UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

MEASURING INDIVIDUAL DIFFERENCES AND CHANGES IN MORAL REASONING: AN ECONOMIC GAME APPROACH

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

DOUGLAS J. BRYANT
Norman, Oklahoma
2018

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/301821

MEASURING INDIVIDUAL DIFFERENCES AND CHANGES IN MORAL REASONING:

AN ECONOMIC GAME APPROACH

A THESIS APPROVED FOR THE
DEPARTMENT OF PSYCHOLOGY

BY

Dr. Luz-Eugenia Fuenzalida, Chair

Dr. Andrea Vincent

Dr. Mike Buckley

Dr. Erin Freeman

© Copyright by DOUGLAS J. BRYANT 2018
All Rights Reserved.

TABLE OF CONTENTS

ABSTRACT v

INTRODUCTION 1

METHODS 9

RESULTS 12

DISCUSSION 15

REFERENCES 22

TABLES 30

Abstract

Motivated reasoning bias and its manifestation in confirmation bias make it unlikely that people will change their minds on matters for which their positions are expressions of their own or their group's identity. Moral scenarios serve as a useful proxy for evaluating motivated reasoning because moral positions often comprise major components of individual and group identity. While many interventions to overcome these cognitive deficits have been tried, they have had only limited success and suffer from the problem of failure in domain transfer. Previous research has identified a subset of individuals who demonstrate a greater than average immunity to cognitive biases. However, characteristics associated with this reduction in susceptibility have yet to be identified. The goal of the present study was two-fold: 1) to validate an economic game approach for measuring changes in moral reasoning and 2) to examine cognitive behavioral covariates, namely, openminded thinking and moral foundations, associated with performance. This approach used an economic game in which participants spent money to reward or punish scenario actors for their moral actions. The scenarios were divided into six epochs of increasing/new information. Following each epoch, participants rendered a judgment based on the information available to them at the time, and changes from previous judgments were measured. Results revealed that thinking in monetary terms significantly changed moral judgments. Open-mindedness was highly correlated with moral foundations and political identity. Each of these predicted reversals in judgment, though in surprising directions. The results suggest that participants who are politically conservative and less openminded reverse their positions more often than more open-minded, politically liberal participants.

Keywords: Motivated reasoning bias, confirmation bias, cognitive biases, economic game

Measuring Individual Differences and Changes in Moral Reasoning: An Economic Game Approach

Human beings are biased thinkers. Our biases have been diligently and precisely uncovered and catalogued (Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1973; Stanovich, Toplak, & West, 2008; Tversky & Kahneman, 1973, 1974). While adaptive in many contexts, these biases can be disastrous in certain situations. Corrections for these biases have proved difficult to discover. Many interventions have been developed and tested experimentally to quantify their effectiveness, though the ameliorative effects are often small, unlasting, and do not transfer well to other domains. Some studies of performance in everyday reasoning have uncovered effects of training and expertise in reasoning (Kuhn, 1991). A key difference here is whether the reasoning under examination is motivated or unmotivated reasoning. In motivated reasoning, an individual is motivated to reach or defend a specific conclusion. When reasoning is not motivated, there is no desire to reach one conclusion over another. Training in reasoning is associated with expert performance in unmotivated but not motivated reasoning tasks (Kahan, 2017; Kuhn, 1991). Motivated reasoning activates certain biases, resistance to which is not much improved by training, but is predicted by certain cognitive tests and attributes discussed below. Here we examine a novel measure of cognitive performance and resistance to bias; namely, the ability to change one's mind. We measured this ability in multiple ways and assessed its association with traits proposed to predict cognitive performance in order to assess whether these traits offer any advantage in combating motivated reasoning biases. Finally, we compared performance on the experimental tasks in two conditions: an economic game condition in which judgments were made monetarily, and a control condition with non-monetary judgments.

Biases and Heuristics

Among our most powerful and pernicious biases are those belonging to the class of motivated reasoning biases (Kunda, 1990). Perhaps the most well-known of these, confirmation bias (Nickerson, 1998) (sometimes termed *myside bias*), is the tendency to seek out evidence confirming one's beliefs, avoid evidence contradiction one's beliefs, and to give more credence to confirming than disconfirming evidence. Confirmation bias is known to be exacerbated, rather than ameliorated, as individual facility with numerical data increases (Kahan, Peters, Dawson, & Slovic, 2017). This is consistent with other findings showing that rationality and intelligence are not the same thing and may not only stand apart, but also be antithetical (Mercier & Sperber, 2011, Stanovich, 2012). Performance on heuristics-and-biases tasks that assess bias like confirmation bias is poorly correlated with measures of general intelligence like IQ (Stanovich & West, 2008a). Cognitive ability does not, for example, predict susceptibility to confirmation bias, one-sided thinking bias, (Stanovich & West, 2008b) or bias blind spot (West, Meserve, & Stanovich, 2012). Performance on these tasks is, however, correlated with so-called thinking dispositions such as need for cognition and actively open-minded thinking (Kokis, Macpherson, Toplak, West, & Stanovich, 2002; Sá, West, & Stanovich, 1999; West, Toplak, & Stanovich, 2008). More predictive still is performance on cognitive tests like the Cognitive Reflection Test (CRT, Frederick, 2005) that require the suppression of automatic, fast, System One responses in order to facilitate deliberative, slower, System Two responses (Kahneman, 2011; Stanovich, Toplak, & West, 2008).

Syllogistic tests demonstrate this in two ways. First are valid but obviously unsound arguments; that is, arguments in which the conclusion logically follows from the premises but in which one or more of the premises is known to be false. Prior beliefs here bias one toward

rejecting the argument as unsound, even if one is explicitly instructed to accept the truth of the premises for the purpose of the test. Such an argument might take the form:

P1: All mammals have hair

P2: Some birds are mammals

C: ∴ Some birds have hair

This argument is unsound—P2 is untrue—but logically valid, and recognizing its validity requires the ability to reason logically when logic conflicts with one's prior beliefs (Toplak *et al*., 2011).

Second are invalid arguments that use known-to-be-true or believed-be-true premises and conclusions. Again, as a standard form categorical syllogism, we formulate the argument:

P1: All dogs have hair

P2: All German Shepherds have hair

C: ∴ All German Shepherds are dogs

The logic of this argument is invalid; the conclusion does not follow from the premises. Seeing this again requires the ability to reason logically—here recognizing a logical fallacy—when the fallacious conclusion and the premises are all consistent with one's prior beliefs. Standard form categorical syllogisms can take 256 forms only 15 of which are valid (Copi, Cohen, & McMahon, 2016). This argument takes the form AAA-2. The two premises and conclusion are universal (All dogs and all German Shepherds). The similarly formed *Barbara* syllogism has the valid form AAA-1. The difference is in the positioning of the middle term "hair." The invalid German Shepherd argument takes the skeleton:

All *P* have *M*.

All *S* have *M*.

∴ All *S* are *P*

where *P* = predicate term, *S* = subject term, and *M* = middle term. The valid *Barbara* form takes the skeleton:

All *M* is *P*.

All *S* is *M*.

∴ All *S* is *P*.

The only formal difference is the placement of the middle term. In the *Barbara* form, the German Shepherd argument is:

P1: All animals with hair are German Shepherds.

P2: All dogs are animals with hair.

C: ∴ All dogs are German Shepherds.

Here again is a logically valid but unsound argument: it is formally correct—the conclusion necessarily follows from the premises—but its premises are untrue. Thus, misleading arguments are easily made by placing false premises into valid forms and by placing true premises into invalid forms.

In addition to esoteric and technical biases and heuristics like ignoring base rates (Tversky & Kahneman, 1981) and failures of numeracy in data interpretation (Kahan, Peters, Wittlin, Slovic, Ouellette, Braman, & Mandel, 2012; Peters, Västfjäll, Slovic, Mertz, Mazzocco, & Dickert, 2006), there are more common failures of everyday reasoning that demonstrate the real-world implications of the cognitive failures demonstrated by heuristics-and-biases tests (Hilton, 2003; Reyna & Farley, 2006; Reyna & Lloyd, 2006; Sunstein, 2005). Among others, these include failure to conceive of evidence for one's view (as opposed to pseudoevidence), of evidence against one's view, of an alternative theory, of a counterargument, failure to rebut an

alternative theory or counterargument, and failure to recognize and reject pseudoevidence (Kuhn, 1991).

Rationality

Humans are biased, but the question remains whether human beings are irrationally or rationally biased thinkers. Despite the decline of the rationalist model (Clark, 2008; Haidt, 2001) popular among mid-twentieth century cognitive psychologists, and the now *in vogue* social intuitionist theory popularized by Haidt (2001), the heuristics and biases literature remains surprisingly controversial (Gigerenzer & Todd, 1999; Klein, 1999). The social intuitionist theory is a theory of specifically moral judgment which holds that moral judgments are the result of intuitions which conform to group identity—an historically adaptive trait for moral animals whose survival is determined both by the wellbeing of the group and one's place within it. Evidence and reasons supporting moral judgments are *post hoc*; however, thinkers are typically oblivious to this and believe the process to have occurred in the reverse direction (Haidt, 2012a; Haidt & Björklund, 2008). The opposing camps on the question of rational or irrational bias as parsed by Stanovich (1999) are the Panglossian, Meliorist, and Apologist positions. What distinguishes them is the way in which they construe the relationship between normative, prescriptive, and descriptive models of rationality. On one extreme is the Panglossian position, which sees no distance between the three models of rationality; in other words, humans are rational. There is thus no distinction to be made between a descriptive model of human rationality and a normative model, and hence no need of a prescriptive model to bridge any such gap. By contrast, the Meliorist and Apologist positions see a large gap between normative and descriptive models and hence find that humans are irrational. What distinguishes these views is where they place the prescriptive model along a continuum. For instance, the Meliorist position

is pessimistic about how much can be done to optimize human rationality, so its prescriptive model is much closer to the normative model than the descriptive model, leaving little room for improvement. By contrast, the Apologist position is optimistic about optimizing human rationality, so its prescriptive model starts at the descriptive model, leaving much room for improvement.

All three approaches hold humans to be biased. The Panglossian approach takes humans to be rationally biased; Meliorists and Apologists hold humans to be irrationally biased. Panglossians see cognitive biases as adaptive traits—evolutionary artifacts that are sometimes anachronistic, but which more often remain adaptive. The Panglossian position need not hold that the biases and heuristics humans rely upon are universally adaptive. More parsimoniously, one can hold that while these reasoning strategies were once adaptive and often still are, they can be misleading and maladaptive today. This is especially so in societies where survival is not largely dependent on group membership and where societies and groups themselves are so large as to hold conflicting positions on many issues.

The biases and heuristics literature has seldom intersected with the social intuitionist model and other decision-making research in social psychology (see Kuhn, 1991 and Kunda, 1990 for exceptions). However, if Haidt's social intuitionist model is correct, the implications for the previously established positions are potentially significant. The heuristics and biases literature is almost exclusively concerned with thinking biases; that is, failures to demonstrate optimal reasoning on what are more or less logic puzzles. The strongest objection to this method of testing, and the argument that has proved so fruitful for the Panglossians, is that performance on logic puzzles is neither adaptive nor more than rarely useful, and hence a poor, if not irrelevant, measure of rationality (Stanovich, 1999). More relevant are tests of everyday

reasoning (Kuhn, 1991) in matters where one is motivated to reach a certain conclusion (Haidt, 2001; Kunda, 1990).

Debiasing Strategies

Many debiasing strategies and interventions have been studied, but these have generally shown meager and temporary results (Aczel, Bago, Szollosi, Foldes, & Lukacs, 2015). Awareness of biases and heuristics is insufficient to surmount them (Babcock & Loewenstein, 1997; Fishhoff, 1981), and training in statistics (Fong, Krantz, & Nisbett, 1986; Nisbett, Krantz, Jepson, & Kunda, 1983) and critical thinking (Baron, 2000; Ennis, 1991; Niu, Behar-Horenstein, & Garvan, 2013; Willingham, 2008;) suffer from the problem of domain transfer (Cheng, Holyoak, Nisbett, & Oliver 1986; Fong & Nisbett, 1991). People can learn the difference between *modus ponens* and *modus tollens* or the law of large numbers, for example, but cannot apply that knowledge outside of the context in which it was learned. Some have made headway against the domain transfer problem with analogical reasoning (Aczel *et al*., 2015; Vendetti, Wu, & Holyoak, 2014); however, improvements are limited and fail to persist.

The human factors approach to debiasing seeks to overcome cognitive failures by changing the environment rather than the user. It is applied in behavioral economics and judgment and decision-making research (Thaler & Sunstein, 2008). The use of visual aids in communicating medical information is one such example (Garcia-Retamero & Cokely, 2011, 2013). This is a useful approach; however, it requires human factors psychologists and decision scientists to remake the world. That is a tremendous undertaking, and yet not all the world is accessible to these decision scientists. In the realm of national security intelligence analysis, for example, decision scientists could surely make improvements. However, it is not possible for all

classified intelligence products to be vetted by psychologists; nor is it possible for all intelligence professionals to become human factors experts.

The human factors approach makes no attempt to uncover the biological causes of cognitive failures and cannot offer a way to correct them at the source. Biological explanations for cognitive failures and differences in reasoning, on the other hand, have flourished (Bryant, Deardeuff, Zoccoli, & Nam, 2016; De Sousa, 2008; Osherson *et al*., 1998; Westen, Blagov, Harenski, Kilts, & Hamann, 2006). We contend that these efforts to reveal the ultimate causes of cognitive failures offer the best hope of correcting those failures. Here we sought to identify and describe individuals who already demonstrate expertise in reasoning as a precursor to biological work.

Individual Differences in Rationality

A number of studies have shown that there are significant individual differences in rationality (Stanovich, 1999); some people have demonstrated perfect reasoning on impersonal and presumably unmotivating subjects (Kuhn, 1991). The expert reasoners identified by Kuhn were graduate philosophy students. They were questioned on three topics: recidivism, school failure, and unemployment. They were not asked such motivating questions as, for example, "Why did the *Journal of Philosophy* reject your last article?" The goal of the present study was to identify traits of individuals demonstrating one aspect of optimal reasoning, the ability to change one's mind, on questions designed to evoke motivated responses. To do this, we designed an economic game that drew on game theoretical approaches (Nowak & Sigmund, 1993) to examine whether and what individual differences correlate with and predict resistance to bias in motivated reasoning. Moral scenarios and dilemmas served as a proxy for questions involving motivated reasoning since most moral positions have a place in one's own and one's group's

identity (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). The following main effects were hypothesized: Susceptibility to motivated reasoning biases as marked by inability to change one's mind would correlate with less openminded thinking, five-foundation morality and political conservatism. Conversely, resistance to bias in motivated reasoning would correlate with open-mindedness, two-foundation morality, and political liberalism.

Methods

Participants

Participants were 112 Amazon Mechanical Turk (mTurk) basic users who responded to a listing soliciting participants for a moral thinking experiment. Participants were predominantly males (63%) and ranged in age from 18 to 69 (M = 34.32, SD = 10.537) (Table 1). All participants consented to participate in the research at the beginning of the experiment using an online unsigned consent form approved by the institutional review board and were paid $5 for their participation.

Procedure

All participants answered a four-item questionnaire to indicate moral positions on one to four topics: the role of free will in moral culpability, use of animal products, insurance coverage for transitioning surgeries, and the amputation of a healthy limb for a person with body integrity identity disorder. Participants then completed the moral reasoning task followed by three self-report questionnaires.

Assessments.

***Moral reasoning task*.** In this task, participants were presented with one of four scenarios. Each scenario was designed to change the participants' minds on an issue by presenting evidence and arguments supporting an alternate position from their own. Scenarios

were designed to extend the logic of one of the initial four positions to what would be an uncomfortable but logically consistent conclusion. Participants who indicated they strongly agreed that free will was required for moral culpability entered a scenario in which arguments against free will were presented as leading to the conclusion that culpability is therefore incoherent. Participants who indicated they were in favor of equality but for or indifferent to the use of animal products proceeded to a scenario in which arguments for extending the principle of equality beyond arbitrary distinctions (sex, race, etc.) were extended to non-human sentients. Participants who indicated they were opposed to discrimination against transgender persons proceeded to a scenario in which a transgender woman tells her fiancée that she was born with male genitalia. Participants who indicated they supported health insurance coverage of transgender sex reassignment surgery proceeded to a scenario in which an individual with body integrity identity disorder (BIID) undergoes an amputation of an otherwise healthy limb in order to bring physical identity into congruency with felt sense of self.

In order to induce motivated moral cognitions related to group identity, we developed the former two scenarios for conservative participants and the latter two scenarios for liberal participants. Since scenario assignment was made on the basis of responses to the initial moral positions questionnaire, the number of participants in each scenario was unequal.

Each scenario provided background information followed by a description of the character's action. Participants were asked to render a moral judgment of the character in the scenario. This initial judgement was then followed by five additional rounds of additional arguments with a new judgment requested subsequent to each. This resulted in a total of six judgements from each participant.

One aim of this study was to evaluate an economic game as a tool for measuring changes of mind in moral judgments. Thus, study participants were divided into two groups: control and treatment. Participants were recruited through two different study announcements on Amazon Turk and assigned to treatment or control groups based on the study announcement to which they responded. In the treatment condition (n = 63) participants played an economic game in which they chose to reward or punish the scenario character for their actions by spending notional money in $10 increments. Participants also had the option to withhold judgment by refraining from spending any money. Over the six rounds, each participant had the opportunity to maximally punish or reward a character by spending $60. After the sixth judgment, participants were again issued $60 of notional money and indicated their final judgment by spending any amount ranging from $0 to $60 to abstain from judgment or punish/reward, respectively.

In the control condition (n = 49), participants responded to the same scenarios but did not play the economic game. Instead, they indicated their moral judgment on a 3-point Likert scale in which -1 indicated punishment, 0 indicated abstain, and +1 indicated reward.

***Actively Openminded Thinking Scale*.** The AOT (Stanovich & West, 1997) is a multi-item questionnaire including several subscales designed to assess openness and flexibility in thinking style. Scores are calculated as difference scores in which the summed totals of the absolutism, dogmatism, and categorical thinking scales are subtracted from the summed totals of the flexible thinking, openness-ideas, and openness-values scales such that higher scores indicate greater openmindedness.

Moral Foundations Questionnaire. The MFQ30 (Haidt, 2012b) consists of 30 items assessing five moral foundations: care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, and sanctity/degradation. Work by Haidt (2012b) has correlated two-

foundation morality with political liberalism and five-foundation morality with political conservatism. In two-foundation morality, care/harm and fairness/cheating are more valued than the other three foundations. In five-foundation morality, all five foundations are valued approximately equally. To determine if participants' values were two- or five-foundation, we calculated difference scores by subtracting the loyalty/betrayal, authority/subversion, and sanctity/subversion summed scores from the summed totals for care/harm and fairness/cheating such that more positive scores indicated two-foundation morality, and more negative scores indicated five-foundation morality.

***Demographics*.** The demographics questionnaire was designed to provide additional information about study participants for post hoc analyses and included items addressing political affiliation, political ideology, and education level, among others.

Results

Using an Economic Game as a Measure of Moral Judgment

Self-report data. Data indicated that the control and treatment groups did not differ in age, sex, or education level. However, the two groups did significantly differ in the 3 trait measures assessing open-mindedness, moral foundations, and political identity (Table 1). Specifically, participants in the control group were more politically conservative, less openminded, and more likely to have a five-foundation morality.

Moral Reasoning task performance. Task performance was operationalized using two metrics: number of changes and number of reversals. The number of *changes* represents the number of times the participant switched from one judgment (reward, abstain, or punish) to another over the course of the scenario. The number of *reversals* represents the number of times the judgment of reward or punish changes to its opposite (ignoring abstentions). Of these two

outcomes, the groups differed only on the mean number of reversals (See Table 2) such that the control group had more reversals on average than the economic game group. This finding did not persist after controlling for the differences in the trait measures between the groups, $F(4, 80) = 3.00$, $p = .09$.

Traits associated with moral reasoning performance. Consistent with previous literature, we examined the correlations between the three trait measures: AOT, MFQ30, and political identity (see Table 3). AOT was correlated with MFQ30, $r(83) = .72$, $p < .001$ and political identity $r(110) = -.51$, $p < .001$ with participants high in AOT also associated with a two-foundation morality and left-of-center political identity. Participants low in AOT were also associated with a five-foundation morality and identified as politically right-of-center. Participants with a two-foundation morality valued care and fairness more than the other three foundations: loyalty, sanctity, and authority. Participants with a five-foundation morality valued all five foundations approximately equally. Political identity was correlated with moral foundations $r(83) = -.67$, $p < .001$ such that participants who were politically left-of-center were associated with a two-foundation morality and participants politically right-of-center were associated with a five-foundation morality.

Multiple regression was used to examine AOT, moral foundations, and political identity as predictors of both the number of changes and reversals (see Table 4). AOT significantly predicted reversals, $\beta = -.28$, $t(110) = -3.10$, $p = .002$ such that participants high in AOT made less reversals on average. Moral foundations also significantly predicted reversals, $\beta = -.22$, $t(83) = -2.08$, $p = .04$ such that participants with a two-foundation morality made less reversals. Participants who scored lower in open-mindedness and who demonstrated a five-foundation morality changed their minds (reversals) more often than more openminded and two-foundation

participants, respectively. Despite the high correlations between our three trait measures of interest, AOT, moral foundations, and political identity, there was not a significant difference in the number of reversals made between politically left-of-center and right-of-center participants; however, this difference was detected in analyses of the whole sample below.

Subset analyses. Since our primary interest was in identifying and understanding people who can change their minds, we recast the analyses by grouping participants as those who made changes and those who did not. We then looked for patterns of variation in trait measures among only those participants who demonstrated the ability to change their minds (See Table 5). In this subset, AOT was the sole measure that significantly predicted reversals, $\beta = -.02$, $t(57) = -3.00$, $p = .004$. Again, participants who scored lower in AOT made more reversals than participants who scored higher in AOT.

Since the most common response across all scenarios and both treatment and control groups was to make no changes in judgement throughout the scenarios, we conducted a series of logistic regression models to identify whether AOT, moral foundations, and political identity were predictive of whether participants made *any* changes or *any* reversals (See Table 6). AOT significantly differentiated reversal behavior. The estimated odds of making at least one reversal were decreased among those with high AOT (OR= 0.96, $p = .03$, $R^2 = .043$, adj. $R^2 = .07$). Political identity also predicted reversals such that participants who identified as right-of-center had increased odds of making one or more reversals in judgement in comparison to those identifying as left-of-center (OR = 1.27, $p = 0.047$, $R^2 = .035$, adj. $R^2 = .057$). Moral foundations did not predict reversals in this analysis, and no trait measure predicted changes.

Discussion

Moral Econ Game

Independent samples were taken for control and treatment groups, yet analyses revealed these to be heterogenous groups. In ANOVAs measuring these groups on the independent trait variables, the groups were significantly different on AOT, $F(1, 110) = 14.11$, $p < 0.001$, with the treatment group scoring higher in AOT. The groups were also heterogeneous on independent variables highly correlated with AOT, namely moral foundations, $F(1, 83) = 9.75$, $p = 0.002$, and political identity, $F(1, 110) = 5.585$, $p = 0.020$, with the treatment group tending toward a two-foundation morality and a left-of-center political identity.

The control and treatment groups also differed by number of reversals made. To validate the economic game as a proxy for measuring shifts in moral judgment, we conducted an ANCOVA and compared the dependent variable reversals from each group (treatment and control) while controlling for trait covariates (AOT, moral foundations, and political identity) and found no significant difference between the groups, $F(4, 80) = 3.00$, $p = .09$).

This suggests that thinking in monetary terms did not significantly alter the moral judgments participants made during the experimental scenarios. For this reason, the economic game may be an impartial measure of changes in moral reasonings—at least in the types of personal scenarios used in this experiment. Whether thinking in monetary terms effects other types of moral reasoning, such as impersonal moral dilemmas, requires further investigation. Since changes of mind are rare, and reversals are rarer, examining the ways in which thinking in monetary terms differentially affects distinct trait differences like cognitive style will require future studies with much larger sample sizes.

Individual Differences in Changing One's Mind

We set out to identify individuals who have increased facility with changing their minds in response to evidence and then measure how these individuals differed by likely covariates. These included actively open-minded thinking (AOT), moral foundations, and political identity. Consistent with previous research, these traits were highly correlated for participants in our samples such that participants high in AOT had a two-foundation morality and were politically left-of-center on average. We hypothesized that participants high in AOT, who had a two-foundation morality, and hence valued care and equality more than they valued loyalty, sanctity, and authority, would change their minds more often and reverse their judgments more often. Instead, we found that participants lower in AOT and who identified as politically right-of-center had more reversals on average.

None of the traits we examined predicted whether participants made changes, as distinct from reversals. That is likely because the majority of respondents did not deviate from their original judgment throughout the experiment. This left a small sample of individuals who made any changes, and an even smaller sample that made reversals. This is consistent with the phenomenon under examination. Changing one's mind on a moral issue merely in response to evidence is not common (Haidt, 2012b).

AOT and moral foundations predicted likelihood of reversals in regression analyses. Given that most respondents made no changes in their judgments, we conducted subset analyses to discover if the trait variables of interest could predict who would make reversals among only those participants who made changes. AOT predicted number of reversals in this subgroup. Surprisingly, however, among those who made changes, it was participants lower in AOT who made more reversals.

To further explore these results, we conducted logistic regressions on the subset of participants who made changes to discover if the traits also predicted who made reversals and who did not. AOT and political identity predicted who likely made reversals in this group. Participants low in AOT and politically right-of-center were again more likely to make reversals than their counterparts.

These results are surprising and worth further exploration. One possible explanation for these results is the order in which the parts of the experiment were carried out. Participants began with one to four sorting questions in which they made moral judgments that determined which of the four scenarios into which they proceeded. They next completed a demographics questionnaire that included political identity. They then completed the main experiment—six iterations of scenario evidence and moral judgment. Finally, they completed the Moral Foundations Questionnaire and the Actively Openminded Thinking Questionnaire. Given the high correlations among political identity, moral foundations, and AOT, it seems unlikely that question order is responsible for the results.

Another possibility is that the intuition that traits like open-mindedness will predict changing one's mind or reversing one's mind is somehow incomplete or an instance of two things that do not relate to one another in the way in which we might expect. This might occur were it the case that openminded individuals were immediately convinced by the first of six arguments and hence only had their positions bolstered by the subsequent five pieces of evidence. Similarly, this could occur were it easier for openminded participants to change their minds at a single point, at any of the six opportunities, and then stick with that judgment. At the same time, or independently, such a relationship between these variables might hold if it was harder for less openminded individuals to change their positions, and they hence found

themselves oscillating—responding to the evidence but then reverting to their original judgment. Several analyses argue against these possibilities (See Table 7). We analyzed the patterns by which the six consecutive decisions were made. These fell into four categories: 1. No change was made, 2. The judgment was solidified, 3. The judgment was moderated, 4. No recognizable pattern. Trait variables were not predictive of any of these patterns. Openminded, politically left-of-center participants were no more or less likely to waffle between judgments than less openminded, politically right-of-center participants.

Another possibility is that there is a disconnect between the sort of questions asked in the Moral Foundations and Actively Openminded Thinking questionnaires and the questions asked in the scenarios in this experiment. The Moral Foundations Questionnaire, for example, asks questions like "When you decide whether something is right or wrong, to what extent are the following considerations relevant to your thinking: Whether or not someone suffered emotionally?, Whether or not some people were treated differently than others?, Whether or not someone showed love for his or her country?, Whether or not someone showed a lack of respect for authority?, Whether or not someone violated standards of purity and decency?" (Haidt, 2012b). These are abstract questions; there is no specific scenario presented, though the questions may call such scenarios to mind for the respondent. The Actively Openminded Thinking Scale is similar in this regard. Participants use a Likert scale to indicate agreement or disagreement with statements like the following: "I think there are many wrong ways, but only one right way, to almost anything, Changing your mind is a sign of weakness, Abandoning a previous belief is a sign of strong character, No one can talk me out of something I know is right, Certain beliefs are just too important to abandon no matter how good a case can be made against them" (Stanovich & West, 1997). One such possibility is that the general questions asked in

these questionnaires capture what people think of themselves and their overarching principles, whereas scenario driven questions may capture what people feel or would do in a more specific and salient matter.

Judgments in our experiment, by contrast, were scenario based. Participants made decisions to abstain, punish, or reward specific characters in these scenarios for their actions. In the BIID scenario, for example, participants judged the rightness or wrongness of a surgeon's decision to amputate an otherwise healthy limb of a BIID patient. In the free will scenario, participants decided whether to abstain, reward, or punish a judge for convicting a defendant despite evidence that the defendant did not have the freedom to do otherwise with respect to the crime. Scenarios such as these can be evaluated from different perspectives. While it might be expected that conservatives will side with the judge or that liberals will side with the defendant, it is easy to find conservative reasons for siding with the defendant and liberal reasons for siding with the judge. The same can be said of individual differences in AOT and moral foundations. More, one can be openminded in general, but less so with regard to a specific scenario.

Another possible explanation for these results; namely, that right-of-center, five-foundation, lower AOT participants showed more reversals than their counterparts, is that self-identified conservatives were responding to stereotype threat—the phenomenon in which consciousness of a stereotype about oneself alters one's behavior (Steele & Aronson, 1995). Support for this possibility comes in part from the ordering of questions; participants identified their place on the political spectrum just before each scenario. Something more is needed to make sense of why question order might manipulate responses in the scenarios but not responses to the Moral Foundations and Actively Openminded Thinking questionnaires. This may be due to differential effects of stereotype threat on responses to specific and personal scenario-driven

questions, and the impersonal, general questions asked in those questionnaires. Further research is required both to replicate the scenario-specific finding and to distinguish between these possible interpretations.

We also set out to validate an economic game model as a tool for measuring changes in moral judgment. The challenge here was that thinking in monetary terms can alter certain cognitions, including the weighing of risks, rewards, and likelihoods. To use such a tool to measure changes in moral cognition, it was necessary to measure its use in a treatment group against the judgments of a control group in which participants did not use money to make moral judgments. To assess our groups for differences, we measured the outcome variables changes and reversals while controlling for group differences in trait measures (AOT, moral foundations, and political identity). Here we found no significant differences; participants in the treatment condition made no more or less reversals than participants in the control condition. Thus, monetary measures may be appropriate as a proxy for measuring moral judgments in at least some circumstances.

This finding has limitations. In our scenarios, participants were restricted in the amount of money they could spend; namely, they could spend $10 to reward, $10 to punish, or they could abstain. Results may vary if participants are given the freedom to select the degree to which they would punish and the degree to which they would reward. Further, participants in this study spent notional currency. Spending one's own money may alter this result. A future study might examine this intersection of monetary and moral thinking by issuing participants actual money and allowing them to spend it in the scenario. Alternatively, scenarios might be constructed to illuminate participants to the ways in which their tax dollars are presently spent, or could be alternatively spent, to punish and reward their fellow citizens for various behaviors.

These dollars, while not in hand, are in fact distributed to encourage and discourage various behaviors.

Whereas previous studies, though few, have examined and attempted to predict who is rational, we chose to identify who can change one's mind in response to evidence. This approach is limited in that we did not measure the reasons for which individuals changed their minds, or did not change their minds, and whether those changes were rational. In all four scenarios, it was possible for individuals who did not change their minds to behave rationally, and it was also possible for individuals who did change their minds to behave irrationally. Studies that have assessed the rationality of participants' thinking have done so in one of two ways. The first is to conduct methodical interviews with a small number of participants. The second is to have many participants answer very few simple questions such as questions about base rates. Each of these approaches has its own advantages and disadvantages. With more resources, it is possible to conduct methodical interviews with complex questions with many participants. Such resources might also allow for examining both changes of mind and the rationale behind such changes.

References

Aczel, B., Bago, B., Szollosi, A., Foldes, A., & Lukacs, B. (2015). Is it time for studying real-life debiasing? Evaluation of the effectiveness of an analogical intervention technique. *Frontiers in Psychology*, *6*: 1120.

Babcock, L., & Loewenstein, G. (1997). Explaining bargaining impasse: the role of self-serving biases. *Journal of Economic Perspectives*. *11*(1), 109–126. doi: 10.1257/jep.11.1.109

Baron, J. (2000). *Thinking and deciding (3rd ed)*. Cambridge, UK: Cambridge University Press.

Boudon, R. (2009). Rational choice theory. In B. S. Turner (Ed), *The new Blackwell companion to social theory* (179-195). West Sussex, UK: Blackwell Publishing Ltd.

Bryant, D. J., Deardeuff, K., Zoccoli, E., & Nam, C. S. (2016). The neural correlates of moral thinking: a meta-analysis. *International Journal of Computational & Neural Engineering*, *3*(2), 28-39.

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology* 18, 293–328. doi: 10.1016/0010-0285(86)90002-2

Clarke, S. (2008). SIM and the city: Rationalism in psychology and philosophy and Haidt's account of moral judgment. *Philosophical Psychology*, *21*(6), 799-820.

Copi, I. M., Cohen, C., & McMahon, K. (2011). *Introduction to logic* (14th ed). New York, NY: Taylor and Francis.

Cucu-Ciuhan, G., & Răban-Motounu, N. (2012). The openness to experience questionnaire: Construction and validation. *Procedia-Social and Behavioral Sciences*, *33*, 717-721.

De Sousa, R. (2008). Logic and biology: Emotional inference and emotions in reasoning. In J. Adler & L. J. Rips (Eds.), *Reasoning: Studies of human inference and its foundations*, (1002-1015). New York: NY: Cambridge University Press.

Dollinger, S. J., Leong, F. T. & Ulicni, S. K. (1996). On traits and values: With special reference to openness to experience. *Journal of Research in Personality, 30*(1), 23-41.

Ennis, R. (1991). Critical thinking: a streamlined conception. *Teaching Philosophy, 14*(1), 5–24. doi: 10.5840/teachphil19911412

Fischhoff, B. (1981). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty*: Heuristics and biases (422-443). New York, NY: Cambridge University Press.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*, *18*, 253–292. doi: 10.1016/0010-0285(86)90001-0

Fong, G. T., & Nisbett, R. E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General*, *120*(1), 34–45. doi: 10.1037/0096-3445.120.1.34

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25-42.

Garcia-Retamero, R., & Cokely, E. T. (2011). Effective communication of risks to young adults: Using message framing and visual aids to increase condom use and STD screening. *Journal of Experimental Psychology: Applied*, *17*(3), 270-287.

Garcia-Retamero, R., & Cokely, E. T. (2013). Communicating health risks with visual aids. *Current Directions in Psychological Science*, *22*(5), 392-399.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart.* New York, NY: Oxford University Press.

Gorman, S. E., & Gorman, J. M. (2016). *Denying to the grave: Why we ignore the facts that will save us*. New York, NY: Oxford University Press.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105-2108.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814-834. doi: 10.1037//0033-295X.108.4.814

Haidt, J. (2012a). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Random House, Inc.

Haidt, J. (2012b). Self-scorable MFQ30. *MoralFoundations.org*. Retrieved from http://moralfoundations.org/MFQ30.self-scorable.doc

Haidt, J., & Björklund, F. (2008). Social intuitionists answer six questions about moral psychology. In W. Sinnot-Armstrong (Ed.), *Moral psychology* (Vol 2), (181-217). Cambridge, MA: MIT Press.

Hilton, D. J. (2003). Psychology and the financial markets: Applications to understanding and remedying irrational decision making. In I. Brocas & J. D. Carrillo (Eds.), *The psychology of economic decisions* (Vol. 1) (273–297). Oxford, England: Oxford University Press.

Kahan, D., Peters, E., Dawson, E., & Sovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy,1*(1), 54-86. doi:10.1017/bpp.2016.2

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature climate change*, *2*(10), 732.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus, and Giroux.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, *80*(4), 237-251.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgments under uncertainty: Heuristics and biases.* New York, NY: Cambridge University Press.

Klein, G. A. (1999). *Sources of power: How people make decisions* (20th Ed). Cambridge, MA: MIT Press.

Kokis, J. V., Macpherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, *83*(1), 26-52.

Kuhn, D. (1991). *The Skills of Argument.* In J. E. Adler & L. J. Rips (Eds.), *Reasoning: Studies of human inference and its foundations*. New York, NY: Cambridge University Press.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480-498.

McElroy, T., & Dowd, K. (2007). Susceptibility to anchoring effects: How openness-to-experience influences responses to anchoring cues. *Judgment and Decision Making*, *2*(1), 48-53.

Mercier, H., & Sperber, D. (2009). Intuitive and reflective inferences. In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond,* (149-170). Oxford, UK: Oxford University Press.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57-74.

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 129-140.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175–220. doi: 10.1037/1089-2680.2.2.175

Niu, L., Behar-Horenstein, L. S., and Garvan, C. W. (2013). Do instructional interventions influence college students' critical thinking skills? A meta- analysis. *Educational Research Review*, *9*, 114–128. doi: 10.1016/j.edurev.2012.12.002

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, *90*(4), 339-363.

Nowak, M., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*, *364*, 56-58.

Osherson, D., Perani, D., Cappa, S., Schnur, T., Grassi, F., & Fazio, F. (1998). Distinct brain loci in deductive versus probabilistic reasoning. *Neuropsychologia*, *36*(4), 369-376.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, *17*(5), 407-413.

Reyna, V. F., & Farley, F. (2006). Risk and rationality in adolescent decision making. *Psychological Science in the Public Interest*, *7*(1), 1–44.

Reyna, V. F., & Lloyd, F. J. (2006). Physician decision making and cardiac risk: Effects of knowledge, risk perception, risk tolerance, and fuzzy processing. *Journal of Experimental Psychology: Applied*, *12*(3), 179–195.

Sá, W. C., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, *91*(3), 497-510.

Sloman, S. A., & Rabb, N. (2016). Your understanding is my understanding: Evidence for a community of knowledge. *Psychological Science*, *27*(11), 1451-1460.

Sunstein, C. R. (2005). Moral heuristics. *Behavioral and Brain Sciences*, *28*(4), 531-541.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, *25*(4), 359-393.

Sperber, D., & Mercier, H. (2012). Reasoning as a social competence. In H. Landemore & J. Elster (Eds.), *Collective wisdom: Principles and mechanisms*, (368-392). New York, NY: Cambridge University Press.

Stanovich, K. (1999). *Who is Rational? Studies of Individual Differences in Reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates.

Stanovich, K. (2012). On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning. In K. J. Holyoak & R. G. Morrison (Eds) *The Oxford handbook of thinking and reasoning*, (343-365). New York, NY: Oxford University Press.

Stanovich, K., & West, R. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*(2), 342-357.

Stanovich, K. & West, R. (1998a). Individual differences in rational thought. *Journal of Experimental Psychology: General*. *127*(2), 161-188.

Stanovich, K. & West, R. (1998b). Who uses base rates and P(D/~H)? An analysis of individual differences. *Memory & Cognition*. *26*(1), 161-179.

Stanovich, K. & West, R. (2008a). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*(4), 672-695.

Stanovich, K. & West, R. (2008b). On the failure of cognitive ability to predict myside and one-sided thinking biases. *Thinking & Reasoning*, *14*(2), 129-167.

Stanovich, K., Toplak, M., & West, R. (2008). The development of rational thought: A taxonomy of heuristics and biases. *Advances in Child Development and Behavior 36*, 251-285.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797-811.

Thaler, R. H., and Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

Toplak, M., West, R., & Stanovich, K. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7): 1275.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1981). *Evidential impact of base rates* (No. TR-4). Office of Naval Research, Arlington, VA.

Vendetti, M. S., Wu, A., & Holyoak, K. J. (2014). Far-out thinking: Generating solutions to distant analogies promotes relational thinking. *Psychological Science*, *25*(4), 928-933.

West, R., Meserve, R. J., & Stanovich, K. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of Personality and Social Psychology, 103*(3), 1-14.

West, R., Toplak, M., & Stanovich, K. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology*, *100*(4), 930-941.

Westen, D., Blagov, P. S., Harenski, K., Kilts, C., & Hamann, S. (2006). Neural bases of motivated reasoning: An fMRI study of emotional constraints on partisan political judgment in the 2004 U.S. presidential election. *Journal of Cognitive Neuroscience*, *18*(11), 1947-1958.

Willingham, D. T. (2008). Critical thinking: Why is it so hard to teach? *Arts Education Policy Review*, *109*(4), 21–32. doi: 10.3200/AEPR.109.4. 21-32

Table 1

Primary Demographics by Group

	Total	Control	Monetary Judgments		
Demographics	M (SD)	M (SD)	M (SD)	t	p
Age (n = 112)	34.32 (10.54)	34.86 (11.73)	33.90 (9.58)	0.46	0.65
Sex (n = 112)	63.4% male	61% male	66% male	0.24	0.82
Education (n = 112)	4.38 (1.39)	4.43 (1.50)	4.33 (1.31)	0.36	0.72
Trait Measures					
AOT (n = 112)	22.38 (15.06)	16.65 (15.91)	26.84 (12.80)	-3.66	<0.01
MFQ30 (n = 85)	-4.74 (21.72)	-12.53 (21.21)	1.55 (20.22)	-3.12	<0.01
Political Identity (n =	3.01 (2.02)	3.51 (2.11)	2.62 (1.87)	2.36	0.02

Note. In the case of frequences, chi-square tests were used in place of t-tests.

Table 2

Outcome & Decision Pattern Differences in Control and Monetary Judgments Groups

	Control	Monetary Judgments		
Outcomes	M (SD)	M (SD)	t	p
Reversals	0.59 (1.10)	0.14 (.43)	2.70	< 0.01
Changes	1.49 (1.68)	1.27 (1.44)	0.74	0.46
Rewards	1.71 (2.20)	1.44 (2.03)	0.67	0.50
Abstentions	2.80 (2.79)	2.95 (2.49)	-0.31	0.76
Punishments	2.47 (2.82)	2.60 (2.86)	-0.25	0.81
Decision Pattern	1.59 (.84)	1.67 (.78)	-0.49	0.63

Table 3

Trait Characteristics for Total Sample

	M (SD)	AOT Difference	MFQ30 Difference	Political Identity
AOT Difference (n=112)	22.38 (15.06)	1	0.72*	-.51*
MFQ30 Difference (n=85)	-4.74 (21.72)		1	-.67*
Political Identity	3.01 (2.02)			1

*$p < .001$

Table 4

Regression Results Correlating Trait Measures with Outcomes

Variable	df	*B*	*SE B*	β	t	*p*	R^2
Changes							
AOT	110	-0.15	0.26	-0.14	-1.52	0.13	0.02
MFQ30	83	-0.01	0.01	-0.17	-1.55	0.13	0.03
Political Identity	110	0.10	0.07	0.13	1.42	0.16	0.02
Reversals							
AOT	110	-0.02	0.01	-0.28	-3.10	<.01	0.08
MFQ30	83	-0.01	<.01	-0.22	-2.08	0.04	0.05
Political Identity	110	0.07	0.04	0.17	1.77	0.08	0.03

Table 5

Subset Analyses: Regression Results for Participants Who Made Changes

Variable	df	*B*	*SE B*	β	t	*p*	R^2
Changes							
AOT	57	-0.02	0.01	-0.23	-1.79	0.08	0.05
MFQ30	48	-0.01	0.01	-0.09	-0.62	0.54	0.01
Political Identity	57	0.07	0.08	0.12	0.94	0.35	0.02
Reversals							
AOT	57	-0.02	0.01	-0.37	-3.00	<.01	0.14
MFQ30	48	-0.01	0.01	-0.24	-1.69	0.10	0.06
Political Identity	57	0.10	0.07	0.19	1.46	0.15	0.04

Table 6

Logistic Regression Results Predicting Reversals

Model	Coefficient (β)	*p* -value	Odds Ratio	95% CI
1. AOT	-0.04	0.03	0.96	0.95-0.98
2. Political identity	0.24	0.05	1.27	1.15-1.40
3. Moral foundations	-0.02	0.21	0.98	0.97-1.00

Table 7

Associations of Response Patterns with Traits

Pattern	df	*B*	*SE B*	β	t	*p*	R^2
AOT	110	<.01	0.01	0.04	0.42	0.68	<.01
MFQ30	83	<.01	<.01	0.06	0.51	0.61	<.01
Political	110	-0.02	0.04	-0.05	-0.56	0.58	<.01

*p < .05