economic Psychology* 32 (5): 818–31.

Kamdar, Amee, Steven D. Levitt, John A. List, Brian Mullaney, and Chad Syverson. 2015. "Once and Done: Leveraging Behavioral Economics to Increase Charitable Contributions." Unpublished.

Konow, James. 2000. "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American Economic Review* 90 (4): 1072–92.

Kőszegi, Botond. 2006. "Ego Utility, Overconfidence, and Task Choice." *Journal of the European Economic Association* 4 (4): 673–707.

Kuhnen, Camelia M. 2015. "Asymmetric Learning from Financial Information." *Journal of Finance* 70 (5): 2029–62.

Kunda, Ziva. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108 (3): 480–98.

Larson, Tara, and C. Monica Capra. 2009. "Exploiting moral wiggle room: Illusory preference for fairness? A comment." *Judgment and Decision Making* 4 (6): 467–74.

Lazear, Edward P., Ulrike Malmendier, and Roberto A. Weber. 2012. "Sorting in Experiments with Application to Social Preferences." *American Economic Journal: Applied Economics* 4 (1): 136–63.

Lin, Stephanie C., Rebecca L. Schaumberg, and Taly Reich. 2016. "Sidestepping the rock and the hard place: The private avoidance of prosocial requests." *Journal of Experimental Social Psychology* 64: 35–40.

Linardi, Sera, and Margaret A. McConnell. 2011. "No Excuses for Good Behavior: Volunteering and the Social Environment." *Journal of Public Economics* 95 (5-6): 445–54.

List, John A. 2003. "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics* 118 (1): 41–71.

Liu, Peggy J., and Stephanie C. Lin. 2018. "Projecting Lower Competence to Maintain Moral Warmth in the Avoidance of Prosocial Requests." *Journal of Consumer Psychology* 28 (1): 23–39.

Madrian, Brigitte C. 2014. "Applying Insights from Behavioral Economics to Policy Design." *Annual Review of Economics* 6 (1): 663–88.

Matthey, Astrid, and Tobias Regner. 2011. "Do I Really Want to Know? A Cognitive Dissonance-Based Explanation of Other-Regarding Behavior." *Games* 2 (1): 114–35.

Mijović-Prelec, Danica, and Drazen Prelec. 2010. "Self-Deception as Self-Signalling: A Model and Experimental Evidence." *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1538): 227–40.

Mobius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. 2022. "Managing Self-Confidence: Theory and Experimental Evidence." *Management Science* 68 (11): 7793–7817.

Nyborg, Karine. 2011. "I don't want to hear about it: Rational ignorance among duty-oriented consumers." *Journal of Economic Behavior and Organization* 79 (3): 263–74.

Oberholzer-Gee, Felix, and Reiner Eichenberger. 2008. "Fairness in Extended Dictator Game Experiments." *B.E. Journal of Economic Analysis and Policy* 8 (1): 1–19.

Olschewski, Sebastian, Marius Dietsch, and Elliot A. Ludvig. 2019. "Anti-Social Motives Explain Increased Risk Aversion for Others in Decisions from Experience." *Judgment and Decision Making* 14 (1): 58–71.

Palma, Marco A., and Zhicheng Phil Xu. 2019. "Shadow of a Doubt: Moral Excuse in Charitable Giving." *Review of Behavioral Economics* 6 (2): 133–46.

Pittarello, Andrea, Margarita Leib, Tom Gordon-Hecker, and Shaul Shalvi. 2015. "Justifications Shape Ethical Blind Spots." *Psychological Science* 26 (6): 794–804.

Rabin, Matthew. 1995. "Moral Preferences, Moral Constraints, and Self-Serving Biases." Unpublished.

Rabin, Matthew. 1998. "Psychology and Economics." *Journal of Economic Literature* 36 (1): 11–46.

Regner, Tobias. 2018. "Reciprocity Under Moral Wiggle Room: Is it a Preference or a Constraint?" *Experimental Economics* 21 (4): 779–92.

Saucet, Charlotte, and Marie Claire Villeval. 2019. "Motivated Memory in Dictator Games." *Games and Economic Behavior* 117: 250–75.

Schwardmann, Peter, Egon Tripodi, and Joël J. van der Weele. 2022. "Self-Persuasion: Evidence from Field Experiments at Two International Debating Competitions." *American Economic Review* 112 (4): 1118–46.

Schwardmann, Peter, and Joël van der Weele. 2019. "Deception and Self-Deception." *Nature Human Behaviour* 10 (3): 1055–61.

Schwartzstein, Joshua. 2014. "Selective Attention and Learning." *Journal of the European Economic Association* 12 (6): 1423–52.

Serra-Garcia, Marta, and Nora Szech. 2022. "The (In)Elasticity of Moral Ignorance." *Management Science* 68 (7): 4815–34.

Shalvi, Shaul, Jason Dana, Michel J. J. Handgraaf, and Carsten K. W. De Dreu. 2011. "Justified Ethicality: Observing Desired Counterfactuals Modifies Ethical Perceptions and Behavior." *Organizational Behavior and Human Decision Processes* 115 (2): 181–90.

Shalvi, Shaul, Ori Eldar, and Yoella Bereby-Meyer. 2012. "Honesty Requires Time (and Lack of Justifications)." *Psychological Science* 23 (10): 1264–70.

Shaw, Alex, Natalia Montinari, Marco Piovesan, Kristina R. Olson, Francesca Gino, and Michael I. Norton. 2014. "Children Develop a Veil of Fairness." *Journal of Experimental Psychology: General* 143 (1): 363–75.

Snyder, Melvin L., Robert E. Kleck, Angelo Strenta, and Steven J. Mentzer. 1979. "Avoidance of the Handicapped: An Attributional Ambiguity Analysis." *Journal of Personality and Social Psychology* 37 (12): 2297–2306.

Spiekermann, Kai, and Arne Weiss. 2016. "Objective and subjective compliance: A norm-based explanation of 'moral wiggle room'." *Games and Economic Behavior* 96: 170–83.

Taubinsky, Dmitry, and Alex Rees-Jones. 2018. "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment." *Review of Economic Studies* 85 (4): 2462–96.

Thaler, Michael. 2019. "The "Fake News" Effect: An Experiment on Motivated Reasoning and Trust in News." Unpublished.

Trachtman, Hannah, Andrew Steinkruger, Mackenzie Wood, Adam Wooster, James Andreoni, James J. Murphy, and Justin M. Rao. 2015. "Fair weather avoidance: unpacking the costs and benefits of 'Avoiding the Ask'." *Journal of the Economic Science Association* 1: 8–14.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124–31.

van der Weele, Joël J., Julija Kulisa, Michael Kosfeld, and Guido Friebel. 2014. "Resisting Moral Wiggle Room: How Robust Is Reciprocal Behavior?" *American Economic Journal: Microeconomics* 6 (3): 256–64.

Wang, Stephanie W., Michelle Filiba, and Colin F. Camerer. 2010. "Dynamically Optimized Sequential Experimentation (DOSE) for Estimating Economic Preference Parameters." Unpublished.

Zimmermann, Florian. 2020. "The Dynamics of Motivated Beliefs." *American Economic Review* 110 (2): 337–61.

This article has been cited by:

1. Lasse S. Stoetzer, Florian Zimmermann. 2024. A note on motivated cognition and discriminatory beliefs. *Games and Economic Behavior* **147**, 554-562. [Crossref]
2. Wei Huang, Yu Wang, Xiaojian Zhao. 2024. Motivated Beliefs, Independence and Cooperation. *European Economic Review* **166**, 104748. [Crossref]
3. Gergely Hajdu. 2024. Excusing Beliefs about Third-party Success. *SSRN Electronic Journal* **98**. . [Crossref]
4. Gergely Hajdu. Excusing Beliefs About Third-Party Success **98**, . [Crossref]
5. Vivek Roy-Chowdhury, Rebecca Heath. Recall Distortion and Past Choices **108**, . [Crossref]
6. Sven A. Simon, Pascal Nieder. 2024. Dishonesty in Complex Environments: Deliberate Lies, Short-Cuts, or Accidental Mistakes? . *SSRN Electronic Journal* **87**. . [Crossref]