Summer 2015

Myside bias in probabilistic ethical decision making

Richard E. Hohn
James Madison University

Myside Bias in Probabilistic Ethical Decision Making

A thesis submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

For the degree of

Master of Arts

Psychological Sciences

August 2015

Myside Bias in Probabilistic Ethical Decision Making

Richard E. Hohn

James Madison University

Author Note

Richard E. Hohn, Department of Graduate Psychology, James Madison University.

This research was supported by a grant from the John Templeton Foundation to Keith E. Stanovich and Richard F. West. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the John Templeton Foundation.

Correspondence concerning this study should be addressed to Richard E. Hohn, Department of Graduate Psychology, James Madison University, Harrisonburg, VA 22807. E-mail: hohnre@jmu.edu

Table of Contents

List of Tables

List of Figures

Abstract

This study examined the extent to which myside bias was observed in the context of ethical decision making. In judgment and decision making research the resistance of myside bias is described as a component of rational thinking. Thus, to some effect, this study aimed to examine the extent to which people act rationally when making ethical decisions. Measures of thinking dispositions (actively open-minded thinking and need for cognition) as well as measures of cognitive reflection and probabilistic knowledge were included in the study to examine their associations with myside bias in ethical decisions. Modest myside bias effects were observed between and within-subjects, however, for within-subjects bias the effect failed to reach statistical significance. Exploration of the data beyond the study's main research questions revealed significant myside bias effects for participants categorized as utilitarian. Finally, actively open-minded thinking was found to moderate myside bias effects.

Keywords: ethics, decision making, heuristics and biases, dual-process theory, myside bias

Myside Bias in Probabilistic Ethical Decision Making

There exists in most cultures a common sentiment that we humans are the only true 'rational beasts' on this planet and that, on the whole, the accumulation of human knowledge and the means by which we reason represent a pinnacle of mental complexity and sophistication. While in many ways that common sentiment rings true, many people are surprised to learn that we humans are not quite as rational as we tend to believe (see Kahneman, 2011; Stanovich, 2013). Even after three decades of research that outlines the deficiencies of human reasoning as well as the proneness we share to exercise biased thinking and form faulty heuristics, there still remains resistance to the notion that humans are not inherently rational (Stanovich, 2011). Moreover, it has been found that many of the systematic thinking bias and heuristics common among people are either not associated, or associated only modestly, with cognitive ability, suggesting that even those of high intelligence are prone to make use of such heuristics and biases (Stanovich, 2009, 2011; Stanovich & West, 2008b).

While the breadth of heuristic and biases research has grown substantially over the last 30 years, the focus of the present study concerns *myside bias*. Broadly defined, "myside bias occurs when people evaluate evidence, generate evidence, and test hypotheses in a manner biased towards their own prior opinions and attitudes" (Stanovich, West, & Toplak, 2013, p. 259). Put another way, the term "myside bias" describes a process in which an individual skews the objectivity of evidence, arguments, or conclusions in way that disproportionally places more strength and significance in

those which favor their own personally held beliefs, opinions, or attitudes. Beginning in 1990s, myside bias began to attract the focus of judgment and decision making researchers. Baron (1995), for example, examined myside bias as it concerned arguments about abortion. Baron found that participants rated arguments that favored their stance on the issue of abortion as significantly better than arguments of equal strength that conflicted with their stance. Baron also found that participants rated one-sided arguments as significantly better than two-sided arguments. Finally, Baron's experiments pertaining to myside bias yielded results that suggested actively open-minded thinking, a multifarious thinking disposition, may be negatively associated with myside bias as well as the preference for one-sided arguments.

The work of Baron is only a sample of his extensive research of myside bias and of the body of research of myside bias as a whole. Like the work of Baron, many examinations of myside bias use the same measurement framework of comparing the rating of written arguments of polarizing issues. There are, however, other ways in which myside bias has been examined in the literature, which have demonstrated that other factors such as group membership can induce myside bias. For example, an instance of group membership induced myside bias occurs when one values products built by their own country more when compared to equivalent products of another country (Stanovich & West, 2008b).

Stanovich and West's (2008b) "German Car Problem," the most relevant empirical study of myside bias to the present study, examined myside bias between-subjects by showing participants one of two similar prompts:

Condition 1:

According to a comprehensive study by the U.S. Department of Transportation, a particular German car is 8 times more likely than a typical family car to kill occupants of another car in a crash. The U.S. Department of Transportation is considering recommending a ban on the sale of this German car.

Condition 2:

According to a comprehensive study by the U.S. Department of Transportation, Ford Explorers are 8 times more likely than a typical family car to kill occupants of another car in a crash. The Department of Transportation in Germany is considering recommending a ban on the sale of the Ford Explorer in Germany.

In the first condition, participants were asked to rate the degree to which they thought the German car should be banned from sale in United States on a 6-point scale (*Definitely Yes* to *Definitely No*), as well as the degree to which they thought the German car should be allowed on U.S. streets. Similarly for the second condition, participants were asked to rate the degree to which they thought the Ford Explorer should be banned from sale in Germany and the degree to which they thought the Ford Explorer should be allowed on the streets of Germany.

Responses between the two conditions were found to be noticeably disparate from one another. In the first condition, 78.4% of participants thought that, to whatever degree, the German car should be banned from sale in the United States. Nearly 74% of those same participants thought that, again to whatever degree, the German car should not be allowed on U.S. streets. Conversely, in the second condition, responses were 51.4% and 39.2%, respectively, for the Ford Explorer. These differences statistically

significant (Stanovich & West, 2008b; Stanovich et al, 2013).

As one can imagine, the presence and prevalence of myside bias can affect a host of decisions not limited to those found in academic settings. Rather, particularly for controversial issues, the presence of myside bias has the potential to distort one's objectivity when evaluating arguments or making decisions that carry with them substantial implications (Baron, 1995; Stanovich & West, 2008a, Stanovich & West, 2008b). There is certainly an appreciable amount of research that examines myside bias in multitude of contexts (for examples see Baron, 1995; Evans, 2002; Stanovich & West, 2007; Toplak & Stanovich, 2003).

Myside bias, like nearly all biases and heuristics in the judgment and decision making literature, is often examined through the lens of *dual-process theory* (DPT). The breadth of DPT is much too expansive to fully disseminate here and there is a large amount of variance surrounding the particulars of the theory across researchers. However, it is important to note that for the purposes of the present study, the default-interventionist perspective espoused by Evans and Stanovich (2013) is maintained. Hence, any mention of DPT refers to the default-interventionist perspective, which is expanded below.

According to dual-process theories in the context of judgment and decision making research, the mind engages in two distinct types of cognitive processing: Type 1 processing and Type 2 processing. Type 1 processes are considered the default element of the default-interventionist perspective. Type 1 processing has been characterized as fast, intuitive, heuristic-based, unconscious, automatic, associative, and autonomous (Evans, 1984, 1989; Kahneman, 2011, Sloman, 1996; Stanovich, 2009, 2011; Wilson, 2002). As

Stanovich writes (2011) the defining feature of Type 1 processing is its autonomy and the implications that follow from having an autonomous processing system (see also Evans & Stanovich, 2013). One such implication mentioned by Stanovich is that executing a Type 1 process is mandatory upon encountering a triggering stimulus. In other words, Type 1 processing describes the instantaneous and implicit associations individuals make when encountering particular types of problems, contexts, situations, or more generally, stimuli. As a result of its automaticity, Type 1 processing is also computationally inexpensive and because of this, Type 1 processes may operate in parallel with one another.

Opposite of Type 1 processing, Type 2 processing is characterized as slow, analytic, rule-based, language-based, goal-oriented, conscious, and reflective (Evans, 1984, 1989; Kahneman, 2011, Sloman, 1996; Stanovich, 2009, 2011; Wilson, 2002). The defining feature of Type 2 processing is that it is a controlled kind of processing which *intervenes* on a default Type 1 response. Commonly used terms such as, 'hypothetical thinking', 'abstract thinking', 'critical thinking', and so on all fall under the umbrella of Type 2 processing. This kind of processing, counter to Type 1 processing, is serial and computationally expensive.

Much of the judgment and decision making literature concerning DPT illustrates that the failure to engage in Type 2 processing and intervene on a Type 1 response is a main source of thinking errors, many of which are the result of learned heuristics and biases associated with Type 1 processing. Therefore, when empirical research concerning DPT suggests that Type 1 processing is favored over Type 2 processing, it is not that individuals necessarily lack the capacity or capability to engage in Type 2

processing. Rather, it is more likely that individuals simply do not often engage in Type 2 processing in the first place and thus do not often intervene on Type 1 responses.

As a result, thinking dispositions such as the tendency to engage in cognitive reflection, consider many alternative possibilities, and enjoy abstract or "deep" thinking have all been found to be associated with higher utilization of Type 2 processing and in many cases resistance to cognitive biases (Stanovich, 2009, 2011). Returning to myside bias, it is thought by some that myside bias is best characterized as a Type 1 response as it is unrelated to many measures associated with Type 2 processing (Stanovich & West, 2008b).

The present study examines myside bias in the context of ethical decision making. Specifically, the present study seeks to examine whether myside bias effects are demonstrated when participants are tasked with making probabilistic decisions in which a strong ethical component is present. In the context of DPT, moral psychologists have proposed that many ethical decisions operate within the same framework. For example, Haidt's (2009) social-intuitionist model postulates that moral decisions are often made using intuitions (Type 1) rather than strategic reasoning (Type 2) and according to the model, moral rationales (Type 2 responses) are made only after a default moral intuition is evoked (Type 1 response). Thus, Haidt's model conforms to the default-interventionist perspective of DPT. Other findings within moral psychology reveal that absolute moral judgments are more associated with Type 1 processing, while utilitarian moral judgments are associated with Type 2 processing (Greene, 2007; Paxton, Ungar, & Greene, 2011).

While myside bias has been evaluated in a number of different ways, most often through the use of argument evaluation tasks, it has seldom been examined in the context

of ethical decision making. As such, the present study makes use of a novel approach to investigating myside bias. Specifically, the present study asks participants to make life-saving decisions by allocating organs among two groups of patients, while each group of patients has associated with it a different chance of survival than the other group. Of the four ethical decision making items that use this format in the present study, one contains a myside bias element. More information regarding these items is provided in the Methods section.

Organ Allocation as an Ethical Decision Making Task

The intended means of evaluating myside bias in probabilistic ethical decision making is the result of combining a number of different item structures and ideas from numerous sources. That is, the conception of my approach is influenced by several prior studies. First among those influences is the work of Ubel and Loewenstein (1995, 1996). Ubel and Loewenstein's research at the time concerned gaining insight into the public's preference for organ allocation. Specifically, Ubel and Loewenstein were interested in evaluating whether the public's preference for organ allocation was commensurate with the national policy, which at the time was to prioritize organ allocation to those in the most critical conditions, rather than necessarily those with the best chance of surviving if given an organ. To do so, Ubel and Loewenstein (1996) created the following item:

> We are going to ask you about a health issue. There are no "right" or "wrong" answers to these questions. As you may know, there is a shortage of livers available for those who need transplants. This problem is especially bad for children. Suppose that 200 children are waiting to receive a liver transplant, none of whom have any other health problems. They need to receive these transplants within one year or they will die. In that time, only 100 usable livers will become available. Children who do not receive a transplant will die.
>
> A blood test is available that divides the children into two groups, each with a different chance of surviving transplant. No other information predicts their

outcomes as reliably as this blood test.

Group 1 (100 patients): 80% chance of surviving if transplanted
Group 2 (100 patients): 50% chance of surviving if transplanted

What percentage of the 100 livers do you think should go to each group?

Group 1: _____%
Group 2: _____ %

Total**:** 100%

The intention of Ubel and Loewenstein (1996) was not to examine ethical decision making, *per se*, but rather to examine whether people's preference to allocate organs in a utilitarian way (i.e., allocate to those with the highest chance of surviving if given the organ, which saves the maximum number of lives, as opposed to giving the organs to those in the most critical conditions). However, their item appears to put participants within an ethical context by resting the responsibility of liver allocation on them. Because there are too few livers to allocate to everyone, the participants must consider their own moral standing, as well as consult the probabilistic evidence (i.e., the transplant survival rates) to come to a final allocation decision.

In the present study, the item from Ubel and Loewenstein (1996) will be modified to use the German Car Problem's myside manipulation as a model, such that group identities will become the United States and Germany. Thus, in this attempt, the Ubel and Loewenstein item will be modified to reflect prior research of myside bias in which the two group identities elicited a noticeable myside bias effect. However, unlike the German Car Problem no *directly* salient name (e.g., the Ford Explorer) is used, though using the United States and Germany as group identities provides more myside salience to the participants than do the Group 1 and Group 2 identities used by Ubel and

Loewenstein, as is reflected in the "German Car Problem" (see Appendices A and B).

The present study contains three purposes for evaluating myside bias in the context of ethical reasoning. The first and broadest purpose is to examine myside bias as it manifests in an ethical decision making task. The second purpose of the present study is to examine the associations between organ allocation tendencies in the ethical decision making task and several measures commonly used in the judgment and decision making literature. The judgment and decision making measures include a measure of cognitive reflections, measures of thinking dispositions associated with Type 2 processing, and a measure of probabilistic knowledge. The final purpose of the present study is to explore further aspects of organ allocation tendencies using supplemental measures included in the study. Towards this purpose, no formal hypotheses are given. Rather, the findings obtained using the supplemental measures are intended to inform future research and generate hypotheses for future research. The supplemental measures include measures of probability matching, the tendency to rely on base rates in predictive decision tasks, and ethical dilemmas, which were used to assess the ethical standing of participants as either absolutist or utilitarian, and other potentially informative tasks.

Derived from the study's first purpose are two research questions. The first research question asks to what extent a between-subjects myside bias effect is observed. It was hypothesized that a between-subjects myside bias effect would be observed and that this effect will favor organ allocation towards the United States group. A second research question asks to what extent a within-subjects myside bias effect is observed. That is, it is asked whether individual allocation tendencies change when allocating for groups *without* a myside element (i.e., Country A and Country B) when compared to their

allocation tendencies for groups with a myside element (i.e., Germany and United States) in a way that favors the United States groups. Towards this research question it was hypothesized that a within-subjects myside effect that favors the United States group would be observed.

A third and final research question was derived from the second purpose of the present study. This research question asks in what ways are organ allocation tendencies towards high survival groups associated with two thinking disposition measures, the Need for Cognition Scale (Cacioppo, Petty, Feinstein, & Jarvis, 1996) and a new version of the Actively Open-minded Thinking Scale (Stanovich & West, *under contract*), as well as two cognitive measures, the Cognitive Reflection Test (Frederick, 2005) and a Numeracy Scale (Lipkus, Samsa, & Rimer, 2001) Further, the third research question ask how the four measures might in isolation or conjunction moderate any observed myside bias effects. It was hypothesized that all four measures would be positively associated with the allocation of more organs to the high survival groups in each organ allocation task. Further, it was hypothesized that one or some combination of these measures would moderate myside bias effects.

Methods

Participants

A sample of 300 participants (175 men, 125 women, M_{age} = 33.87, SD = 9.99, age range: 18-71 years) was collected using Amazon's Mechanical Turk (Mturk) online service.[1] Participants were compensated four dollars for roughly twenty minutes of their time (M_{time} = 22.87 minutes). In order to facilitate the myside bias manipulation, only

participants from the United States were allowed to participate. A majority of the sample had received some degree of college education, as 36.8% had completed 'some college', 40.1% had obtained a bachelor's degree, and 8.6% had obtain a graduate level degree. Of the remaining participants, 13.2% had complete high school or earned their GED.

Materials

A myriad of different tasks were used in the present study. Among them are measures of organ allocation, thinking dispositions, cognitive reflection, and numeracy. Further, two follow up questions to the allocation tasks, a measure of probability matching, and two ethical dilemmas are included for exploratory purposes. Each of these measurement instruments (sans the Actively Open-minded Thinking Scale) are included in the Appendices.

Materials Related to the Research Questions

Measures of allocation (MoA) – Myside: Livers. Drawing from past research involving allocation decisions (Ubel & Loewenstein, 1995, 1996), an item was modified to include a myside bias element used in prior myside bias research (Stanovich & West, 2008b). Specifically, the myside bias element is the salience of the group identities as being the United States and Germany. For this particular item, participants were asked to distribute 100 livers between two groups of 100 children each, 100 children in the United States and 100 children in Germany. Participants were told that without these livers for transplant, all of the children will die. However, each group had attached to it a different probability of survival for transplant procedures. In condition 1, participants were given a version of the problem in which the United States has attached to it an 80% chance of transplant survival while Germany had attached to it a 50% chance of transplant survival.

In condition 2, the survival rates were reversed between the two groups (i.e., US 50% ; GER 80%). Participants were given the question prompt and asked to enter the raw number of livers they choose to allocate to each group into two text fields, one for the United States and one for Germany, which together were required to total 100.

Measures of allocation – Non-myside: Kidneys, heart valves, and surgical procedures. Beyond the myside MoA, three other MoAs given to the participants. Those other three MoAs did not incorporate a myside bias element, such that the group identities were not myside salient to the participants (e.g., Town A and Town B). Further, the amount of children in need of transplants (as well as group sizes) and the probability of survival for each of the groups also varied. The first non-myside MoA is intended to be directly compared to the myside MoA. Thus, the total number of children (200) and the group sizes (100 each) are identical in the two problems. The probability of surviving the transplant for the groups are also identical (e.g., 80% and 50%). For the first non-myside MoA, however, the organs being allocated are kidneys as opposed to livers. Note that because the groups are not myside salient in the non-myside MoAs, there is no need for two conditions of the items. Therefore, each condition from the myside MoA sees the same non-myside MoAs. The second non-myside MoA asked participants to allocate artificial heart valves among a total of 800 children (400 in each group) at survival rates of 40% and 25%. Finally, the third non-myside MoA asked participants to allocate surgical procedures (instead of organs) among 200 total children (100 in each group) at survival rates of 70% and 30%. The final non-myside MoA intends to induce probability matching, wherein participants who probability match think that giving 70 surgical procedures to the 70% survival rate group and 30 surgical procedures to the 30% survival

group saves the most lives. Responses for these items were recorded in the same manner as the myside MoA described above.[2]

Cognitive Reflection Test. The Cognitive Reflection Test (CRT) is designed to measure the tendency to override an intuitive response alternative that is incorrect and instead engage in further reflection that leads to a correct response. It consists of three items such that the correct response to each item requires at least some reflection on behalf of a participant. Participants responded to the items by providing their answer in a blank text field. Scores for participants range from 1-3, the totaled sum of correct responses. An example item from the CRT is, "If a bat and a ball together cost $1.10 and the bat costs $1 more than the ball, how much does the ball cost?"

Actively Open-minded Thinking Scale. An updated version of the Actively Open-minded Thinking (AOT) Scale was used (Stanovich & West, *under contract*) in the present study. The AOT Scale is composed for 30 items that tap flexible thinking, openness, dogmatism, categorical thinking, and counterfactual thinking. Responses for the scale were recorded on a 6-point scale: *Strongly Agree* (6), *Moderately Agree* (5), *Slightly Agree* (4), *Slightly Disagree* (3), *Moderately Disagree* (2), and *Strongly Disagree* (1). Total scores were generated by taking the sum of each of the 30 items, whereby higher scores reflect higher measures of actively open-minded thinking. Some example items include, "People should always take into consideration evidence that goes against their beliefs," and, "Certain beliefs are just too important to abandon, no matter how good a case can be made against them" (reverse scored).

Need for Cognition Scale. The Need for Cognition (NFC) Scale measures "the tendency for an individual to engage in and enjoy thinking" (Cacioppo et al, 1996).

Cacioppo et al surveyed 2797 participants over 11 studies and found that alpha coefficients for the scale ranged from .81 to .97. The NFC Scale consists of 18 items and the responses were reported on a 6-point scale: *Strongly Agree* (6), *Moderately Agree* (5), *Slightly Agree* (4), *Slightly Disagree* (3), *Moderately Disagree* (2), and *Strongly Disagree* (1). Total scores were generated by taking the sum of the 18 items, whereby higher scores reflect higher measures of need for cognition. Some example items include, "Learning new ways to think doesn't excite me very much" (reverse scored), and, "The notion of thinking abstractly is appealing to me."

Numeracy Scale. An 11-item numeracy scale was used to assess participants' prior knowledge of probability. The numeracy scale to be used in this study is the amalgam of two numeracy measures used by Lipkus et al (2001). Lipkus et al found in initial studies using the scale an alpha coefficient if .78. Responses to these items are provided in either a multiple choice or free response format. Items were scored as either correct or incorrect and a summed total score was calculated. Some example items are, "If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?" and, "The chance of getting a viral infection is .0005. Out of 10,000 people, about how many of them are expected to get infected?"

Materials for Exploratory Purposes

Measure of probability matching – The marble task. An item was written for this study that intends to assess participants' susceptibility to engage in probability matching when evaluating a problem involving the use of probabilistic thinking. The item asked participants to suppose they are playing a game in which they draw marbles from a bowl and will be rewarded a hypothetical $5 for each correct guess. The participants were told

that there were nine marbles in the bowl, three black and six white. They then were told that in this game marbles were replaced (i.e., after you draw a marble, it is put back into the bowl) and that they were to go through nine rounds of guessing which color marble they would draw. This item is not scored in a typical way (e.g., correct or incorrect, Likert scale, etc.) but instead flags those who probability matched. Thus, if participants guessed three black marbles and six white marbles over the nine rounds, then they were labeled as probability matchers for this item. Participants who guess all white marbles are considered "maximizers."

Measure of allocation follow up questions. Following the final non-myside MoA were two questions. First, participants were asked if they believe that their allocation strategy saved the most lives possible. For this question, responses were recorded on a 4-point scale with the option of *Definitely Yes* (4), *Probably Yes* (3), *Probably No* (2), and *Definitely No* (1). Second, the participants were asked if they would prefer to ignore the given survival rates and instead have the surgical procedures randomly distributed between the two groups. Responses for the second follow up question were recorded on a simple 2-point scale of options *Yes* and *No.*

Ethical dilemmas – The Trolley Car Problem. A modified version of the "Trolley Car Problem" was used (Greene, Morellia, Lowenberg, Nystorm, & Cohen, 2008). Participants were presented with two dilemmas in which they must decide whether or not to intervene in a situation in which a runaway trolley is going to kill five people. In the first problem, participants were given the option of pulling a lever that redirects the trolley onto a different set of tracks. However, in doing so, one person caught on those tracks would be killed. In the second problem, participants had to decide whether to push

a fat man off of a footbridge and onto the tracks in front of the runaway trolley. If the participants refused, five people further down the tracks would die. If the participants pushed the man onto the tracks, his size would be enough to stop the trolley and save the five people, but he would die. Though the context for each problem is markedly different, in either case the participants had to choose to saving the lives of five people or saving the life of one person. These ethical dilemmas were used by the researchers to give some indication of the participants' moral stance given the conditions (i.e., absolutist or utilitarian).

Procedure

United States users of Amazon's Mturk selected themselves into participation of the present study by selecting the study's posting among other Mturk offers presented to them. After selecting the present study, participants were redirected to a Qualtrics survey containing the materials listed above. Participants were first shown informed consent documents in which they needed to electronically acknowledge prior to beginning the survey. Upon receiving consent, participants were then shown the myside MoA, in which they were randomly selected into conditions (153 in condition 1 and 147 in condition 2). Following the (1) myside MoA, all participants completed the survey in the following order: (2) the CRT, (3) the NFC scale, (4) the Marble Task, (5) the AOT scale, (6) demographic information, (7) the non-myside MoAs (kidney, heart valves, and surgical procedures), (6) the MoA follow up questions, (9) the Numeracy Scale, and (10) the ethical dilemmas.

The tasks in the survey were ordered strategically in hopes that their placement would reduce the potential for priming effects. In particular, measures that were

suspected of priming Type 2 processing were distanced from the non-myside MoAs. It has been found in one study, for example, that the CRT acts as a successful prime of utilitarian moral reasoning (Paxton et al, 2011). That is, the CRT has been found to prime reflective thinking (i.e., Type 2 processing), which is more associated with utilitarian moral decisions than intuitive Type 1 responses, which are more associated with absolute moral decisions. More generally, tasks that require calculation or deliberation, such as the Numeracy Scale, prime Type 2 processing. The NFC scale, consisting of affective items rather than cognitive items like the CRT or the Numeracy Scale, was also a concern for priming. While no research has been found to support the notion that the NFC scale may prime Type 2 processing, it was the judgment of the researcher that it may have been problematic, as asking participants whether they are deep thinkers may lead some to have been self-conscious and perhaps to think more deeply about problems that followed the NFC scale. The ethical dilemmas were also suspected of eliciting a priming effect by potentially causing participants to think more deeply about their moral standing. For all of these reasons, the presentation of the tasks used in the present study was not counterbalanced. Instead, tasks were placed strategically to allow for a maximum amount of space between the myside and non-myside MoAs, as well as to prevent potential priming tasks from appearing before the non-myside MoAs.

Categorization of Participants

To explore the data for trends that might influence future research, participants were categorized into several ways. First, participants were categorized dependent upon how they responded to the MoAs. Three categories were created. Participants were

considered “maximizers” if they allocated all of the organs to the groups with the highest survival rates. Participants were categorized as “equal chancers” if they split the organs 50/50 among the two groups in the MoAs. Finally, participants were put into an “other” category if they used allocation strategies other than those mentioned. Only participants who were consistently maximizers, equal chancers, or others across all of the MoAs were categorized, resulting in a subset of 202 of the 300 participants (42 maximizers, 60 equal chancers, and 100 other).

Participants were also categorized as either utilitarian or absolutist dependent upon their responses to the ethical dilemmas. Utilitarians are those participants who chose to save five lives over one life in the ethical dilemmas. Absolutists are those that chose to save only one life in the ethical dilemmas. These labels are admittedly crude. The term absolutist refers to one who is likely think ethically in terms of absolute moral truths. Therefore, in the ethical dilemmas, the act of pulling the switch or pushing the large man off of the footbridge would be considered murder, which to absolutists is absolutely unethical in any instance. The term utilitarian refers to one who believes that saving the maximum number of lives in the ethical choice, regardless of the means used to do so. Under the utilitarian view, then, pulling the switch and pushing the large man off of the footbridge are necessary actions to save the most lives and are therefore ethical.

While it is impossible to completely infer the degree to which a given participant is either absolutist or utilitarian using the ethical dilemmas, the categorization used in the present study was conservative, as only participants that responded in the same way to both dilemmas (n = 206, 90 absolutists, 116 utilitarians) were categorized. Thus, the judgment of the researcher is that using this categorization provides a useful distinction

for examining differences between the two ethical positions. Furthermore, the MoAs bare some semblance to the ethical dilemmas, as the participants' decisions could be based on saving the most lives (maximizers) or based on maintaining a position of absolute fairness by giving children of both groups equal chances of receiving and organ (equal chancers).

Results

Between-subjects Myside Bias

Towards the first research question, an independent samples *t*-test was conducted to assess the difference between the average number of organs allocated to the US80% group ($M = 69.29$, $SD = 19.27$) in condition 1 and the GER80% group ($M = 71.58$, $SD = 20.25$) in condition 2. The results of the one-tailed *t*-test showed a significant difference between the two conditions, $ts(298) = -1.714$, $p = .044$, $d = -.198$, $r = .099$, 95% CI [$-\infty$, -.157]. Following the test for between-subjects myside bias, analyses were conducted to assess the second research question that asked to what extent a within-subjects myside bias in favor of the United States would be observed.

Within-subjects Myside Bias

A two-way mixed analysis of variance (ANOVA) was conducted to assess the second research question. For the analysis, the 80% groups from both conditions (US80% and GER80%) were compared to the 80% group in the comparison non-myside MoA. Because the myside MoA and comparison non-myside MoA items were identical except for the group identities, these two items were collapsed into a within-subjects variable in which the myside MoA was thought of as time 1 and the comparison non-myside MoA was thought of as time 2.

Results of the mixed ANOVA are displayed in Table 1 and Figure 1. Though the interaction between time (i.e. myside versus non-myside) and condition (i.e. U.S. 80% versus Germany 80%) failed to be statistically significant $F(1, 298) = 2.826$, $p = .094$, $\eta_p^2 = .009$, it was deemed worthy of subsequent investigation. A follow-up t-test supported the between-subjects myside bias effect in allocation rates for the myside MoA, ts(298) = -1.714, $p = .044$, $d = -.198$, $r = .099$, 95% CI [-∞, -.157]. Individuals assigned to the United States favored and German favored condition reported similar allocation rates for the comparison non-myside MoA, ts(298) = -.357, $p = .721$, $d = -.041$, $r = .021$, 95% CI [-5.428, 3.761]. Individuals assigned to the German favored condition allocated more organs to the non-myside MoA ($M = 70.75$, $SD = 20.18$) than the myside MoA ($M = 65.07$, $SD = 23.320$), ts(146) = -3.567, $p < .001$, $d = .59$, $r = .283$, 95% CI [-8.828, -2.533]. Allocation rates failed to change across the myside MoA ($M = 69.29$, $SD = 19.27$) and non-myside MoA ($M = 71.58$, $SD = 20.25$) for the United States favored condition, ts(152) = -1.825, $p = .07$, $d = -.148$, $r = .074$, 95% CI [-4.764, .189].

The Relationships of CRT, AOT, NFC, and Numeracy on Organ Allocation

The third research question asked to what extent CRT, AOT, NFC, and Numeracy are associated with allocating organs to the high survival groups in the MoAs and to what extent might these variables or a combination of these variables moderate any observed myside bias effects. Bivariate correlations were obtained to examine these relationships (see Table 2). Nearly all correlations between the scales and MoAs were positive with the exception of the relationship between NFC and one of the non-myside MoAs ($r = -.021$, $p = .716$). Furthermore, none of the associations between NFC and any of the MoAs were significant.

The associations among the CRT, AOT, and Numeracy and the MoAs were all found to be statistically significant with two exceptions – CRT and AOT both failed to significantly correlate with the same non-myside MoA concerning the allocation of heart valves to groups with 40% and 25% survival rates (r = .092, p = .112 and r = .112, p = .053, respectively). The heart valves MoA was the only MoA not to use survival rates of 80% and 50%. Examination of the distributions of CRT and Numeracy scores revealed noticeable ceiling effects, with the majority of participants scoring the highest possible scores on these measures. Numeracy in particular displayed limited variability (M = 9.46, SD = 1.83, range: 1-11).

After observing the bivariate correlations and the distributions of the judgment and decision making variables, it was determined that AOT was the only variable that could viably moderate myside bias. This is because participants did not vary much on CRT or Numeracy and NFC was not associated strongly with the MoAs, but was strongly associated with AOT. Any moderation on the part of NFC, then, would potentially be spurious and reflect the moderating influence of AOT.

To test the moderating influence of AOT on the between-subjects myside bias effect, a hierarchical regression analysis was conducted that entered AOT, the between-subjects grouping variable from the myside MoA, and their interaction in three separate and sequential steps (see Table 3). The model with all three terms was found to be significant, $F(3, 296) = 10.707$, $p < .001$, $R^2 = .098$, with a significantly higher amount of variance explained by the interaction term, $\Delta F = 10.117$, $p = .002$, $\Delta R^2 = .031$. The interaction can be viewed graphically in Figure 2.

A similar hierarchical regression analysis was conducted to evaluate the

moderating influence of AOT on the within-subjects myside bias effect by using the difference scores of the MoAs as a dependent variable (see Table 3). The difference scores were calculated by subtracting the number of organs allocated to the 80% groups on the myside MoA from the number of organs allocated to the 80% group of the comparison non-myside MoA. The results of the hierarchical regression analysis showed that the model containing AOT, the between-subjects grouping variable, and the interaction was significant, $F(3, 296) = 4.387$, $p = .005$, $R^2 = .043$, with a significantly higher amount of variance explained by the interaction term, $\Delta F = 8.444$, $p = .004$, $\Delta R^2 = .027$. The interaction can be viewed graphically in Figure 3 and the points of intersection and regions of significance for both the between and within-subjects interactions can be viewed in Table 4.

Results of Exploratory Analyses

A number of other analyses were conducted to explore the data for findings that could generate future research questions and hypotheses. These analyses made use of the measures in the survey described as supplemental above. Further, various categorizations were made (see the Methods section for explanations) and used in the exploratory analyses.

First, a number of one-way ANOVAs were conducted to assess whether there were significant differences among various categories of participants for CRT, AOT, NFC, and Numeracy. There are reasons to suspect some categories of participants might differ from others on these measures. For example, those who can be called utilitarians and absolutists, as identified by the ethical dilemmas, might differ on measures such as CRT, as the engagement of cognitive reflection has been linked to utilitarian moral

judgments (Paxton et al, 2011).

A set of one-way ANOVAs were conducted that used a grouping variable derived from the participants' allocation tendencies. The ANOVAs compared maximizers, equal chancers, and those who used an alternate strategy (n = 202, 42 maximizers, 60 equal chancers, 100 "others") on their average scores on the judgment and decision making measures (see the Methods section). It was found that only the AOT and NFC variables resisted violations of normality and homogeneity of variances and among those two variables, only the ANOVA on AOT revealed significant differences among the three groups, $F(2, 201) = 10.998$, $p < .001$, $\eta_p^2 = .099$. Tukey pairwise comparisons for the ANOVA showed that for AOT, maximizers had significantly higher scores than the equal chancers and the other group, but the equal chancers and the other group did not statistically differ from each other.

A set of independent samples t-tests were also conducted that compared the means of the judgment and decision making variables between participants categorized as utilitarians and absolutists (n = 206, 116 utilitarians, 90 absolutists). No significant differences among AOT, CRT, NFC, and Numeracy were found. Following the t-tests, absolutists and utilitarians were separately analyzed to assess each subsets extent of observed myside biases.

Mixed-ANOVAs identical to the one conducted for the second research question were conducted separately for absolutists and utilitarians. For absolutists, no significant interaction or main effects were observed. For the utilitarians, however, a significant interaction was observed, $F(1, 116) = 9.422$, $p = .003$, $\eta_p^2 = .076$ (see Figure 4). This interaction suggests that utilitarians exhibited a within-subjects myside bias in favor of

the United States and the size of the effect was greater than the interactions found in the complete sample. Moreover, mean allocations made by utilitarians were higher than allocations made by absolutists (see Table 5), significantly so for the comparison non-myside MoA, *t*s(204) = -4.195, $p < .001$, $d = .58$, $r = .28$, 95% CI [-16.320, -5.885].

As with the analysis of the second research question, follow up tests for the interaction were conducted. Similar patterns as found for research question two were obtained, as the same pairs of cell means were found to be significant in the follow up tests. The *t*-test comparing the mean allocations to the 80% groups in the myside MoA revealed a significant difference between the two myside conditions, *t*s(114) = -2.347, $p = .021$, $d = -.44$, $r = .21$, demonstrating a significant between-subjects myside bias effect. A *t*-test comparing the myside and non-myside MoAs for the Germany favored condition also yielded significant results, *t*s(56) = -4.005, $p < .001$, $d = -.57$, $r = .27$.

As with the full sample, the moderating influence of AOT on the myside biases was investigated using the same procedures. The hierarchical regression model examining the influence of AOT in the between-subjects myside bias was found to be highly significant, $F(3, 112) = 11.59$, $p < .001$, $R^2 = .216$, with a significantly higher amount of variance explained by the interaction term, $\Delta F = 12.176$, $p = .001$, $\Delta R^2 = .083$, that in the models without it (see Table 6 and Figure 5). When performing the same analysis on the difference scores to examine the moderating influence of AOT on the within-subjects myside bias, the model containing the interaction was also highly significant, $F(3, 112) = 8.12$, $p < .001$, $R^2 = .179$, with the interaction term explaining a significant amount of variance over and above the other models, $\Delta F = 12.905$, $p = .001$, $\Delta R^2 = .095$ (see Table 6 and Figure 6). The points of intersection and regions of

significance for the moderation analyses can be viewed in Table 7.

Discussion

The present study consisted of three goals. The first goal was to examine the extent to which myside bias effects were observed both between and within subjects. It was hypothesized that both kinds of bias would be observed in a manner that favored the United States. Analysis of the data showed that a statistically significant between-subjects myside bias effect was observed while a nonsignificant within-subjects myside bias was observed. In either case, however, the meaningfulness of the effects is difficult to interpret. The between-subjects myside effect, for example, was found to have rather small effect sizes even though the *t*-test used to test for the between-subjects bias was statistically significant. In total for the between-subjects myside effect, only about 1% of the overall variance in organ allocation to the 80% group of the myside MoA can be explained by the condition of the myside MoA participants were assigned to ($r^2 = .0098$).

Conversely for the within-subject myside bias, statistical significance was not obtained but the follow up tests of the nonsignificant interaction revealed interesting behaviors between the two myside conditions. Specifically, it appeared that condition 2 (Germany favored) exhibited a within-subjects myside bias in favor of the United States. Moreover, it is not clear whether condition 1 (United Stated favored) exhibited a within-subjects myside. When the United States is favored in condition 1 allocation strategies do not necessarily need to be adjusted to compensate for United States children with low survival rates. That is, probabilistically it makes just as much sense to allocate organs to the US80% group in the myside MoA as it does to allocate to the 80% group of the non-myside MoA. It is therefore hard to parse out whether condition 1 allocated more organs

because of a myside bias or if allocations in that condition were different as a result of the favored survival rate. For the both the between and within subjects myside bias analyses, it is apparent that some degree of myside bias was observed. Nevertheless, the ambiguity surrounding these effects make coming to conclusions about the study's corresponding hypotheses difficult.

A second goal of the study was to examine the relationships between a set of judgment and decision making variable and allocations to the high survival groups in the MoAs, as well as determine if any of the judgment and decision making variables moderated any observed myside bias effects. While positive associations were generally found between CRT, AOT, NFC, Numeracy and the allocation of organs to the high survival groups, the magnitude of the correlations were modest, typically between $r = .1$ and $r = .2$. In some cases, the correlations were confounded by low variability and ceiling effects, specifically for the CRT and Numeracy variables. The issues of ceiling effects and low variability in those variables undermined their amenability towards being used in inferential statistics, as violations of statistical assumptions arose in all attempts to do so.

Regarding moderation, it was determined that AOT was the only viable measure whose moderating influence could be investigated. Interestingly, it was found that AOT moderated both the between and within subjects myside bias effects. Points of intersection and regions of significance were obtained for the two effects and it was found that at thresholds of 130 and 128.39 on AOT for the between and within subjects effects, significant differences in organ allocation dissipated between the two myside conditions. For both effects, organs allocation was found to be no different when AOT

score were around 143.

The moderating influence of AOT on myside bias is a particularly intriguing finding of the present study. Results of the one of the exploratory analyses revealed that AOT significantly differed for a subset of categorized participants. In particular, it was found that maximizers scored significantly higher on AOT than did equal chancers or other organ allocation strategies. It could be the case, then, the moderating influence of AOT reflects this difference and that those who are lower on AOT tend to be equal chancers or use other strategies. Furthermore, significant negative correlations between observed myside bias and AOT have been found in other research (Stanovich & West, 2008a), the findings of which align with the between and within subjects myside biases found in condition 2 (Germany favored). Given these results the third hypotheses can be thought of as partially corroborated.

A third and final goal of the present study was to explore the data for findings that might inform future research. In doing so, other measures not directly relevant to the research questions were analyzed. One particularly notable finding emerged from the exploratory analyses, which concerned the differences between participants categorized as absolutist and those categorized as utilitarian. It was found that absolutists to nearly no extent committed any kind of myside bias while utilitarians to large extent exhibited both between and within subject myside biases. On its surface this finding seems to be counterintuitive, as utilitarianism is often thought to be related to rational thinking, as it involves calculation and is more associated with Type 2 processing (see Paxton et al, 2001). In retrospect, however, it makes sense that absolutists would be firm and consistent in their allocations because they are likely adhering to an absolute ethical

value, which by its definition does not change. The notion that utilitarians where to a large extent biased, however, remains puzzling.

Unlike the findings of the entire sample, the subset of utilitarians exhibited between and within subjects myside bias effects that were highly statistically significant in both cases. The effect sizes for the analyses of utilitarians were also much larger than those obtained using the full sample. A couple of potential conclusions can be drawn from these findings. First, it is likely the case that the utilitarians, which make up over a third of the overall sample, are the driving force if the biases observed in the overall sample. Second, it could be the case that while some participants may be utilitarian in spirit, they fail at being utilitarian in practice, that is, they lack the proper knowledge of how to act in a utilitarian way. Observing a cross tabulation of the ethical dilemma and marble task categorizations shows that less than half of the utilitarians properly maximized in the marble task and over 27% (32 of 166) probability matched. Interestingly, as reported above, utilitarians and absolutists did not significantly differ on a measure of Numeracy. Thus, it is possible that even with an assumed knowledge of how to maximize, utilitarians still did not do so. It is also possible that tasks like the MoAs and the marble task that use two probabilities in a question create and environment in which probability matching is more readily applied, whereas the Numeracy Scale does not provide that same context.

Moderation analyses on the subset of the utilitarians also showed the amplified effect of AOT on the observed myside biased. The points of intersection were found to be higher than in the full sample. The regions of significance were also higher for the utilitarian subset, suggesting that even at higher levels of AOT, myside bias effects start

to emerge for utilitarians compared to the full sample.

Limitations

Throughout the research project, a number of limitations have become apparent and though the goals of the present study were all to some degree met, it is important to be cautious when interpreting the findings of the study. One of the study's limitations concerns the information provided by the MoAs. Largely, the distributions of responses for the MoA were bimodal as the two most common responses were to either split the organs 50/50 or allocate all of the organs to the high survival group (maximize). Therefore, while the MoAs were conceived to be continuous variable, their distributions indicate the items acted somewhat like categorical variables. There was still a feasible amount of variability in the MoA as the skewness and kurtosis values indicated no violations of normality. However, the large concentration of responses around several specific allocation strategies suggests some analyses cannot be interpreted as liberally as they can under regular circumstances. In particular, interpretation the moderation analyses must acknowledge that the prediction lines that indicate the interactions are not as accurate as they would be in analyses with more normally distributed dependent variables. The need to qualify these interpretations is also reflected in the regression models for the moderation analyses. Although the models presented were significant, the variance explained in the regression models tended to be low.

There was another limitation discovered concerning the MoAs, specifically the myside MoA. At the time of the myside MoA's creation it was not known by the researcher that livers are unlike other organs used in transplant situations. Transplant patients needing livers, in fact, do not always require an entire liver to survive and

recover from a liver transplant. There are many documented cases in which people donated half of their livers to someone in need with both the donors and recipients surviving and recovering. The extent to which this fact was known in the sample is unknown. Theoretically, it is plausible that a participant allocating organs 50/50 assumed the organs could be halved and thus, all children in each group would have a chance at survival. Such a strategy has the potential to save more lives than does a maximizing strategy.

Another limitation regards the categorizing of participants (e.g., utilitarian/absolutist and so on). Although participants were categorized conservatively based on their consistency, there is still not a high degree of clarity as to whether the categorization strongly represent that categories they refer to. For example, the categorization of utilitarians and absolutists might not accurately reflect the extent to which individuals are either utilitarian or absolutist. Furthermore, the ethical dilemmas used to categorize participants are not typically used for that purpose, at least not in isolation. These same concerns are echoed in the maximizer/equal chancer/other categorization, as well as the categorizations made by the marble task. When those two categorical variables were cross tabulated, it was found that some participants who maximized on the MoAs were flagged as probability matchers in the marble task. It is possible that the situation or context in which participants are placed in plays a role in whether they maximize or probability match.

Finally, it is important to acknowledge the novelty of this instance of myside bias research. Because the present study examines myside bias in a novel way, there exists no context in which the size of the effects can be compared, as there is among myside bias

research that use similar argument tasks. Taken together, all of these limitations suggest that while exciting effects were found, the interpretation of those effects must be made cautiously. Ultimately, it is the judgment of the researcher that some degree of myside bias effects were observed in an ethical decision making context, but the specificity of those effects are uncertain.

Future Direction of Myside Bias Research in Ethical Decision Making

There are a number of ways in which the present study is influential for future research of myside bias in ethical contexts. For example, modifications to the MoAs could be made and compared. One way in which the MoAs could be altered is by using countries other than Germany. Germany was originally chosen because of the countries relative compatibility with the United States. Both countries have democratic governments, capitalist economies, are technological leaders, and so on. The decision to use a compatible countries was rooted in avoiding potential cofounds that participants might have experience when comparing the two countries in the MoAs. However, it would be interesting to compare the magnitude of myside bias effects as they apply to different nation comparisons. For instance, one might expect a larger magnitude of between-subjects bias is the United States and a communist country were paired in a myside MoA.

Another alternative for future research could be to revise the items used in the present study. The effects observed in this study suggest that some degree of myside bias is present in the context of ethical decision making and there are perhaps better ways of measuring myside bias in this context than by using the MoAs. The same is true for the categorization variables, particular the utilitarian/absolutist distinction. Instead of

dichotomizing subjects as either utilitarian or absolutists, it would be prudent for future research to find continuous measures of utilitarianism and absolutist. If such continuous measures could be found or created, then those measures could be used to make more accurate predictions and distinctions. For example, one could examine how the extent to which a person is utilitarian moderated organ allocation or the results of some other ethical myside task.

Finally, future research could examine the role that educational level and cognitive ability play in the resistance of myside bias in ethical contexts. There does exist a large body of literature to suggest increases cognitive ability do not necessarily result in decreases in myside bias (see Stanovich & West, 2008a, 2008b). However, the extent to which the relationship between myside bias and cognitive ability persists in an ethical decision making context is not established.

References

Baron, J. (1995). Myside bias in thinking about abortion. *Thinking & Reasoning*, 1, 221–235. doi: 10.1080/13546789508256909

Buhrmester, M. D., Kwang, T., & Gosling, S. D. (in press). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3-5. doi: 10.1177/1745691610393980

Cacioppo, J. T., Petty, R. E., Feinstein, J., & Jarvis, W. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*, 197- 296. doi: 10.1037/0033-2909.119.2.197

Evans, J. St. B. T. (2002). The influence of prior belief on scientific thinking. In P. Carruthers, S. Stich, & M. Siegal (Eds.), The cognitive basis of science (pp. 193–210). Cambridge, England: Cambridge University Press.

Evans, J. St. B. E. (1989). *Bias in human reasoning: Causes and consequences*. Hove, UK: Lawrence Erlbaum Associates.

Evans, J. St. B. E. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 75*, 223-240.

Evans, J. St. B. E. & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223-241. doi: 10.1177/1745691612460685

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25-42. doi: 10.1257/089533005775196732

Greene, J.D., Morelli, S.A., Lowenberg, K., Nystrom, L.E., Cohen, J.D. (2008) Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, 107, 1144-1154. doi: 10.1016/j.cognition.2007.11.004

Greene, J.D. (2007). Why are VMPFC patients more utilitarian?: A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*. *11*(8), 322-323.

Ipeirotis, P. G. (2010b). Demographics of Mechanical Turk (Tech. Rep. No. CeDER-10-01). New York: New York University. Retrieved from http://hdl.handle.net/2451/29585. March 2015.

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

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making* *21*(37). 37-44. doi: 10.1177/0272989X0102100105

Mason, W. & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavioral Research, 44*(1), 1-23. doi: 10.3758/s13428-011-0124-6

Paxton, J. M., Ungar, L., & Greene, J. D. (2011). Reflection and reasoning in moral judgment. *Cognitive Science.* 1-15. doi: 10.1111/j.1551-6709.2001.01210.x

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3-22.

Stanovich, K. E. (2013). Why humans are (sometimes) less rational than other animals: Cognitive complexity and the axioms of rational choice. *Thinking & Reasoning, 19*(1), 1-26. doi: 10.1080/13546783.2012.713178

Stanovich, K. E. (2011). Rationality and the reflective mind. New York: Oxford University Press.

Stanovich, K. E. (2009). What intelligence tests miss: The psychology of rational thought. New Haven, CT: Yale University Press.

Stanovich, K. E., & West, R. F. (under contract) *The rationality quotient (RQ): Towards a test of rational thinking*. Cambridge, MA: MIT Press.

Stanovich, K. E., & West, R. F. (2008a). On the failure of intelligence to predict myside bias and one-sided bias. *Thinking & Reasoning*, 14, 129–167. doi: 10.1080/13546780701679764

Stanovich, K. E., & West, R. F. (2008b). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, 94, 672–695. doi: 10.1037/0022-3514.94.4.672

Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*, 13, 225–247. doi: 10.1080/13546780600780796

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, 89, 342−357. doi: 10.1037//0022-0663.89.2.342

Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. *Current Directions is Psychological Science*, *22*(4) 259-264. doi: 10.1177/0963721413480174

Toplak, M. E., & Stanovich, K. E. (2003). Associations between myside bias on an informal reasoning task and amount of post-secondary education. Applied Cognitive Psychology, 17, 851–860. doi: 10.1002/acp.915

Ubel, P. A. & Loewenstein, G. (1996). Public perceptions of the importance of prognosis in allocating transplantable livers to children. Medical Decision Making, 16, 234-241. doi: 10.1177/0272989x9601600307

Ubel, P. A. & Loewenstein, G. (1995). The efficacy and equity of retransplantation: an experimental survey of public attitudes. Health Policy, 34, 145-151. doi: 10.1016/0168-8510(95)00714-4

Wilson, T. D. (2002). *Strangers to ourselves*. Cambridge, MA: Harvard University Press.

Footnotes

[1] Recent research has suggested that use of the Mturk is appropriate and in many regards beneficial for psychological survey research. For example, it has been found by a number of researchers that Mturk samples are generally older than university participant pool samples (*M*age = 32), more internally motivated to take the survey seriously, and are oftentimes more attentive and diligent when completing surveys (Buhrmester, Kwang, & Gosling, 2011; Ipeirotis, 2010; Mason & Suri, 2012)

[2] During the data collection process it was brought to the attention of the researcher than an error was present in one of the non-myside MoAs. Specifically, the third and final non-myside MoA concerning the allocation of surgical procedures contained a typographical error than made the groups receiving the procedures unable to be differentiated. Therefore, any subsequent results that reference non-myside MoAs do not reference the erroneous item as it was dropped. Further, any categorizations of participants based on their responses to the MoA do not reference the surgical procedure MoA.

Table 1.

Means and Standard Deviations from the Time by Condition Mixed ANOVA

	N	*M*	*SD*
Myside MoA			
Condition 1 (US80%)	153	69.29	19.27
Condition 2 (GER80%)	147	65.07	23.32
Non-myside MoA			
Condition 1 (US80%)	153	71.58	20.25
Condition 2 (GER80%)	147	70.75	20.18

Table 2.

Intercorrlations between Measures of Allocation, CRT, NFC, AOT, and Numeracy

	MoA1	MoA2	MoA3	CRT	NFC	AOT	Numeracy
MoA1 (80% Group)	--	.647**	.566**	.161**	.064	.235**	.139*
MoA2 (80% Group)		--	.749**	.158**	.011	.187**	.226**
MoA3 (40% Group)			--	.092	-.021	.112	.178**
CRT				--	.240**	.407**	.490**
NFC					--	.413**	.497**
AOT						--	.230**
Numeracy							--

Note : MoA1 = MoA: Myside - Livers; MoA2 = MoA: Non-myside - Kidneys; MoA3 = MoA: Non-myside - Heart Valves

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Table 3.

Results of the Hierarchical Regressions Demonstrating the Moderating Influence of Actively Open-minded Thinking on Myside Bias

	Between-subjects Myside Bias					
Predictor Variables	*B*	SEB	β	sr^2	R^2	ΔR^2
Step 1					.055**	
Actively Open-minded Thinking	.296	.071	.235	.055		
Step 2					.067**	.012
Actively Open-minded Thinking	.301	.071	.240	.057		
Myside Condition	2.318	1.2	.108	.012		
Step 3					.098**	.031**
Actively Open-minded Thinking	.300	.069	.238	.057		
Myside Condition	31.677	9.306	1.481	.035		
Interaction	-.221	.069	-1.384	.031		

	Within-subjects Myside Bias					
Predictor Variables	*B*	SEB	β	sr^2	R^2	ΔR^2
Step 1					.005	
Actively Open-minded Thinking	-.075	.059	-.073	.053		
Step 2					.015	.010
Actively Open-minded Thinking	-.079	.059	-.077	.059		
Myside Condition	-1.750	1.009	-.100	.010		
Step 3					.043*	.027*
Actively Open-minded Thinking	-.078	.059	-.076	.056		
Myside Condition	-24.359	7.844	-1.392	.031		
Interaction	.170	.059	1.302	.027		

* $p < .05$

** $p < .001$

Table 4.

Points of Intersection and Regions of Significance of AOT Moderation Analyses

		Region of Significance	
AOT Moderation Analysis	Point of Intersection	Lower Limit	Upper Limit
Between-subjects Effect	143.33	130.00	187.95
Within-subjects Effect	143.29	128.39	211.53

Table 5.

Means and Standard Deviations from the Time by Condition Mixed ANOVAs for Utilitarians and Absolutists

	N	*M*	*SD*
Utilitarians			
Myside MoA			
Condition 1 (US80%)	59	73.25	18.06
Condition 2 (GER80%)	57	63.60	25.72
Non-myside MoA			
Condition 1 (US80%)	59	74.44	19.15
Condition 2 (GER80%)	57	76.58	18.30
Absolutists			
Myside MoA			
Condition 1 (US80%)	44	64.09	17.16
Condition 2 (GER80%)	46	65.43	22.65
Non-myside MoA			
Condition 1 (US80%)	44	63.86	19.26
Condition 2 (GER80%)	46	64.89	19.02

Table 6.

Results of the Hierarchical Regressions Demonstrating the Moderating Influence of Actively Open-minded Thinking on Myside Bias for Utilitarians

	Between-subjects Myside Bias					
Predictor Variables	*B*	SEB	β	sr^2	R^2	ΔR^2
Step 1					.095*	
Actively Open-minded Thinking	.384	.111	.308	.095		
Step 2					.154**	.059*
Actively Open-minded Thinking	.411	.108	.330	.108		
Myside Condition	5.491	1.954	.244	.059		
Step 3					.237**	.083*
Actively Open-minded Thinking	.409	.103	.328	.106		
Myside Condition	53.072	13.763	2.36	.101		
Interaction	-.361	.103	-2.135	.083		

	Within-subjects Myside Bias					
Predictor Variables	*B*	SEB	β	sr^2	R^2	ΔR^2
Step 1					.004	
Actively Open-minded Thinking	-.074	.111	-.063	.004		
Step 2					.084*	.080**
Actively Open-minded Thinking	-.104	.107	-.088	.008		
Myside Condition	1.93	1.93	-.284	.080		
Step 3					.179**	.095**
Actively Open-minded Thinking	-.101	.102	-.086	.007		
Myside Condition	-54.306	13.553	-2.544	.118		
Interaction	.366	.102	2.281	.094		

* $p < .05$

** $p < .001$

Table 7.

Points of Intersection and Regions of Significance of AOT Moderation Analyses for Utilitarians

		Region of Significance	
AOT Moderation Analysis	Point of Intersection	Lower Limit	Upper Limit
Between-subjects Effect	147.01	134.11	190.91
Within-subjects Effect	148.38	135.66	191.22

Figure 1.

Plotted Cell Means from the two-way ANOVA Examining Myside Bias

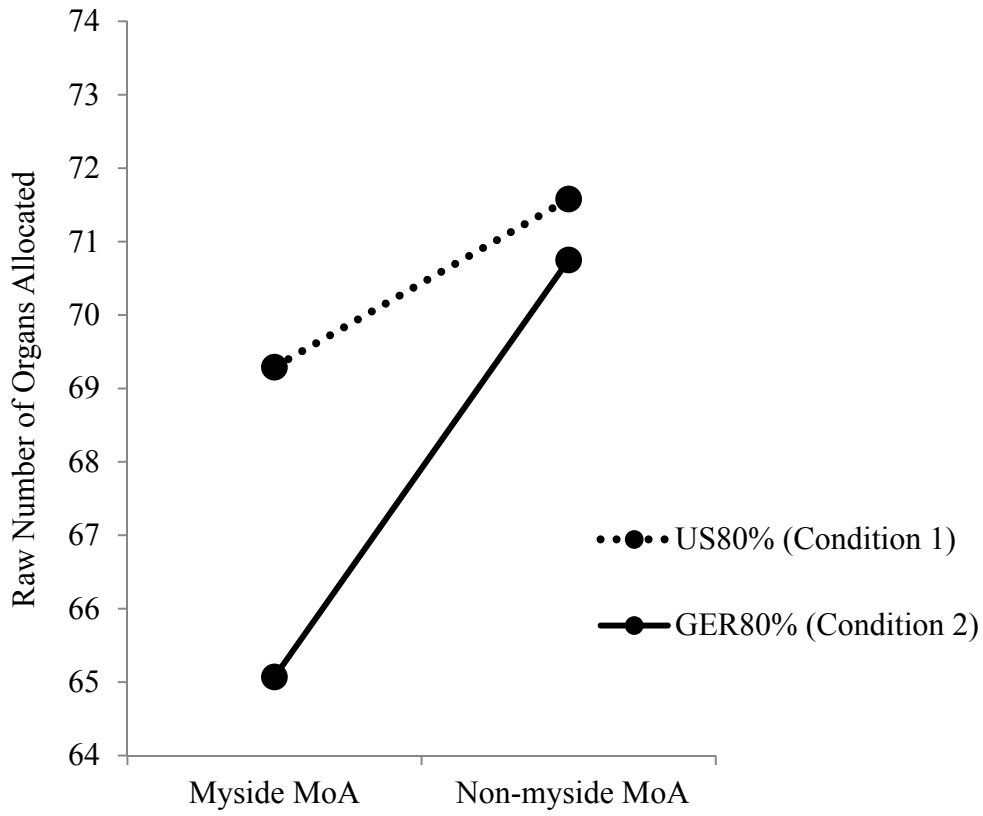


Figure 2.

The Moderating Influence of AOT on Between-subjects Myside Bias

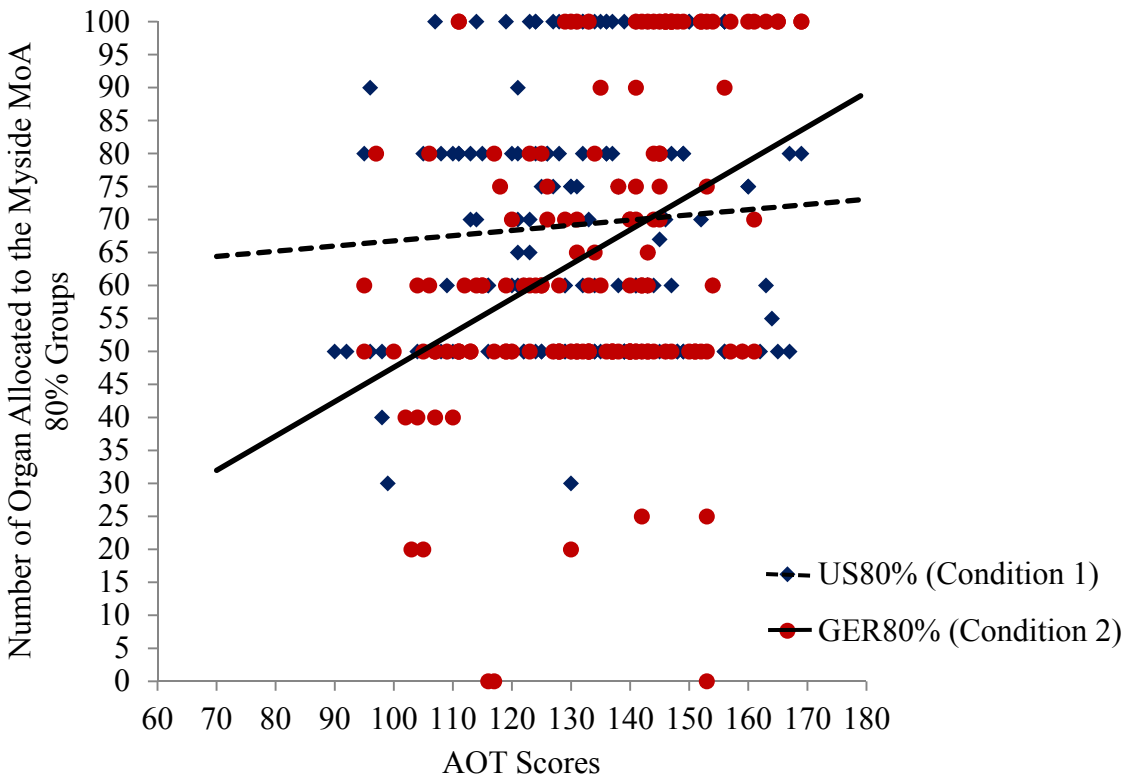


Figure 3.

The Moderating Influence of AOT on Within-subjects Myside Bias

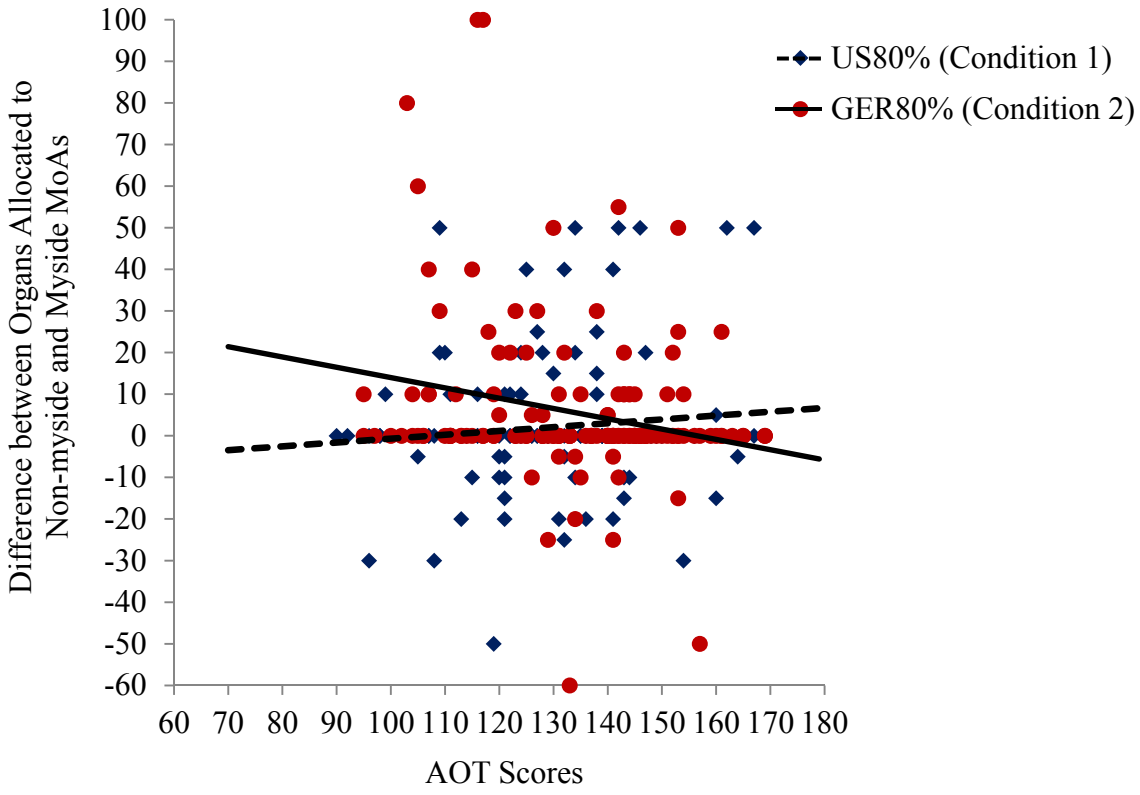


Figure 4.

Plotted Cell Means from the two-way ANOVA Examining Myside Bias for Utilitarians

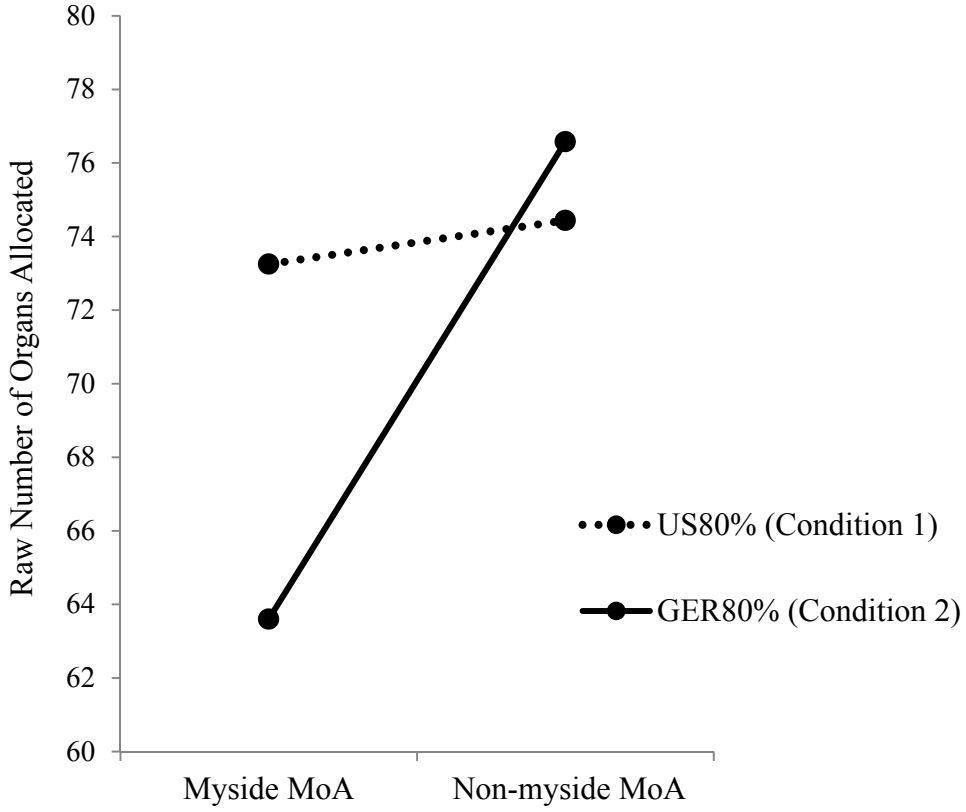


Figure 5.

The Moderating Influence of AOT on Between-subjects Myside Bias for Utilitarians

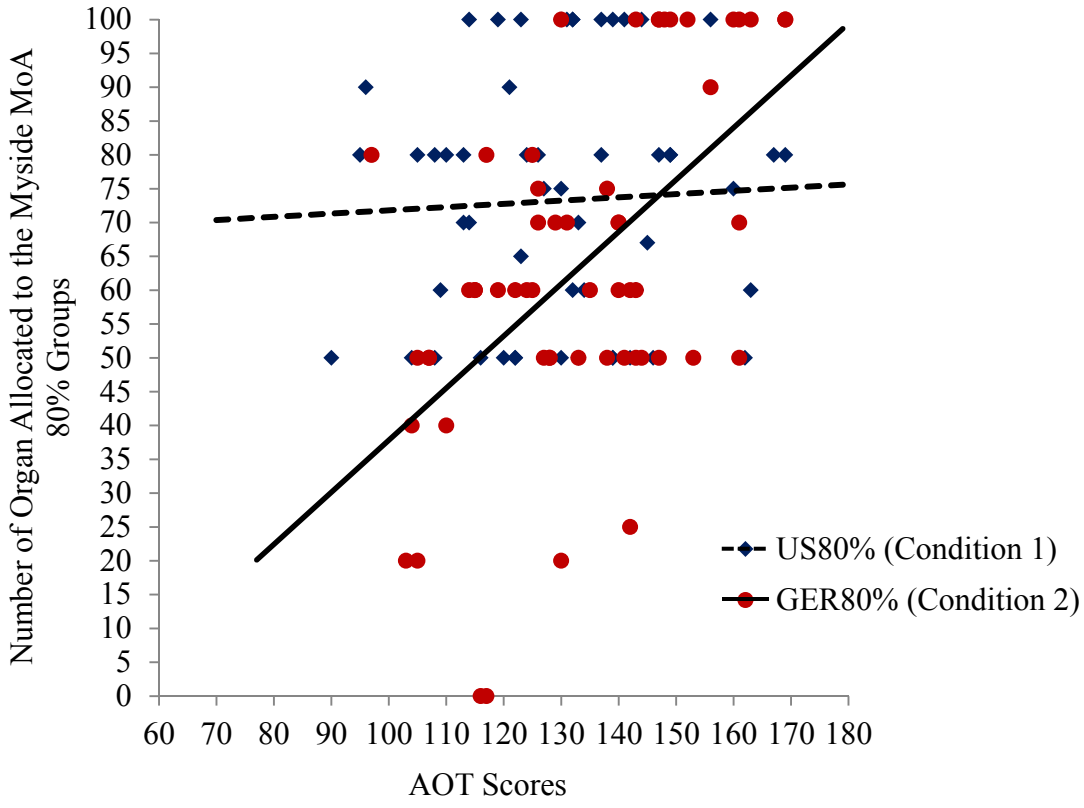
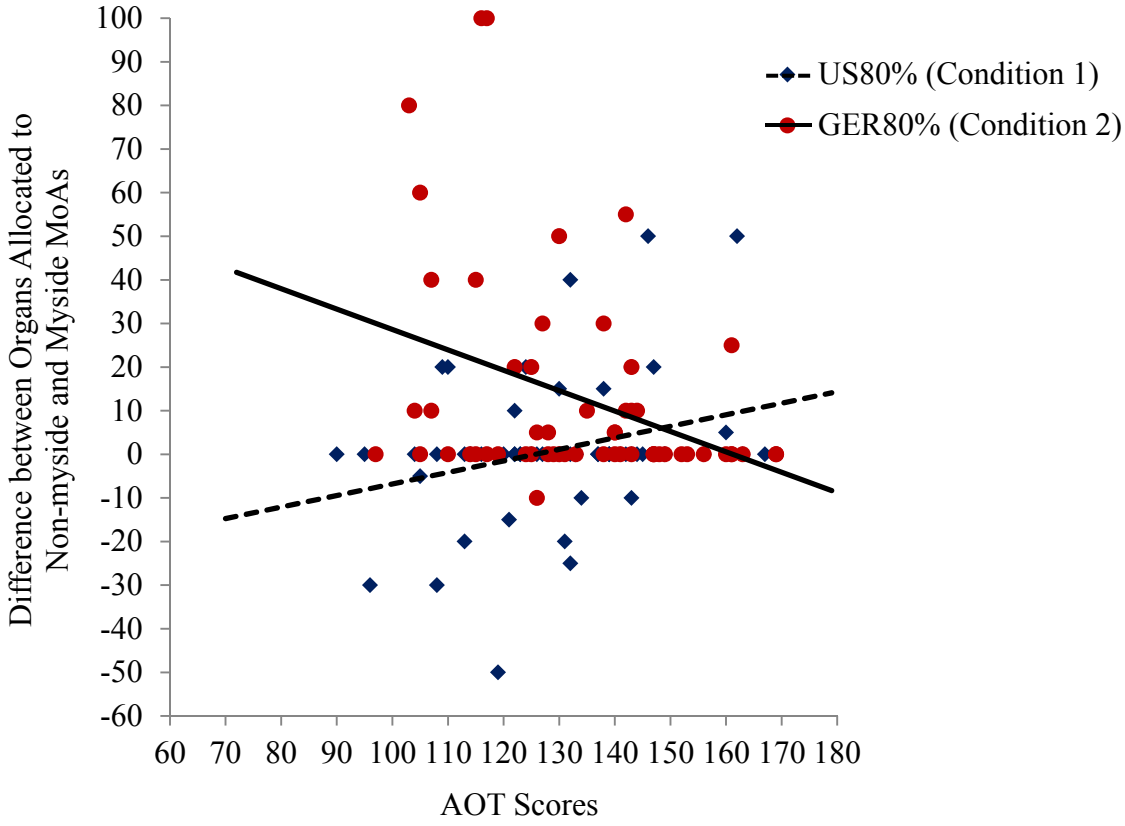


Figure 6.

The Moderating Influence of AOT on Between-subjects Myside Bias for Utilitarians



Appendix A

Measure of Allocation: Myside - Livers

We are going to ask you a question about a health issue. There is no "right" or "wrong" answer to this question.

Imagine that 100 children in the United States and 100 children in Germany have a very rare liver disease and will die if they do not receive a liver transplant within one year. Unfortunately, at this time no livers are available in either country for children with this very rare liver disease. However, 100 of the needed type of livers have become temporarily available for use in these transplants from a source outside of these two countries.

Because different surgical procedures are used in each country, transplant patients have survival rates of 80% (50% in condition 2) in the United States and 50% (80% in condition 2) in Germany:

United States survival rate:	80% (50% in condition 2)
Germany survival rate:	50% (80% in condition 2)

Assuming that the 100 life-saving livers could be allocated in any way among the 100 children in the United States and the 100 children in Germany, how would you allocate them?

United States	
Germany	
Total	100

Appendix B

Measures of Allocation: Non-myside - Kidneys

We are going to ask you a question about a health issue. There is no "right" or "wrong" answer to this question.

As you may know, there is a shortage of kidneys available for those who need transplants. This problem is especially bad for children. Suppose that 200 children are waiting to receive a kidney transplant. These children have no other health problems. They need to receive these transplants within one year or they will die. In that time period, only 100 usable kidneys will become available. Children who do not receive a transplant will die.

A blood test is available that divides the children into two equally sized groups of 100 children each. Children with blood factor X have a transplant survival rate of 80% and children with blood factor Y have a transplant survival rate of 50%. No other information predicts their outcomes as reliably as the blood test.

Children with blood factor X survival rate:	80%
Children with blood factor Y survival rate:	50%

Assuming that the 100 life-saving kidneys could be allocated in any way among the 200 children, how would you allocate them?

Children with blood factor X:	
Children with blood factor Y:	
Total:	100

Appendix C

Measures of Allocation: Non-myside – Heart Valves

Imagine that 800 children from two foreign countries, A and B, are suffering from a very rare heart condition that can only be treated by the implantation of an unusual type of heart valve. Unfortunately, only a limited number of these special artificial heart valves are currently available. In fact, only 400 of these heart valves will be available in time to save the children's lives.

Imagine also that the lifesaving heart valve surgery has a survival rate of 40% in children from Country A and 25% in children from Country B. There are 400 children who need a heart valve in Country A and 400 children that need a heart valve in Country B .

Children in country A survival rate:	40%
Children in country B survival rate:	25%

Assuming that the 400 life-saving artificial heart valves could be allocated in any way among the 400 children in Country A and 400 children in Country B, how would you allocate them?

Children in Country A:	
Children in Country B:	
Total:	400

Appendix D

Measures of Allocation: Non-myside – Heart Valves

Imagine that 200 children are suffering from a rare form of lung cancer and will die within a year if they do not have their tumors removed. Unfortunately, there are few surgeons with the very specialized skill needed to successfully operate on children with this rare form of lung cancer. In fact, there are only enough surgeons to perform 100 surgical procedures in the year the 200 children have to live.

100 of the children have what is classified as a Type 1 tumor and 100 children have what is classified as a Type 2 tumor. The chances of surviving the operation are dependent on the type of tumors the children have. Type 1 tumor patients have a 30% chance of surviving the operation and Type 2 tumor patients have a 70% chance of surviving the operation.

Children with Type 1 tumor survival rate:	30%
Children with Type 2 tumor survival rate:	70%

Assuming that the available procedures could be allocated in any way among the 100 children with a Type 1 tumor and the 100 children with a Type 2 tumor, how many procedures would you allocate to each group?

Children with a Type 1 tumor:	
Children with a Type 2 tumor:	
Total:	100

Note: This item was dropped due to a typographical error in the survey version of the item.

Appendix E

Measures of Allocation: Non-myside – Heart Valves Follow Up Questions

Do you think the way you allocated the surgical procedures to the children with lung cancer would be likely to save the maximum number of children's lives?

a. Yes
b. Probably yes
c. Probably no
d. No

If you were given the option of ignoring the survival rates for the two groups of children (Type 1 = 30%; Type 2=70%) and instead have the procedures *randomly* distributed among the two groups, would you prefer that option?

[ ] Yes
[ ] No

Note: These items was dropped due to a typographical error in the survey version of the MoA they referred to.

Appendix F

Cognitive Reflection Test

A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?

____5____ cents

If it takes 5 machines 5 minutes to build 5 widgets, how long does it take 100 machines to build 100 widgets?

____5____ minutes

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long does it take for the patch to cover half of the lake?

____47____ days

Appendix G

Need For Cognition Scale

1. I would prefer complex to simple problems.
2. I like tasks that require little thought once I've learned them.
3. The idea of relying on thought to make my way to the top appeals to me.
4. Learning new ways to think doesn't excite me very much.
5. I prefer to think about small, daily projects to long-term ones.
6. I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.
7. I prefer my life to be filled with puzzles that I must solve.
8. I feel relief rather than satisfaction after completing a task that required a lot of work.
9. I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.
10. It's enough for me that something gets the job done; I don't care how or why it works.
11. I like to have the responsibility of handling a situation that requires a lot of thinking.
12. I really enjoy a task that involves coming up with new solutions to problems.
13. I usually end up deliberating about issues even when they do not affect me personally.
14. Thinking is not my idea of fun.
15. I try to anticipate and avoid situations where there is a likely chance I will have to think in depth about something.
16. I only think as hard as I have to.
17. I find satisfaction in deliberating hard and for long hours.
18. The notion of thinking abstractly is appealing to me.

Appendix H

The Marble Task

Marble Color Prediction Game:
Imagine that a bowl contains the following 9 marbles that are randomly mixed up:

3 black marbles
6 white marbles

In Round 1 of a game, a marble is randomly picked from the bowl and its color is noted. The ball is then put back into the bowl, and the 9 marbles are once again randomly mixed up. This process is repeated for a total of 9 rounds.

Imagine that you will win $5 each time you correctly predict the color of the randomly selected marble.

Indicate the color that you would predict for each of the 9 rounds:

Round 1: black or white
Round 2: black or white
Round 3: black or white
Round 4: black or white
Round 5: black or white
Round 6: black or white
Round 7: black or white
Round 8: black or white
Round 9: black or white

[Optimal response would be nine consecutive white marbles choices]

Appendix I

Demographics

1. What is your age? _____

2. What is your sex?
   a. Male
   b. Female

3. Is English your first language?
   a. Yes
   b. No

4. Are you currently a college student?
   a. Yes
   b. No

5. What is the highest level of education you have received?
   a. Less than a high school diploma
   b. High school diploma or GED
   c. Some college
   d. B.A. degree
   e. M.A. degree
   f. Ph.D.
   g. Professional degree (e.g., law, medicine)
   h Not applicable

6. Politically, I would consider myself to be:
   a. Very conservative
   b. Conservative
   c. Slightly conservative
   d. Slightly liberal
   e. Liberal
   f. Very liberal

7. If the 2012 presidential election were held today, I would vote for:
   a. Mitt Romney
   b. Barack Obama
   c. I am not eligible to vote

Appendix J

Numeracy Scale

1. Which of the following numbers represents the biggest risk of getting a disease? 1 in 100, 1 in 1000, 1 in 10?
   a. 1 in 100 b. 1 in 1000 c. 1 in 10
2. Which of the following represents the biggest risk of getting a disease? 1%, 10%, 5%?
   a. 1% b. 10% c. 5%

11. If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 100? [10]

11. If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000? [100]

5. If the chance of getting a disease is 20 out of 100, this would be the same as having a ___% chance of getting the disease. [20%]

6. If Person A's risk of getting a disease is 1% in ten years, and Person B's risk is double that of A's, what is B's risk? [2%]

7. If Person A's chance of getting a disease is 1 in 100 in ten years, and Person B's risk is double that of A, what is B's risk? [2 in 100]

8. In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1,000 people each buy a single ticket from BIG BUCKS? [10]

9. Imagine that we roll a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up even (2, 4, or 6)? [500 times]

10. The chance of getting a viral infection is .0005. Out of 10,000 people, about how many of them are expected to get infected? [5]

11. In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car? [.1%]

Appendix K

Ethical Dilemmas

Trolley Switch:

You are at the wheel of a runaway trolley quickly approaching a fork in the tracks. On the tracks extending to the left is a group of five railway workmen. On the tracks extending to the right is a single railway workman.

If you do nothing the trolley will proceed to the left, causing the deaths of the five workmen. The only way to avoid the deaths of these workmen is to hit a switch on your dashboard that will cause the trolley to proceed to the right, causing the death of the single workman.

Footbridge:

A runaway trolley is heading down the tracks toward five workmen who will be killed if the trolley proceeds on its present course. You are on a footbridge over the tracks, in between the approaching trolley and the five workmen. Next to you on this footbridge is a stranger who happens to be very large.

The only way to save the lives of the five workmen is to push this stranger off the bridge and onto the tracks below where his large body will stop the trolley. The stranger will die if you do this, but the five workmen will be saved.