Essays on Judgment and Decision Making

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:42029732">http://nrs.harvard.edu/urn-3:HUL.InstRepos:42029732</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Essays on Judgment and Decision Making

A dissertation presented

by

Bradley R. DeWees

to

The Committee on Higher Degrees in Public Policy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Public Policy

Harvard University

Cambridge, Massachusetts

May 2019
Abstract

Judgment and choice under uncertainty does not occur in a vacuum—instead, it occurs amidst a rich web of social and contextual cues that profoundly shape how individuals navigate their worlds. Each of the essays in this dissertation, while primarily intended as an independent research project, examines how social and contextual factors shape fundamental judgments and choices. The first examines how accountability pressures shape decisions characterized by ambiguity, when the probabilities associated with decision options are vague or unknown. While prior work has shown that accountability amplifies ambiguity aversion, this essay, based on the flexible contingency model, predicts that accountability amplifies ambiguity aversion only under certain conditions. It finds in four experiments that, while accountability amplifies ambiguity aversion when decision makers lack knowledge about the decision domain, accountability actually attenuates ambiguity aversion when decision makers are highly knowledgeable. The second essay examines the influence of observation on decisions involving trade-offs between the values of equality and efficiency. In an extension of the value pluralism model, it finds that observation increases equal (yet inefficient) allocations, even with real financial stakes. Further, it finds that choosing equality is, on average, a politically savvy response to observation—observers trust equal allocators more than efficient allocators. The third essay examines how task flow affects judgments of the ideas and contributions of others. Results from seven experiments with both lay and expert individuals show that committing to one’s own point of view before evaluating a peer’s judgment or choice (relative to the reverse order), leads evaluators to
derogate peers’ ideas and make negative inferences about them. The results reveal a social cost to independent judgments within groups.
# Table of Contents

Title Page ........................................................................................................................................ i  
Abstract.......................................................................................................................................... iii  
Table of Contents .......................................................................................................................... v  
Acknowledgements ....................................................................................................................... vii  
Dedication ......................................................................................................................................... x  

## Introduction ................................................................................................................................. 1  
Results through the Lens of Leadership ......................................................................................... 2  
Dissertation Overview ..................................................................................................................... 3  
References......................................................................................................................................... 7  

Essay 1: Portrait of the Politically-Savvy Decision Maker: Ambiguity Aversion Reverses when Ambiguity Seeking is Justifiable ................................................................. 9  
Introduction ........................................................................................................................................ 9  
Study 1 ............................................................................................................................................. 17  
Study 2 ............................................................................................................................................. 24  
Study 3 ............................................................................................................................................. 26  
Study 4 ............................................................................................................................................. 33  
Discussion ......................................................................................................................................... 34  
References......................................................................................................................................... 39  

Introduction ........................................................................................................................................ 44  
Study 1 ............................................................................................................................................. 50  
Study 2 ............................................................................................................................................. 58  
Study 3 ............................................................................................................................................. 65  
Discussion ......................................................................................................................................... 78  
References......................................................................................................................................... 84  

Essay 3: ‘I Was First and I Was Right’: The Effects of Task Order on Evaluations of Peer Judgments ................................................................................................................................. 92  
Introduction ........................................................................................................................................ 92  
Study 1 ............................................................................................................................................. 99
| Studies 2A & 2B | 103 |
| Study 3 | 108 |
| Study 4 | 113 |
| Studies 5A & 5B | 118 |
| Discussion | 125 |
| References | 131 |

**Appendix: Supplementary Material**

| Essay 1 Supplementary Material | 135 |
| Essay 2 Supplementary material | 153 |
| Essay 3 Supplementary material | 161 |
Acknowledgements

At the beginning of this journey I would have said that praise for collaboration in research was mostly feigned humility. Surely no self-respecting academic actually believes their success is due to “standing on the shoulders of giants.” A few short months of confronting the daunting challenge of producing original, valuable research, however, quickly disabused me of that notion. To the extent that this work is valuable, it is due to the supportive and brilliant network of people with whom I have had the honor of working. While I cannot thank everyone in this network, what follows is my best attempt to thank those who have had an especially important impact on me or the work in this dissertation.

First, I owe a thanks to those mentors and sponsors whose belief that I was capable of pursuing a dissertation provided the critical dose of confidence necessary to undertake and complete the journey. Among these mentors, many of whom are in the Air Force, are: Colonel Cheryl Kearney, Colonel Bridget Gigliotti, Rear Admiral (ret.) Peg Klein, Brigadier General Andy Armacost, Dr. Paul Bolt, Ambassador Roger Harrison, Dr. Dave Sacko, Dr. Damon Coletta, Lt General (ret.) Erv Rokke, Lt General (ret.) Chris Miller, Dr. Steve Pierce, Dr. Stephen Wright, Colonel (ret.) Harmon “Saint” Lewis, Colonel (ret.) Dale Condit, Lt Colonel Joe Foster, and Lt Colonel Ryan Guiberson. Quite literally in some cases, I would not have been able to undertake the journey without these individuals’ support. Also, I owe a special thanks to Dr. Mike Nelson, whose early encouragement and regular support was always a bright spot over the past three years.

Once the journey began, the support of peers and advisors made its completion feasible. When I look back on this experience, I think what I will miss the most is the vigorous intellectual back-and-forth and friendly banter that defines life in Lerner and Minson labs. I greatly benefited
from the feedback, advice, help, and moral support of each of the following wonderful individuals: Joowon Kim, Chelsea Zabel, Seunghee Han, Dilhan Perera, Christina Li, Roberto de Oliveira, Lieutenant Commander Tommy Shannon, Major Joshua Stinson, Jen Hegarty, Shan Shan, Klara “the boss” Kabadian, Ke Wang, Molly Moore, Ayse Yemiscigil, Zoe Rahwan, Hanne Collins, Major Chris Umphres, Alki Aliopoulou, Ariella Kristal, Martha Jeong, Hayley Blunden, Julian Zlatev, Jenn Logg, and Mike Yeomans.

One member of these labs deserves a special thank-you. Charlie Dorison’s more-than-able mind, tireless work ethic, and okay (yes, just okay!) sense of humor made the past three years both fun and productive. He was my primary collaborator on Essay 2 in this dissertation. The strengths in that project are due to him, the shortcomings due to me.

I have heard it said that an advisor and graduate student relationship is like a marriage—it makes all the difference, and, for better or worse, you’re stuck with the one you choose. That analogy is imperfect given that I have two advisors who get along well with each other, but it could not be more right about how important advisors are. They really do make all the difference. Jenn Lerner and Julia Minson, in what was effectively a volunteer capacity, sacrificed their schedules for (at least) weekly meetings in order to meet the three-year timeline, walked me through everything from the basics to the finer points of research, and did their best to teach me what they know. What impressed me most of all was their patience. They took my many novice errors in stride, and managed to make them learning opportunities without making them put-downs. For this and everything else they have done, I hold them as exemplars of how one ought to conduct oneself—as a scientist, a professional, and as a decent person.

Last, I owe a thanks to the two people who were by my side the most the past three years: Kate and Jackson (technically, Kate was there more than Jackson since he didn’t show up until
the halfway mark). I owe them thanks for patiently listening to the half-baked ideas that would eventually come to form in this dissertation. More importantly, I owe them for always helping me keep things in perspective. Put simply, I could not have done it without them.
For Kate and Jackson,

my strength and joy.

Thank you for your steadfast love.
Introduction

“O me, the word ‘choose!’”

– Shakespeare, The Merchant of Venice*

The empirical record of human judgment and choice often paints a bleak picture. In judgment, humans are overconfident (Lichtenstein, Fischhoff, & Phillips, 1982; Moore & Healy, 2008; Moore, Tenney, & Haran, 2015), easily swayed by irrelevant information (Kahneman & Tversky, 1974), and slow to update their beliefs (Edwards, 1962), if they update them at all (Griffin & Tversky, 1992). In choice, humans can be irrational in their consistency (e.g., sunk cost bias; Arkes & Blumer, 1985) as well as in their inconsistency (e.g., violations of Savage’s independence axiom; Slovic & Tversky, 1974). A fair upshot of the field of judgment and decision making could be that humans are, at best, sub-optimal (Simon, 1956), but often biased, from their initial processing of information (Kahneman, 2003) to their ultimate choices (Kahneman & Tversky, 1979). “O me,” indeed.

Despite this empirical record, however, the picture is not all dark. At least two sources of hope emerge. First, humans are corrigible (Laland, 2018). As someone committed to learning in general and improving decision making in particular, my own self-vindication biases me toward that claim. Yet, the claim is empirically true (e.g., Hogarth, 2001; Payne, Bettman, & Johnson, 1993), even in settings where humans are typically prone to bias (Morewedge et al., 2015). As just one example, one-shot training sessions that provide personalized feedback can help people learn to avoid common biases for up to at least two months after initial training (Morewedge et al., 2015).

* Portia is both lamenting her inability to choose a suitor of her own as well as the poor choices of those who do come calling. See Act I, Scene II.
A second reason for hope is that humans, even if biased individually, can be reasonable in the aggregate through the decision making tools and procedures they adopt. They can reduce the error of group judgments, for example, by embracing the error in independent individual judgments (Einhorn, 1986; Minson, Mueller, & Larrick, 2017). They can deliberately alter contextual cues to curb counter-productive impulses (John, Donnelly, & Roberto, 2017). They can, through accountability relationships, tilt the drive to “protect and enhance one’s social image” (Tetlock, 1991, p. 455) toward generating more effortful cognition in their decision making (Lerner & Tetlock, 1999). While lone individuals may have many shortcomings as decision makers, they can address their shortcomings in their capacity as decision architects—they can build organizational processes, cues, and incentives that foster good decision making.

Results through the Lens of Leadership

Leaders play an outsized role in creating good decision environments. They are, metaphorically, the chief decision architects for their organizations (Beshears & Gino, 2015). In their capacity to inspire subordinates and to hold them accountable, leaders are key to motivating the often extremely hard work of *not* relying on one's intuition when making a decision, and instead laboring through a structured process. More broadly, leaders hold power over perhaps the single most important factor underlying a good decision making environment, whether the organization rewards decision makers for adherence to a rational process, as normative decision theory suggests should be the case (Keeney & Raiffa, 1983; Raiffa, 1968), or whether the organization rewards decision makers for fortunate outcomes.

For judgment and decision making theorists interested in leadership, the metaphor of leader as decision architect raises important questions: What levers do leaders have to systematically shape judgment and choice? What systematic effects do those levers have? And
what are the key moderators of those effects? On one level, this dissertation can be read as a contribution to these questions—each of its essays suggests mechanisms through which leaders can shape judgment and choice. Essay 1 suggests that leaders, who likely have some control over the accountability relationships within their organizations, can use the interaction between accountability and decision maker subjective knowledge to make their organizations more or less open to ambiguity, as the situation demands. Essay 2 suggests that leaders aiming to increase efficiency within their organizations should insulate their decision makers from social pressure during key trade-off decisions. Essay 3, in a focus on judgment rather than choice, suggests that leaders can structure task flows to increase group judgment accuracy without creating negative social consequences.

At a more abstract level, the empirical approach of the three essays can be read as its own contribution to leadership. The essays’ approach represents an ideal epistemology for leaders—adherence to the principles of causal inference that are embedded in the scientific method constitutes the best means of by which leaders can distinguish knowledge from opinion. In an uncertain world, leaders should both construct decision making environments that foster unbiased judgment and choice, and insist on rigorous empirical approaches to determine what they and their organizations know.

**Dissertation Overview**

While this dissertation can inform the practice of leadership, its primary function is to contribute to knowledge within the field of judgment and decision making (JDM). In pursuit of that goal, the dissertation is organized into three independent essays that each contribute to a distinct theoretical debate within JDM:
Essay 1: Portrait of the politically-savvy decision maker: Ambiguity aversion reverses when ambiguity seeking is justifiable, with Jennifer S. Lerner

Essay 2: When waste pays: Equality versus efficiency in social context, with Charles A. Dorison, Zoe Rahwan, Chris Robichaud, and Jennifer S. Lerner

Essay 3: “I was first, and I was right”: The effects of task order on evaluations of peer judgments, with Julia A. Minson

The first essay begins with the recognition that important human decisions are often characterized by ambiguity—that is, the probabilities associated with their occurrence are vague or unknown. When given a choice between an option with specific odds and an option with ambiguous odds, humans tend to gravitate toward specific odds. This tendency is known as ambiguity aversion (Ellsberg, 1961). Numerous studies have found that accountability—the expectation of having to justify one’s choices (Lerner & Tetlock, 1991)—amplifies ambiguity aversion. The literature casts decision makers as defensive, selecting the least ambiguous options because they deem that option the easiest to explain. Based on the flexible contingency model, however, Essay 1 predicts that accountability amplifies ambiguity aversion only under certain social/structural conditions, and that accountability will in fact attenuate ambiguity aversion under different conditions. In four incentive-compatible experiments (N = 2,362), examining both accountable individuals and the supervisors to whom those individuals are accountable, the results of Essay 1 reveal that accountability amplifies ambiguity aversion when decision makers lack knowledge about the decision domain, but attenuate it when decision makers are highly knowledgeable. Importantly, accountability triggers these patterns because it allows decision makers to signal that they are standing by their prior predictions. Taken together, the results revise prior conclusions about human responses to ambiguous odds, elucidate underlying
mechanisms, and present a less defensive, more socially/politically savvy portrait of human
decision makers than prior studies have portrayed.

The second essay also examines how social context affects choice. Specifically, it studies
how observation—as opposed to anonymity—affects trade-off decisions between the values of
equality and efficiency. Resource allocations often entail a tension between the values of equality
and efficiency. In a world of scarce resources, allocators often cannot treat everyone the same
while they also use resources efficiently. As a complement to research examining decision
features and individual differences, the three experiments (N = 1,095) in Essay 2 extend the
value pluralism model to show how observation systematically affects equality-efficiency trade-
offs. Study 1 predicts and finds that observation increases equal (yet inefficient) allocation
decisions, even with real financial stakes. Consistent with predictions from the value pluralism
model, changes to the ordinal ranking of values accounts for this effect: observation shifts one’s
focus from lives saved to equality. Study 2 conceptually replicates the effect of observation on
equal (yet inefficient) allocation decisions, varying how observation is operationalized and
introducing a decision domain involving life and death choices. Study 3 focuses on observers,
and documents three key results. First, with real financial stakes, observers trust equal allocators
more than efficient allocators. Second, while observers who favor equality financially punished
allocators who prioritized efficiency, observers who favored efficiency did not. Finally, this
financial punishment by equality-preferring observers derives from different underlying social
perceptions of warmth. Taken together, the experiments in this essay extend understanding of
how the social context shapes value-laden judgment and choices.

The third essay examines how the contextual factor of task flow affects judgments of the
ideas and contributions of others. According to prior research, the generation of independent
judgments prior to group interaction improves group judgment accuracy—a near consensus recommendation to groups seeking to improve the accuracy of their judgments, therefore, is to first make independent judgments before forming an overall group opinion. The third essay, however, shows that this recommendation comes with a social cost. The results from seven experiments reveal that committing to one’s own point of view before evaluating a peer’s judgment or choice (relative to the reverse order), leads evaluators to derogate peers’ ideas and make negative inferences about them. Study 1 of this essay demonstrates the effect. Studies 2A and 2B show that the effect is both moderated and mediated by disagreement. Study 3 shows that disagreement is interpreted in an egocentric manner: disagreement implies that other’s judgments—rather than one’s own—are less accurate. Studies 4, 5A, and 5B investigate the effect in two complex decision-making scenarios. Results from those studies reveal that, after committing to a decision themselves, both lay participants and national security experts derogate peer decisions and the quality of the reasoning behind them.

The subsequent chapters of this dissertation include, respectively, the three essays listed above, followed by an appendix containing the supplementary materials for each essay.
References


Essay 1

**Portrait of the politically-savvy decision maker**

_Ambiguity aversion reverses when ambiguity seeking is justifiable_¹

with Jennifer S. Lerner

Human choice is often rife with uncertainty that defies precise quantification (Knight, 1921). Specifically, choice options are often ambiguous—their likelihood is vague or unknown (Ellsberg, 1961). One would say, for example, that a suitor seems promising, not that the suitor holds precisely a 30% chance of becoming one’s spouse. Similarly, a job may offer the possibility of advancement, yet no precise quantification of the likelihood of making manager may be possible. Navigating choice—whether of suitors, jobs, or other matters of consequence—often means navigating ambiguous odds.

Inspired by Ellsberg (1961) and Knight (1921), numerous studies have found that human decision makers display ambiguity aversion—that is, they prefer a known probability of winning over an unknown probability of winning (see, for example: Akay, Martinsson, Medhin, & Trautmann, 2012; Curley & Yates, 1989; Dimmock, Kouwenberg, & Wakker, 2012; Einhorn & Hogarth, 1986; Slovic & Tversky, 1974). Ambiguity-averse choices do not constitute a bias, per se, but can violate Savage’s axiom of independence (Slovic & Tversky, 1974), which, in dynamic settings, can lead to irrational inconsistencies and preference reversals (Al-Naar & Weinstein, 2009).² Importantly, ambiguity aversion represents a fascinating regularity of human choice in a world replete with vague or unknown odds.

---

¹ Funding from the National Science Foundation (PI: Lerner, Grant SES-1559511) and Harvard University’s Center for Public Leadership (PI: DeWees) partially supported this research.

² Ambiguity aversion may also be costly for individuals. Individuals who show ambiguity aversion, for example, are less likely to invest in equities, a practice that can limit their investment returns over the long run (Dimmock, Koewenberg, Mitchell, & Peijnenburg, 2016). When ambiguity-averse individuals do invest in equities, they are likely to under-diversify, preferring home-country to foreign equities and own-company equity to all others (Dimmock et al., 2016).
Before reviewing psychological explanations for ambiguity aversion, it is useful to move up several levels of analysis to assumptions about human nature. Psychological explanations for empirical regularities such as ambiguity aversion ultimately invoke functionalist assumptions—often personified through metaphor—about the objectives humans seek to achieve with their judgments and choices (Lakatos, 1970, Tetlock, 1991). Beyond lending intuitive coherence to a research program’s assumptions, such metaphors guide theoretical inquiry by pointing toward relevant research questions.

**Metaphors for Human Decision Making**

Representing hundreds, if not thousands, of path-breaking studies, the *intuitive scientist* research program has been the dominant underlying metaphor for at least five decades of psychology research. Describing people as aiming to understand and predict reality (Kelley, 1971; Ross, 1977), researchers implicitly or explicitly working within the intuitive scientist program might explain ambiguity aversion as resulting from a desire to make precise predictions. The *intuitive economist* research program, by contrast, describes people as aiming to maximize subjective utility (Edwards, 1962). It might explain ambiguity aversion as a conservative strategy for maximizing payoffs in situations characterized by ambiguity.3

Although distinct in many ways, both of the foregoing metaphors implicitly depict individuals in a Thoreauvian state of isolation seeking intrapsychic goals (i.e., understanding the external environment or maximizing subjective utility; Tetlock, 2002). While useful for illuminating key aspects of human nature, these particular metaphors focus less on the social/political nature of human decision making (Cooley, 1902; Lieberman, 2013; Mead, 1934;.....
Tetlock, 2002; Tetlock & Lerner, 1993; Lerner & Tetlock, 2003). Choices that occur amidst a web of relationships may lead decision makers to anticipate the reactions of key constituencies and factor in how others will perceive their choices (Vygotsky, 1980; Tetlock, 1992; Lerner & Tetlock, 1999).

Recognizing the social context surrounding choice is the starting point for what Tetlock (1991; 2002) originally termed the intuitive politician research program. According to this view, accountability—i.e., the expectation of justifying one’s views to a potentially evaluative other (Lerner & Tetlock, 1999)—is a key driver of individual judgment and choice (Adelberg & Batson, 1978; De Dreu & van Knippenberg, 2005; Sedikides, Herbst, Hardin, & Dardis, 2002; Simonson, 1989; Tetlock, Skitka, & Boettger, 1989; for reviews, see: Hall, Frink, & Buckley, 2015; Lerner & Tetlock, 1999; Patil, Vieider, & Tetlock 2014). Within this metaphor, the theoretical focus becomes identifying the motives and strategies involved in coping with accountability pressures, and how well those strategies perform in addressing audience goals. Decision makers may, for example, be defensive and insecure about their choices, or they may instead be confident, flexible, and responsive to cognitive and social contingencies. Individuals may be savvy judges who accurately read their constituency, or they may be inept judges who cling to broken mental models (e.g., Epley, Keysar, van Boven, & Gilovich, 2004).

The Intuitive Politician and Decision Making

Interestingly, although the intuitive politician metaphor is relatively new in judgment and choice research, a clear portrait of the intuitive politician’s behavior has already emerged. Several decades of research portray the intuitive politician as a reflexively defensive (perhaps even spineless) decision maker. For example, rather than undertaking the difficult task of evaluating all features associated with options in a choice set, accountable decision makers tend
to select the option supported by superficially justifiable reasons (Simonson, 1989), even if the reason is merely that the option appears to be a compromise between two other options (Simonson & Nowlis, 2000). Similarly, when given a choice between developing strong arguments to defend their initial opinions and the easy option of merely adjusting their opinions to those of an audience, decision makers adopt the acceptability heuristic—i.e., they reflexively shift their opinions toward the views of the audience to whom they will have to justify themselves (Tetlock, et al., 1989). Accountable decision makers also become prone to using all possible cues in a judgment task, even if they are irrelevant cues that bias a prediction (Siegel-Jacobs & Yates, 1996; Tetlock, Lerner, & Boettger, 1996), for fear of leaving some stone unturned.

Importantly, all studies to date regarding the effects of accountability find that ambiguity aversion becomes even stronger when decision makers expect to have to justify themselves, implying that people try to choose the option that seems easiest to justify (Curley, Yates, & Abrams, 1986; Tetlock & Boettger, 1994; Trautmann, Vieider, & Wakker, 2008; for a review, see: Trautmann & van de Kuilen, 2015). Curley, Yates, and Abrams (1986) were the first to show this, documenting that accountable decision makers were more likely than unaccountable decision makers to prefer options with known odds over options with ambiguous odds, even at equal expected value. Building on this finding, Trautmann, Vieider, and Wakker (2008) showed in a clever series of studies that “fear of negative evalaution,” in particular, is a necessary precondition for ambiguity aversion (p. 225, emphasis added). In their studies, as long as decision makers were insulated from blame for betting on ignorance—a reasonable concern in most experimental paradigms testing ambiguity (Tetlock, 1991)⁴—the decision makers did not

⁴ Most empirical tests withhold information from decision makers in order to create ambiguity; that is, they reveal the odds associated with some options but not others. This creates a state of “comparative ignorance” for the
show ambiguity aversion. In sum, the foregoing results suggest that accountable decision makers react defensively to their evaluating audiences by rigidly choosing options that insulate themselves from negative evaluation. Choosing between options with known or ambiguous odds, they adopt the easier-to-justify options with known odds, even though choice options with ambiguous odds may be more beneficial over the long run.

**The Flexible Contingency Model of Choice under Accountability**

While this portrait of a defensive decision maker must hold true under the conditions employed in prior studies, it may not turn out to be a full and fair representation of accountability’s effects in choice settings overall and in ambiguous choice settings in particular. To provide a fair test of whether decision makers who face accountability are as reflexively defensive as the literature suggests, at least three conditions must be met:

1) Accountable decision makers need to be incentivized for accuracy so they have reason to demonstrate optimal choices;

2) Decision makers need to be accountable to audiences with the authority to tangibly award or withhold award; and

3) Perhaps most importantly, accountable decision makers need to make choices at varying levels of knowledge—not just ignorance—in order to assess whether they flexibly adjust their choices according to domain knowledge. Without this last condition, studies cannot detect whether the effects of accountability differ when decision makers believe they can gain credit for their knowledge.

---

decision maker relative to the experimenter (Fox & Tversky, 1995, p. 585). Trautman et al. (2008) avoided placing decision makers in this position by allowing them to keep their preferences over the outcomes associated with different options as private information. Different outcomes were randomly assigned to options with known or unknown odds; participants then made their choice. While the experimenter could see whether a participant chose options with known or unknown odds, the experimenter could not know whether the choice was due to a preference for the outcome or the odds associated with the outcome.
To hypothesize alternatives to the portrait of decision makers as reflexively defensive, we draw on the flexible contingency model (FCM) of choice under accountability (Lerner & Tetlock, 1999). The FCM assumes that accountability’s effects hinge on interactions among several situational factors, including characteristics of the accounting audience, the timing and focus of accountability, and characteristics of the accountable decision maker. Changes in any of these dimensions can shift, or even reverse, accountability’s effects (for a review, see Lerner & Tetlock, 1999). The FCM suggests that the effects of accountability on choice under ambiguity are more nuanced than the current empirical record suggests.

One critical variable in both the FCM and decision making under ambiguity is an individual’s level of knowledge. The FCM holds that accountability’s effects result from an interaction between individual knowledge and the expectation of justifying one’s views. Awareness of Bayes theorem, for example, is a precondition for accountability to motivate decision makers to incorporate base rates in judgments of probability (Simonson & Nye, 1992). For decision makers facing ambiguity, prior research has shown that subjective knowledge of a given decision domain is a key moderator of ambiguity attitudes (Heath & Tversky, 1991; Fox & Tversky, 1995). Heath and Tversky (1991), for example, found that decision makers showed ambiguity aversion only when they had low subjective knowledge of a domain. When they had high subjective knowledge, decision makers actually sought ambiguity (i.e., they preferred options with vague odds over options with precise odds of equal expected value). Heath and Tversky (1991, pp. 7-8) speculated that decision makers in their studies may have anticipated blame for betting on ambiguous options when they had low subjective knowledge and credit for

---

5 The FCM shares key conceptual similarities with other models of choice under accountability. The feature-based model of Hall, Ferris, Bowen, and Fitzgibbons (2007), for example, holds that accountability’s effects depend crucially on such factors as “accountability source, accountability focus, accountability salience, and accountability intensity” (for a review, see: Hall, Frink, & Buckley, 2015).
doing so when they had high subjective knowledge. In other words, Heath and Tversky (1991) speculated that decision makers were only willing to stick their necks out when it was reasonable to do so (because of high-domain knowledge). While we find this to be a compelling explanation for the role of subjective knowledge as a moderator in decision making under ambiguity, we note that the studies did not examine the potential causal effects of accountability to an important audience—e.g., a supervisor who can assign financial rewards or punishments.

Beginning where Heath and Tversky left off, we therefore combine the nuanced approach of the FCM with research on the role of subjective knowledge in decision making under ambiguity. We reasoned that accountability could both amplify and attenuate ambiguity aversion, depending on a decision maker’s subjective knowledge in the underlying decision domain. Specifically, we reasoned that accountable decision makers would anticipate *evaluative peril* for choosing ambiguous decision options about which they had low subjective knowledge. However, in a significant revision of prior literature, we also reasoned that accountable decision makers would anticipate *evaluative promise*—positive evaluation for judging well—for choosing ambiguous decision options about which they had high subjective knowledge. Thus, the present research aimed to elucidate whether accountability pressures could signal not only peril but also promise when dealing with ambiguous odds.

**Hypotheses**

As noted, prior research consistently found that accountable decision makers were reflexively defensive: They avoided the peril of negative evaluation when faced with ambiguity.

---

6 The studies in Heath and Tversky (1991) operationalized subjective knowledge as the estimated probability in a prediction, with higher probabilities representing higher subjective knowledge. Higher probability indicates closer psychological distance (Waksłak, Trope, Liberman, & Alony, 2006). Decision makers may have inferred credit for betting on options that were psychologically proximal but blame for betting on objects that were psychologically distant.
We predict that decision makers will take a more flexible approach to accountability. Namely, defensiveness should hold only when decision makers have low subjective knowledge in a given domain. When accountable decision makers have high subjective knowledge, they should seek the promise of positive evaluation. We thus predict a crossover interaction between accountability and subjective knowledge in determining ambiguity aversion.

The Present Research

To study meaningful accountability with true financial consequences, we conducted incentive-compatible experiments in pairs, focusing first on decision makers and second on the evaluating audiences themselves. In Study 1, we compared the effects of accountability on incentive-compatible choices in low- and high-subjective knowledge domains. In Study 2, actual supervisors evaluated decision makers’ choices and determined their financial rewards. We used this study to measure whether the decisions made in Study 1 successfully appeased evaluating supervisors. This paired approach, which we continue for Studies 3 and 4, offers a holistic view of the costs and benefits of decision making under real accountability rather than just under anticipated accountability. In addition to examining the direct payoffs resulting from particular choices under ambiguity, we also examine the indirect payoffs resulting from the evaluating audience’s rewards of particular patterns of choices under ambiguity. To our knowledge, these are the first studies to examine systematically both the decisions made by accountable and unaccountable decision makers as well as the responses of actual supervisors who evaluate (and potentially reward) the accountable decision makers.

Open Science & Power Considerations

7 The examination of high-knowledge settings is important for the alignment of research on ambiguity with real-life decision making. High knowledge is likely typical for people operating in their work environments, for instance.
In keeping with best practices for fully reproducible science (Simmons, Nelson, & Simonsohn, 2012), we report all methodological decisions (e.g., determining sample size), manipulations, and measures. We have also posted data and all materials necessary to replicate our studies. We preregistered the hypotheses, methods, and analyses. We sought to maximize power in our studies in a number of ways. In studies 1 and 3, we utilize a repeated-measures design and a within-subjects treatment for subjective knowledge (more details below). Further, we based all sample size calculations on simulations from pilot data. Last, we verified that manipulated variables were successful, and report manipulation checks at the start of results sections.

**Study 1**

**Method**

**Participants**

We sought to enhance the diversity and inclusiveness of our sample through the use of an online subject pool, Amazon’s Mechanical Turk (mTurk), which is typically more diverse than undergraduate research pools (Buhrmester, Kwang, & Gosling, 2011; Casler, Bickel, & Hackett, 2013). To obtain participants in a high-knowledge domain, we recruited mTurk workers who claimed to be NFL football fans (mTurk), offering $1.50 and bonus possibilities. In order to ensure our study was adequately powered, we conducted simulations with pilot data; these simulations suggested the need for a sample of approximately five hundred. Of the 568 participants who began the study (194 female, $M_{age} = 38$, $SD_{age} = 11.89$), one dropped out after receiving initial instructions but before the accountability manipulation. Thirteen participants dropped out after randomization: 10 from the accountable condition and three from the unaccountable condition. To assess attrition in response to treatment conditions (Zhou &
Fishbach, 2016), we conducted a chi-square test. The results approached significance ($p=.052$). However, we note that the marginal difference across conditions (n=7) represented only 1.2% of the sample, an amount unlikely to change the results.

**Design and Procedure**

**Design.** Adapting the classic paradigm from Heath and Tversky (1991), Study 1 took the form of a 2 (accountability: control, accountable) X 2 (subjective knowledge: low, high) mixed factorial design with subjective knowledge as a within-subjects factor. Accountable decision makers anticipated justifying their choices to a real supervisor who had the authority to allocate resources based on her/his evaluations. Without deception, we tested the moderating role of subjective knowledge on the effects of accountability by asking football fans to estimate and then bet on ambiguous probabilities in domains representing relatively high (football) versus low (stock returns) subjective knowledge.

**Procedure.** After giving informed consent, participants stated whether they felt that they had higher knowledge of football or stock price movements. Participants received brief training in probabilistic prediction; for each prediction, we asked them to imagine “making the same prediction 100 times” and stated that their judged probability (which we called “confidence” in the survey materials) should be the number of times out of 100 that they expected to make the right prediction. Participants then made six predictions, three high-knowledge (football) and three low-knowledge (stock return). We asked them, “Which team will win?” (high knowledge) or “Which stock will gain more?” (low knowledge). Because simultaneous display encourages

---

8 Though we targeted NFL fans, 92/568 participants (16.2%) claimed more knowledge about stocks than about the NFL. Participants who claimed relatively higher knowledge in the stock market after signing up for a survey for NFL fans presumably felt knowledgeable in both domains. To compare accountability’s effects across high and low knowledge, we studied only participants who said football was their higher-knowledge domain. As a check on robustness, we report analyses with the entire sample.
comparison between options (Basu & Savani, 2018), we displayed the high-knowledge and low-
knowledge questions together in pairs in order to make the knowledge differences between the
two salient. The order of the pairs was random. We counterbalanced which appeared first, high
knowledge or low knowledge, when reading from left to right.

After making predictions, participants entered the bonus phase of the survey. In this
phase, we offered them a choice between a certain amount, which we based on their judged
probability for a given prediction, or a higher payment tied to the accuracy of their prediction.
This higher payment was $.50 if the prediction was correct, $0 otherwise. As a specific example,
if a participant gave an 80% judged probability for a prediction, we offered her a choice between
$.50 if her prediction was correct ($0 otherwise), or $.40 for certain (i.e., 80% of $.50). We
offered this choice for each of the six predictions, with the specifics of each certain alternative
tied to the decision maker’s judged probability for a specific prediction. Notably, when decision
makers initially offered their judged probabilities, they were unaware that higher judged
probabilities would lead to higher certain alternatives.

**Accountability manipulation.** After participants made their initial predictions, but
before they made their bonus phase choices, we randomly assigned them into accountable and
unaccountable conditions.

Participants in the *accountable* condition read that they would have to justify their
choices: “Your choices—as well as the justification you provide for your choices—will be
evaluated by another mTurk worker (a supervisor) in a separate survey on rewards.” We
explained, truthfully, that the supervisor would allocate a double bonus to either the participant
or another mTurk worker randomly paired with the participant. We emphasized that the
supervisor could use any criterion s/he deemed appropriate to make this decision. We also stated
that the supervisor would know whether a participant had stated that s/he knew more about
football than stocks, or vice versa. In sum, accountability had meaningful consequences:
accountable participants anticipated justifying their choices between their self-judged ambiguous
prospect and the matched, certain alternative to a supervisor with the authority to allocate
resources based on her/his evaluations.

Participants in the unaccountable condition read that their choices were confidential.
They also read that half of them would be randomly selected to receive a double bonus (to match
the incentive structure in the accountable condition, in which half of participants could expect a
double bonus).

Participants in both accountability conditions answered comprehension-check questions
for their respective accountability-treatment pages. The questions were parallel, with the
exception that the role of the supervisor was highlighted in the accountable condition and the
confidentiality of the predictions was highlighted in the unaccountable condition.

Dependent measure. Following the accountability treatment, participants chose between
the ambiguous prospect and the certain alternative. The measure contained explicit reminders of
each of their predictions, their judged probability (confidence) for a given prediction, and the
monetary value of the certain alternative.

Manipulation checks and exploratory individual-difference measures. Participants
completed an accountability manipulation check after making their bonus choices. The check
consisted of a pair of Likert items that asked to what extent participants considered (1) justifying
their choices and (2) how others would view those choices. Participants then completed a writing
task and a series of individual difference scales, inserted for exploratory purposes: numeracy
(i.e., ability to process numerical concepts; Lipkus, Samsa & Rimer, 2001; Peters et al., 2006),
honor orientation (i.e., belief in the importance of remaining true to one’s word, Kertzer, 2017), and self-monitoring (i.e., tendency to regulate social behavior to accommodate social cues; Snyder, 1974). Last, participants answered demographic questions before receiving a survey completion code. We paid bonuses as stated in the survey and did not use any deception.

Results

Preliminary Analyses

Accountability manipulation. The manipulation was successful; accountable participants reported higher levels of agreement with the statements that (a) they considered how they would justify their choices (95% CI [0.37, 0.96], $t(460) = 4.39, p < .001$) and (b) they considered how others would view their decisions (95% CI [0.64, 1.35], $t(460) = 5.47, p < .001$).

Subjective knowledge manipulation. This manipulation also succeeded; participants reported higher confidence in their predictions of football games ($M = 77.86$) than in their predictions of stock market events ($M = 63.08, 95\% \text{ CI} [-15.61, -13.95], z = -34.88, p < .001$).

Order of display manipulation. As expected, the order in which high- and low-knowledge predictions appeared on the screen had no effect on judged probabilities or the likelihood of betting on one’s ambiguous predictions. Given the absence of an effect, we collapsed across ordering conditions for subsequent analyses.

Primary Analyses

We hypothesized that accountable decision makers with high subjective knowledge would be more likely to seek a positive evaluation by betting on their ambiguous predictions and that they would be less likely to bet on ambiguous predictions with low subjective knowledge. We tested this hypothesis by entering accountability, subjective knowledge, and their interaction into a regression model. Given a repeated-measures design and a binary dependent variable, we
used a generalized multi-level model with the binomial distribution as the link function. We assigned choosing the ambiguous option (one’s own judgment) a value of 1 and choosing the certain alternative a value of 0. Higher values, then, signified more openness to ambiguity.

Consistent with our hypothesis, we observed a significant interaction between accountability and knowledge \((\log\text{-}odds \ 95\% \ CI \ [1.95, 4.31], \ z = 5.27, \ p < .001)\). To isolate the locus of the interaction, we tested the simple effect of accountability at both knowledge levels. Consistent with our hypothesis, as compared to unaccountable participants, accountable participants were significantly more likely to bet on ambiguous options at high levels of knowledge \((\log\text{-}odds \ 95\% \ CI \ [1.14, 2.58], \ z = 2.63, \ p = .009)\) and significantly less likely to do so at low levels of knowledge \((\log\text{-}odds \ 95\% \ CI \ [0.38, 0.91], \ z = -2.40, \ p = .016; \) see Figure 1.1).

![Figure 1.1](image_url)

**Figure 1.1.** As predicted, the effect of accountability depended on a decision maker’s subjective knowledge in a given domain. Accountable decision makers were more likely to avoid negative evaluation by choosing the certain option at low knowledge (left) while being more likely to bet on ambiguous options at high knowledge (right).
Interestingly, the choices made by accountable participants in the high-knowledge domain were statistically indistinguishable from indifference between their ambiguous predictions and the certain alternative (95% CI [0.48, 0.56], p = .38). In other words, they showed no ambiguity aversion. Ambiguity aversion appeared in all other conditions: Unaccountable and accountable decision makers with low knowledge all preferred certainty over their ambiguous predictions (p < .001 for all cells). Thus, the tendency to be ambiguity averse held for decision makers except when they had the opportunity to bet on what they knew.

**Payoff from bets.** A final key question was whether the accountability treatment led to reckless bets: Did accountable participants’ greater likelihood of choosing their ambiguous prediction over a certain alternative lead them to walk away with less bonus money than unaccountable participants? To examine this, we compared financial earnings across accountability conditions. Accountable participants were, in fact, no worse off than unaccountable participants in terms of money earned at both high and low levels of knowledge (respective ps > .40).

**Discussion**

Consistent with the flexible contingency model of choice under accountability, decision makers’ reactions to an evaluating audience were moderated by their characteristics. Specifically, accountability’s effects depended on decision makers’ level of subjective

---

9 As a robustness check, we re-ran the model to include those participants who claimed more knowledge about the stock market than the NFL and still observed a significant interaction (95%-CI[1.45, 2.91], z=4.02, p<.001) as well as a significant simple effect of accountability at high knowledge (95%-CI[1.03, 2.14], z=2.17, p=.03). However, the simple effect of accountability at low levels of knowledge drops from significance (95%-CI[0.50, 1.06], z=-1.66, p=.096). Again, these individuals who claimed higher knowledge of the stock market after signing up for a study for NFL fans likely had high knowledge of both.

10 A final set of exploratory analyses examined the extent to which individual differences might moderate key relationships. Because these analyses were explicitly exploratory, we opted to increase the sample size for analyzing individual difference data by collapsing across Studies 1 and 3. The results revealed that none of the measures served as a reliable moderator of the effects. We elaborate on these analyses in our Supplementary Materials.
knowledge. Accountable decision makers were more likely than unaccountable decision makers to favor their ambiguous predictions, potentially seeking a positive evaluation, when they had high levels of subjective knowledge. The reverse was true at low levels of subjective knowledge.

The results also showed that accountable decision makers were able to adapt to audience pressures without sustaining a direct financial cost. Their choices between certain amounts and ambiguous predictions led to similar bonus winnings as the choices of unaccountable participants. In the next study, we examine another dimension of the costs and benefits of choice under ambiguity in an accountable context by testing whether accountable participants sustain an indirect financial cost or benefit from the rewards of their evaluating audience.

**Study 2**

In Study 1, the presence of a supervisory audience in the accountable condition triggered a novel behavioral pattern among decision makers: Those with high knowledge bet on ambiguous options, while those with low knowledge avoided ambiguous options. In Study 2, we pivot from the decision makers themselves and focus instead on how the supervisors who evaluate them react to their decisions. Do accountable decision makers accurately infer what their supervisors will reward?

**Method**

**Participants**

After calculating statistical power based on estimates of effect size, we recruited 361 participants via mTurk (48% female, $M_{age} = 36$). We offered $.50 for a five-minute survey with the possibility of a $1 bonus for carefully reading instructions. Two participants dropped out, leaving us with a final sample of 359.

**Design and Procedure**
To isolate the effects of particular ambiguity choices across high and low knowledge, we defined four patterns for choice behavior among decision makers in Study 1: A decision maker (1) bets on ambiguity at both high and low knowledge, (2) bets on ambiguity at high knowledge while taking the sure thing at low knowledge, (3) chooses the sure thing at high and low knowledge, or (4) takes the sure thing at high knowledge while betting on ambiguity at low knowledge. To test the effect of choosing one pattern over another on rewards allocated by supervisors, we took each possible permutation of the two patterns (e.g., 1-versus-2, 2-versus-1, 1-versus-3, etc.) and presented it to an individual supervisor. We then allowed the supervisor to select the pattern that s/he believed deserved the bonus. In effect, we employed a 6 (pattern combination) X 2 (pattern ordering) between-subjects design. The outcome measure was a supervisor’s binary choice of which pattern to reward.

Supervisors saw the choices of a pair of hypothetically named decision makers and, using any criteria they deemed appropriate, selected a pattern to reward. After their selection, supervisors completed basic demographic questions and then concluded the survey. Within a week of survey completion, we paid bonuses to a randomly selected 10% of supervisors who correctly answered all comprehension questions.

Results

Preliminary Analysis

Order-of-display manipulation. Using a dummy variable for order-of-display, a series of logistic regressions did not find a significant result for any of the six pattern combinations. Given this, we collapsed across ordering for all subsequent analyses.

Primary Analyses
We found that supervisors rewarded the behavior we observed in Study 1. Using a series of binomial tests (tests of the true proportion of supervisors that favored a given strategy over another) we observed that supervisors tended to reward individuals who bet on their own (ambiguous) predictions at high knowledge while avoiding them at low knowledge (values above 0.50 suggest a favorable preference: 95% CI [.64, .77], p < .001). In the supplementary materials, we present analyses of how supervisors rewarded all other choices.

**Discussion**

While the results of Study 1 showed that accountable decision makers did not sustain a financial cost for making reckless bets, the results of Study 2 suggest that decision makers sustained an indirect financial benefit for choosing ambiguity at high subjective knowledge while avoiding it at low subjective knowledge. Supervisors were more likely to reward this choice pattern than other patterns of choice under ambiguity. Taken together, Studies 1 and 2 suggest that accountable decision makers’ flexible approach allowed them to effectively manage multiple goals: In Study 1, they earned the same amount as unaccountable participants, and in Study 2, their pattern of choice under ambiguity increased their standing in their supervisors’ eyes.

**Study 3**

In Study 3, we return our focus to decision makers and consider two possible interpretations of the choice patterns we observed in Study 1. In one interpretation, decision makers’ flexible reactions to accountability resulted from the application of a fairly simple knowledge-heuristic strategy: Avoid ignorant bets and embrace knowledgeable ones. If this were a decision maker’s strategy, it would suggest that subjective knowledge is the primary guidepost for navigating ambiguity and that accountability pressures simply moderate the salience of subjective knowledge. In an alternative interpretation, decision makers’ reactions may result
from application of a primarily credit-and-blame focused strategy: Avoid ignorant bets and embrace knowledgeable ones because this approach is likely to maximize audience approval. If this were a decision maker’s strategy, it would suggest that anticipated evaluation—a focus on the peril and promise of ambiguity—is the primary guidepost for navigating ambiguity and that subjective knowledge moderates which bet is most likely to garner positive evaluations.

Our design in Study 1 did not allow us to differentiate between these two strategies. We varied subjective knowledge, which is common across the knowledge-heuristic and peril-and-promise strategies, but we did not vary the possibility of negative and positive evaluation, which distinguishes the two strategies. In Study 3, therefore, we test which strategy decision makers are more likely to employ: the knowledge-heuristic strategy or the peril-and-promise strategy.

To test which strategy decision makers are more likely to employ, Study 3 compares decision makers’ choices across two contexts: (a) one that varies knowledge and allows for negative and positive evaluation, and (b) one that varies knowledge but does not allow for negative and positive evaluation. If decision makers are using the knowledge-heuristic strategy to cope with accountability pressures, we should see the choice pattern observed in Study 1 in both contexts. If, by contrast, decision makers are using the peril-and-promise strategy to cope with accountability, we should only see the pattern when decision makers can be evaluated (the first context).

**Method**

The methods of Study 3 largely mimic those of Study 1; we highlight only important differences.

**Participants**
Again based on simulations from pilot data, we recruited NFL fans (N = 710) using a similar approach as before (259 female, $M_{\text{age}} = 38, SD_{\text{age}} = 11.35$). Four of the 710 participants dropped out of the study after giving their consent but before the accountability treatment. After we randomized participants into accountability treatments, 15 participants dropped out, 11 of whom were in the accountability condition. This was unlikely the result of chance ($\chi^2(1) = 3.27, p = .071$). However, in the event attrition was systematic, we again note the small marginal difference relative to our sample size (7 participants, 0.9% of total sample). As in Study 1, in order to operationalize the knowledge-level manipulation, we limited our test to individuals who declared that they knew more about the NFL than about the stock market (584/691, 84.5%).

**Design and Procedure**

**Overview.** Study 3 added an additional experimental condition to test whether the pattern of behavior observed among accountable decision makers in Study 1 appears only in contexts where audiences’ choices can be positively or negatively evaluated. Specifically, half of participants chose between their own prediction and a certain equivalent (as in Study 1), while the other half chose between their own prediction and a lottery whose odds were matched to the decision-maker’s confidence level (judged probability).

In supplementary studies, we found that both decision makers and evaluators believed that choosing one’s own prediction over a guaranteed alternative provided a clear basis for evaluation, while choosing one’s own prediction over a matched lottery did not. The latter, though technically less ambiguous than one’s own prediction (Heath & Tversky, 1991), provided a second form of betting on one’s judged probability in the context of uncertainty. Choosing one’s prediction over the safe space of certainty provided a basis for evaluation in the eyes of the audience; choosing one’s prediction over a risky alternative did not.
Adding the matched lottery condition led us to a 2 (Accountability: control, accountable) X 2 (Depth of Domain Knowledge: low, high) X 2 (Alternative Choice: certainty, matched lottery) mixed factorial design, with domain knowledge again a within-subjects factor.

Procedure and dependent measure. Participants in the certainty option condition followed the same procedure as in Study 1. Participants in the matched-lottery option condition read that they would face a choice between basing the bonus payment on their prediction ($.50 if correct, $0 otherwise) or on a lottery that “hit” as frequently as the confidence they stated for a given prediction. As an example, if a participant stated that the New England Patriots NFL team had an 80% chance of winning on Sunday, she faced a choice between $.50 if her prediction was correct ($0 otherwise) or a lottery that offered exactly an 80% chance of winning $.50 ($0 if they were unsuccessful in the lottery). All choices—one’s own prediction, the certain equivalent, and the matched lottery—were equal in expected value; only the certain alternative offered a safe alternative from the uncertainty inherent in one’s initial predicted likelihood.

The accountability treatment was unchanged. As before, we concluded with individual difference and demographic measures. For individual difference measures, we again administered the numeracy and concern for honor scales, as well as the social and financial components of the domain-specific risk scale (i.e., one’s tendency to take risks in social and financial domains; Weber, Blais, & Betz, 2002) and the short form of the need-for-cognition scale form (i.e., tendency to engage in and enjoy effortful cognition; Petty, Cacioppo, & Kao, 1984).

Results

Preliminary Analysis
Accountability manipulation. As expected, accountable participants were more likely than unaccountable participants to state that they considered how they would justify their views (95% CI [0.55, 1.10], t(582) = 5.89, p < .001) and how others would view their decisions (95% CI [0.93, 1.54], t(582) = 7.90, p < .001).

Subjective knowledge manipulation. As expected, participants reported higher confidence (judged probability) in the high (M = 78.40) versus low (M = 64.35) domains (95% CI [-14.78, -13.33], z = -38.04, p < .001).

Order of display manipulation. We observed no effect for order of display (whether high or low knowledge predictions appeared first on the screen); as in Study 1, we collapsed across ordering levels in subsequent analyses.

Alternative choice manipulation. Finally, given a pre-established preference for certainty, we found, as expected, that decision makers in the certain-alternative condition were less likely to bet on their predictions than decision makers in the matched lottery-alternative condition (log-odds 95% CI [0.35, 0.72], z = -3.84, p < .001).

Primary Analyses

Study 3 sought to test whether decision makers responded to accountability with a knowledge-heuristic strategy or a peril-and-promise strategy. If decision makers were using the first strategy, we should observe participants reacting to accountability by avoiding bets on ignorance and embracing knowledgeable ones regardless of the choice alternative they faced. However, if decision makers were reacting to accountability with the peril-and-promise strategy, we should observe differences in bets on knowledge only when audiences have a basis for evaluating them (e.g., only when they faced a certain equivalent).
The results revealed behavior consistent with the peril-and-promise strategy. Replicating Study 1, we found a significant interaction between accountability and knowledge among participants who faced a certain alternative (log-odds 95% CI [2.61, 8.24], $z = 5.23, p < .001$), with accountable participants being more likely to bet on their predictions at high knowledge (log-odds 95% CI [1.17, 4.81], $z = 2.39, p = .017$). When we examined the participants who faced the matched lottery, however, this pattern of results dropped away, and we did not observe a significant interaction between accountability and knowledge (log-odds 95% CI [0.65, 1.74], $z = 0.25, p = .806$). This implied a three-way interaction between accountability, knowledge level, and alternative choice, which we observed (log-odds 95% CI [0.25, 0.53], $z = -3.63, p < .001$; see Figure 1.2).

![Figure 1.2](image)

**Figure 1.2.** Decision makers facing a certain alternative (left panel) were more likely to avoid the ambiguous option at low knowledge and more likely to bet on the ambiguous option at high knowledge. Accountable decision makers facing a risky alternative did not show a change in preference toward ambiguity relative to unaccountable controls.
As in Study 1, accountable participants in the certain-alternative condition were indifferent between their predictions and the certain alternative at high knowledge (95% CI [0.47, 0.56], \( p = .586 \)), while all other participants in the certain-alternative condition preferred certainty (\( p < .001 \) for all cells). In the matched lottery-alternative condition, replicating prior work (Heath & Tversky, 1991; Fox & Tversky, 1995), both accountable and unaccountable decision makers preferred their own predictions to risky lotteries when they had high knowledge (95% CI [0.54, 0.61], \( p < .001 \)), but not low knowledge (95% CI [0.32, 0.39], \( p < .001 \)).

**Payoffs from bets.** Also as in Study 1, accountable decision makers earned as much money from their predictions as unaccountable decision makers (95% CI [-.015, .016], \( z = 0.40, p = .952 \)). This was true among participants facing a certain alternative (95% CI [-.012, .007], \( z = -0.45, p = .653 \)) and among those facing the matched lottery (95% CI [-.032, .049], \( z = 0.40, p = .689 \)).

We conducted exploratory analyses of individual differences across Studies 1 and 3. We observed a significant, negative relationship between numeracy and the tendency to bet on the ambiguous option (95% CI [-0.12, -0.29], \( z = -5.03, p < .001 \)). The more numerate respondents were, the less likely they were to bet on the ambiguous option. We found no other significant differences and no interactions across demographic variables or individual differences. As before, supplementary materials elaborate on these analyses and report measures of confidence and demographic variables.

**Discussion**

Study 3 sought to disentangle two possible decision-making strategies as explanations for accountability’s effects. One strategy, the knowledge-heuristic strategy, would suggest that accountability moderates the salience of subjective knowledge, making decision makers more
likely to avoid ignorant bets and to embrace knowledgeable ones. A second strategy, the peril-and-promise strategy, would suggest that accountability leads decision makers to avoid ignorant bets and to embrace knowledgeable ones only when doing so will affect audience approval.

The results suggest that the peril-and-promise strategy is the more likely explanation for accountability’s effects. Participants inferred that only when they turned from the safe space of certainty would their audience evaluate them negatively or positively, and only in this setting did accountability have an effect. In our final study, we test whether these inferences were accurate by measuring evaluator rewards of different patterns of decision-maker choice.

**Study 4**

Does this newly observed pattern of behavior under ambiguity cost decision makers in the eyes of their supervisors? To test this, we randomly varied whether supervisors evaluated decision makers who could turn to a certain equivalent or not (e.g., a matched lottery equivalent).

**Method**

**Participants**

Based on estimates of effect size and power analyses, we recruited 780 participants via mTurk (46% female, $M_{age} = 37$) using the same approach as in Study 2. Fifty-five participants dropped out after receiving treatment, and we observed no differential attrition across any of our treatment variables. Our final sample totaled 725.

**Design and Procedure**

The only design change was the addition of an alternative-choice factor. Mirroring Study 3, supervisors evaluated either decision makers who faced certain alternatives or decision makers who faced matched-lottery alternatives. This resulted in a 6 (pattern combination) X 2 (pattern
ordering) X 2 (alternative-choice) between-subjects design. Procedures were unchanged with the exception of an open-ended response question asking supervisors to explain their choices.

**Results**

**Preliminary Analysis**

**Order-of-display manipulation.** A series of logistic regressions using a dummy variable for ordering again showed no significant ordering effects within any of the choice pattern treatments. We collapsed across the variable for all subsequent analyses.

**Primary Analyses**

As before, when decision makers faced certain equivalents, supervisors systematically rewarded them for betting on ambiguous options at high knowledge and not low knowledge (probabilities greater than 0.50 suggest a favorable preference: 95% CI [.59, .73], *p* < .001). When decision makers faced matched lotteries, however, supervisors showed no systematic preferences (probabilities overlapping 0.50 suggest indifference: 95% CI [.45, .60], *p* = .495).

**Discussion**

Study 4, like Study 2, again showed that decision-makers’ inferences were not financially costly in the eyes of supervisors. In fact, decision makers’ reactions to accountability appeared finely tuned to the inferences that evaluators would draw from choices. Supervisors rewarded decision makers for betting on ambiguous options at high levels of subjective knowledge while avoiding them at low levels of subjective knowledge, but only when they could evaluate decision makers in the light of a certain alternative.

**General Discussion**

At the broadest level, these studies underscore the strong degree to which social factors drive choice processes. Revising prior conclusions, the present studies show that accountability
can either amplify or attenuate ambiguity aversion, depending on the circumstances. When decision makers were highly knowledgeable in a domain and could publicly signal that their initial predictions were sincere and reliable, accountability pressures led them to embrace (rather than reject) ambiguity. More broadly, the results suggest a view of decision makers as flexibly responsive intuitive politicians. Decision makers base their responses to evaluating audiences on, at least, their level of knowledge in a decision domain and the extent to which the decision setting allows them to gain audience approval. While they react defensively when deciding under low knowledge, they take more positive approaches to audience approval when deciding under high knowledge—that is, they stick their necks out and stand by their prior predictions rather than backing down.

The findings advance a growing body of work that incorporates motivational influences on decision making (e.g., Larrick, 1993), particularly motives to protect the self from negative evaluation. The findings also advance understanding of a social-contingency model for the effects of accountability, which posits that decision makers tailor their strategy for coping with evaluation to the situational context (Tetlock, et al., 1989). Our results expand the social-contingency model in at least four distinct ways. First, the experiments provide original systematic evidence that accountability does not necessarily increase ambiguity aversion. Second, the experiments show that accountable decision makers rely on contextual cues as they navigate ambiguous choices, namely their level of knowledge and whether the decision-making setting allows them to be evaluated for backing down. Third, because the studies financially incentivized all judgments and choices, they provide some of the first evidence that experimentally induced accountability alters choice behavior even when choices involve financial stakes. Finally, because these studies included actual supervisors who evaluated the
choices, they offer evidence that accountable decision makers were fairly adept in judging how their audience would reward them.

The findings also suggest the possibility of a relationship between accountability and theories concerning psychological distance. Higher subjective knowledge, reflected in higher probability estimates in participant predictions, likely entailed closer psychological distance (Wakslak, Trope, Liberman, & Alony, 2006). One complementary explanation for the findings is, therefore, that accountability may have led participants to choose psychologically proximal options while avoiding psychologically distant options as a means of garnering credit and avoiding blame with evaluative audiences.

**Methodological Innovation.** Across four studies we focus on both the accountable decision maker as well as the evaluating audience (supervisor). This pairing of studies examining both decision makers and audiences represents a new method in research on choice amidst accountability pressures. Most studies examine only the effects that accounting audiences have on their decision makers (for reviews, see Hall, Frink, & Buckley, 2015; Lerner & Tetlock, 1999), leaving open the question of how well (or poorly) decision makers placate their audiences. For the intuitive politician research program this question is central, similar to questions of the accuracy of intuitive scientists or the rationality of intuitive economists. Our paired decision maker/audience studies allow us to examine this question, and suggest a portrait of the intuitive politician as relatively savvy.

**Limitations.** To be sure, the present studies have limitations. First, they did not consider how social contingencies may differ across cultures. Preferences for consistency differ across relatively collectivist and individualistic cultures (Uchida, Savani, Hitokoto, & Kaino, 2017; Yates, Lee, & Shinotsuka, 1996; Yates et. al., 1998), and different preferences for consistency
may result in different levels of willingness to stand by one’s judgment. Further, decision makers in the present studies assumed that their supervisors would reward them for sticking their neck out based on their individual knowledge. This assumption may be more common in Western cultures than in Eastern cultures. In the United States, for example, independence and interpersonal disengagement are central to subjective well-being; whereas, in Japan, interdependence and engagement are central (Kitayama, Markus, & Kurokawa, 2000). One’s willingness to stick one’s neck out may differ, and depend on different factors other than subjective knowledge, across Western and Eastern cultures. Future studies could test these hypotheses.

Future studies could also differentiate distinct forms of accountability. Accountability for one’s decision-making process, for example, may have different effects than accountability for the outcome of one’s decision (Siegel-Jacobs & Yates, 1996; Patil, Tetlock & Mellers, 2016). Finally, future studies could consider how effects may vary in face-to-face accountability. While many modern workplaces involve the kind of online (distal) supervision present in our studies, face-to-face accountability should be examined in future studies.

Conclusion. Because high-stakes decisions rarely take place in a “social vacuum,” it is crucial to understand the effects that social/political factors such as accountability exert on judgment and choice (Tetlock, 1991, p. 454; Weigold & Schlenker, 1991), and how well decision maker reactions appease respective audiences. Prior studies taking this approach have painted a picture of decision makers in accountable contexts as defensive and rigidly reflexive in response to their social context. Drawing upon the flexible contingency model of choice under accountability (Lerner & Tetlock, 1999) as well as classic work on choice under ambiguity (e.g., Heath & Tversky, 1991), however, we hypothesized and found that decision makers are actually
flexibly responsive to their social context. Specifically, the present studies revealed that accountable decision makers were not uniformly more ambiguity averse than their anonymous counterparts, as prior work suggested. We found, instead, that accountability amplified ambiguity aversion only under certain social/structural conditions, and could in fact attenuate it under different conditions.

To our knowledge, the present fully-randomized experiments are the first, or at least among the first, to systematically examine choice under ambiguity when decision makers face audiences who can allocate or withhold financial rewards. Further, the present experiments are among the first to focus the microscope not only on accountable decision makers, but also on their evaluating audiences. By studying both, the present studies tested whether decision makers were savvy or inept judges of their audiences, revealing a more complete picture of accountable choice. Ultimately, the key novel finding in this research—that decision makers are not uniformly averse to ambiguity when they are held accountable—contributes to a growing theoretical understanding of socially-situated choice. The web of relationships surrounding choice can flexibly guide decisions. Moreover, because most decisions of consequence, from financial investments to military operations, occur in a social context and involve ambiguous odds, practical applications are myriad. Shaping choice in applied settings calls for a multifaceted portrait of decision makers, not least one that captures their political savvy.
References


overconfidence? *Organizational Behavior and Human Decision Processes*, 74(2), 89-117.

Essay 2

**When waste pays:**

*Equality versus efficiency in social context*¹¹

with Charles A. Dorison, Zoe Rahwan, Chris Robichaud, and Jennifer S. Lerner

“Don’t tell me what you value. Show me your budget, and I’ll tell you what you value.”
– Vice President Joseph Biden¹²

Scarce resource allocations force trade-offs between competing values. As former vice president Biden’s remarks suggest, amidst scarce resources it is less important to know that one values something than it is to know how much one values it in comparison to other things (Peterson, 1994). The vice president’s remarks—which he credits to his father, a used car salesman (Broder, 2008)—apply beyond elected politicians. Humans from all walks of life face limited resources, and thus the need to trade-off competing values that cannot be satisfied concurrently (Gordon-Hecker et al., 2017a).

A cardinal trade-off in resource allocations is the trade-off between the values of equality and efficiency (Choshen-Hillel, Shaw, & Caruso, 2015). On the one hand, individuals cherish equality (Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000; Fehr, Naef, & Schmidt, 2006). Historically, “All men are created equal” was a self-evident truth that helped motivate American colonists to risk their lives (US Declaration of Independence, 1776). Today, “equal pay for equal work” motivates social justice movements that seek to advance the same ideal (Lewin, 1997). At

¹¹ Funding from the Pershing Square Fund for Research on the Foundations of Human Behavior (PI: DeWees), the National Science Foundation (PI: Lerner; SES 1559511) and Harvard University’s Institute for Quantitative Social Sciences (PI: Lerner) supported this research.

the same time that individuals cherish equality, they also condemn wasting resources (Charness & Rabin, 2002; Engelmann & Strobel, 2002). Waste implies some degree of inefficiency.

“Waste not, want not” is an over-generalized rule, often observed even when its breach would be in an individual’s best interest (Arkes, 1996).

Equality and efficiency come into tension when some individuals can make better use of resources than others, or when resources cannot be divided equally without creating waste. This tension dates to humanity’s evolutionary past and the distribution of foraged resources (Hawkes, 1993). In more recent times, the trade-off remains pervasive and important (Messick, 1995; Mitchell, Tetlock, Newman, & Lerner, 2003). The trade-off is central to philosophers who debate the proper role of the state (Aristotle, 350 BC/2013; Locke, 1690/1980; Marx 1848/1906), firm managers who set prices (Bertsimas et al., 2012; Kahneman, Knetsch, & Thaler, 1986), doctors who triage patients (Guindo et al., 2012), and parents who divide their time among children (Arkes & Ayton, 1999). Given the ubiquity of equality-efficiency tradeoffs and their importance for decisions of consequence, it is important to understand how individuals manage when equality and efficiency tug in different directions.

**Theoretical Aims and Overview**

As they trade off values, individuals rarely occupy a “social vacuum” (Tetlock, 1991, p. 454; Weigold & Schlenker, 1991). Instead, trade-offs occur amidst a web of relationships that lead decision makers to anticipate the reactions of key constituencies and factor in how others will perceive their choices (Aberle et al. 1951; Tetlock, 2002; Vygotsky, 1980). The current research, as a complement to research examining decision features and individual differences, tests whether this social context systematically affects the trade-offs people make.
Specifically, three pre-registered experiments (N = 1,095) extend the value pluralism model (Tetlock, 1986; Tetlock, Peterson, & Lerner, 1996) to investigate the effect of social factors on equality-efficiency trade-offs. We test the hypothesis that observation increases equal (yet inefficient) allocation decisions (Studies 1-2). We also test whether observation shifts attentional focus from lives saved to equality (Study 2). Finally, we examine how observers draw inferences from allocator choices (Study 3).

In the following sections we review the value pluralism model, describing its early work and later revisions. Second, we narrow the focus of the model to our current domain, moving from value trade-offs broadly to equality-efficiency trade-offs in particular. In doing so, we draw on prior equality-efficiency research to theorize how the social environment affects allocator choices. Third, we offer falsifiable hypotheses and report three studies (N = 1,095) that test them. We conclude by discussing how our findings advance the value pluralism model and, more generally, how they reinvigorate research on values as key to understanding meaningful judgments and choices.

The Value Pluralism Model

The value pluralism model postulates that individuals possess an implicit internal hierarchy of core or terminal values, and that the ordinal ranking of these values serves as a starting point for trade-off judgments (Tetlock, 1986; Seufeld, Bluck, Loewen, & Elkins, 1994; Tetlock, Michelitti, & Hanum, 1984; Tetlock, Peterson, & Lerner, 1996). An individual’s internal ranking of equality relative to efficiency, for example, offers a useful starting point in predicting that individual’s allocation choices. The value pluralism model further holds that individuals differ in the extent to which they acknowledge high-priority values as conflicting. Two individuals may each value equality and efficiency when the values occur in isolation, yet
differ in whether they acknowledge that the values conflict during a trade-off. Rule-based value systems, for example, could generate such differences—if one believes that others should always receive equal treatment, an equal-versus-efficient trade-off does not pose a conflict even if efficiency is desirable in isolation.

Early work on the value pluralism model focused on predicting individual reasoning style when making trade-offs (Suedfeld, Bluck, Loewen, & Elkins, 1994; Tetlock, 1986; Tetlock, Michelitti, & Hanum, 1984). While the original model offered predictive power for trade-off reasoning style, the value pluralism model eventually grew to incorporate two key insights: both 1) the “social context” of a decision maker and 2) the “social content” of colliding values each affect how decision makers navigate trade-offs (Tetlock, Peterson, & Lerner, 1996).

Specifically, the revised value pluralism model stipulated that accountability pressures from one’s social context—the implicit or explicit expectation of justifying one’s views to others (Lerner & Tetlock, 1999)—could alter the cognitive processing style a decision maker uses to navigate trade-offs. An audience with unknown views, for example, is likely to trigger complex processing as decision makers preemptively address a wide range of possible audience positions (Tetlock, Skitka, & Boettger, 1989). Further, the model stipulates that the social content of values affects the likelihood of complex trade-off reasoning. Individuals who define a particular value as sacred, “possessing infinite or transcendental significance” (Tetlock et al., 2000), are unlikely to engage in complex trade-off reasoning when such values collide with secular values (Tetlock, 2003). To compare a sacred value against the value of money, for example, would destroy the transcendence that defines the sacred value in the first place.

Social Context and Allocator Choice
In the present work, we draw on the social context postulate to argue that observation from others systematically affects allocator choices. The premise of the social context postulate is that, when making trade-offs of real consequence, individuals are accountable to important constituent audiences (Tetlock, Peterson, & Lerner, 1996). A variety of research programs across multiple disciplines supports this premise in the context of equality and efficiency. Neuroscientific research, for example, suggests that considering social norms raises the value of equality (Hsu, Anen, & Quartz, 2008). Increased activity in the caudate/septal subgenual region of the brain, a region associated with monitoring social norms, is associated with a greater preference for equality (Hsu et al., 2008).

Separately, classic work in social psychology suggests that accountability—expecting to justify one’s views to an evaluating audience (Lerner & Tetlock, 1999)—raises the value of equality relative to efficiency. Adelberg & Batson (1978) showed that allocators who were accountable to either recipients of an allocation or the owners of allocated resources were more likely to prefer equal allocations than were anonymous controls. We note two cautionary points in extrapolating from this finding, however. First, the test was not incentive compatible, in multiple ways: allocators dispersed hypothetical resources and were accountable only to hypothetical individuals. Second, the decision makers were accountable to audiences who had stakes in the allocation (i.e., those receiving the resources and those who owned the resources). It is not clear that accountability to these audiences will have the same effect as observation by 3rd-parties who have no stake in the allocation. More generally, the effects of accountability can vary dramatically based on the accounting audience, and an effect of accountability with one type of audience does not necessarily translate to an effect with a different type of audience (for a review, see Lerner & Tetlock, 1999).
Another clue about the social context’s effects lies in the apparent motives on which decision makers act when they trade-off equality and efficiency. One important motive is the desire to appear impartial: individuals are more likely to prefer efficient allocations when the decision context allows them to avoid showing favoritism toward themselves or close others (Choshen-Hillel et al., 2015; Shaw et al., 2016; for a review, see Shaw, 2013). For example, when individuals can make efficient allocations that leave themselves relatively worse off than other allocation recipients, they are more likely to prefer efficiency (Choshen-Hillel et al., 2015). Similarly, individuals are averse to taking responsibility for unequal allocations, even if they would accept such allocations as a result of randomness (Gordon-Hecker et al., 2017b). These findings appear rooted in interpersonal concerns. One wants to appear impartial to others. To the extent that observation triggers such concerns, its effects may parallel the effects of decision contexts that force impartiality, increasing the likelihood of equal yet inefficient allocations.13

Hypotheses & The Present Research

We hypothesize that pressure from the social context—in particular, third-party observation—raises the value of equality relative to efficiency and increases the likelihood of equal (yet inefficient) allocations, even when efficiency is Pareto-optimal. Two key points about this hypothesis merit note. First, by focusing on third-party observation rather than observation by those with a direct interest in a resource allocation, we study a more general effect of social pressure beyond pandering to resource stakeholders. Second, by examining Pareto-optimal

13 While the desire to appear impartial may ultimately be rooted in interpersonal concerns, its effects may result from intrapsychic factors that are present regardless of whether an individual is under observation. Though interpersonal and intrapsychic explanations can be notoriously difficult to disentangle (Tetlock & Manstead, 1985), there are important theoretical and practical reasons for doing so (Leary, Raimi, Jongman-Sereno, & Diebels, 2015). Theoretically, the two explanations imply substantially different mechanisms underlying human behavior (i.e., the maintenance of internal states versus the bolstering of one’s standing with others). Practically, decision architects focused on maximizing group benefit, for example, would need to know whether the mere presence of others has an effect on equal-efficient preferences.
efficiency, we likely construct a conservative test. Pareto-optimal allocations make at least one recipient better off while making no recipients worse off (Lockwood, 1987). Such allocations are rare in real-world allocations (Denhardt, 2004), where efficiency often requires a reduction in benefits for some people (though a reduction outweighed by benefits elsewhere). If third-party observation increases the likelihood of equal allocations even when efficiency makes no one worse off, we likely underestimate the social pull toward equality in settings where efficiency does make some worse off.

We test this hypothesis in Studies 1 and 2. The two studies extend the value pluralism model in several ways. They extend the model to incentive-compatible, equal-efficient trade-offs. They extend the model to a general form of pressure from the social context—observation by third-parties to a resource allocation. Last, they extend the value pluralism model from predicting reasoning style to predicting ultimate choice.

**Open Science**

In keeping with best practices for fully-reproducible science (Simmons, Nelson, & Simonsohn, 2011), we report all methodological decisions (e.g., determining sample size), manipulations, and measures. Data, code, preregistrations, and materials are available [here](#). We pre-registered all studies and did not use deception.

**Study 1**

Study 1 tests the hypothesis that third-party observation increases equal (yet inefficient) allocation decisions, even when efficiency is Pareto-optimal. We give allocators a choice between giving the same amount of resources to all, or, relative to this equal allocation, giving additional resources to some without taking resources away from others. The efficient allocation
is Pareto-optimal in terms of material resources.\textsuperscript{14} We randomly assign allocators to make choices either under the observation of future trust game partners or anonymously. We predict that observation will raise the value of equality relative to efficiency, leaving allocators in the observation condition more likely to make equal (yet inefficient) allocations than participants in the anonymous condition.

We use a real-stakes trust game to place observation in an incentive-compatible setting. This approach aligns our manipulation of the social context with a “core social motive” (Fiske, 2003) that underpins human cooperability (Baumard et al., 2013). In our paradigm, both allocation choice and the relationship with observers involve real-stakes. Last, the use of a trust game makes our findings on observer inferences (Study 3) more comparable to similar work involving decisions in other distinct moral dilemmas (e.g., Everett et al., 2018; Jordan et al., 2016).

Additionally, we use this study to rule out alternative explanations, namely the possibility that observation could lead to more equal allocations \textit{without} raising the value of equality. If an allocator were to infer that her/his observing partner would be more likely to reward an equal allocation than an efficient allocation, for example, s/he could use an equal allocation strategically without raising her valuation of equality. We use a combination of approaches to rule out this explanation, including probes of allocator motives, measures of individual differences in strategic reasoning, and comparisons of allocator reciprocity across observation conditions.

\textsuperscript{14} We specify Pareto-optimality in terms of material resources in particular because Pareto-optimality can be a values-plural concept. When decision makers derive utility from multiple sources of value, Pareto-optimality occurs when one cannot improve on any single source without making another source worse off (for a discussion, see Grandy, 2009). Understood in this broader sense, an equality-versus-efficiency can never allow for Pareto-optimal decision options as long as decision makers value both equality and efficiency.
Method

Participants

Simulations from pilot data suggested that a sample size of 450 would adequately power our hypothesis. We recruited 457 participants via mTurk for a “Survey on Judgment and Decision-Making.” In line with our pre-registered analysis plan, we included only participants who correctly answered at least five out of seven basic comprehension questions. Approximately 11% (n = 53) participants failed to meet this criterion, resulting in a final sample of 404 (M_age = 37.25, 52.5% female). Importantly, robustness checks on our data reveal that the overall results remain consistent with the patterns reported below regardless of how we adjust this exclusion criterion (all data available via the link above).

Design and Procedure

Design. We utilized a 2 (Observation: Observed, Anonymous) X 2 (Order of Allocation Options: Equal-first, Efficient-first), between-subjects design. In addition to randomly assigning participants to make their allocation decision while being observed by a future trust game partner or anonymously, we also randomly assigned the order in which the allocation options appeared. Our dependent variable was an incentive-compatible binary choice between an equal (yet inefficient) allocation or an efficient (yet unequal) allocation.

Procedure. For this and all subsequent studies we obtained informed consent. After informed consent, participants, whom we refer to as “allocators” in this study, read that they would participate in a two-stage task (see Figure 2.1). In the first task, they would allocate bonus

---

15 In order to avoid deception, we recruited a similar number of participants to play in the opposite side of the trust game (explained in more detail below). As these participants were not pertinent to our hypotheses, we describe that sample and their choices in supplementary material.

16 The research was reviewed and approved by the Harvard University-Area Committee on the Use of Human Subjects, protocol number IRB17-1343.
payments to other mTurk workers (whom we call “receivers”). In the second task, they would
play a game with another mTurk worker (whom we call a “partner”). We clarified that the
receivers and the partner were not the same.

The bonus allocation decision (the first task), which we generically described as a
“dilemma,” came before the trust game (the second task). However, in order to introduce the
prospect of observation by a trust game partner, we first instructed allocators on the trust game
and the role their allocation would play in the game. In sum, allocators learned about the second
task, undertook the first task, and then undertook the second task.

The trust game instructions stated that another mTurker—the partner—would begin with
$.30 and would have the opportunity to transfer any amount, including none at all, to the
allocator. We would triple any amount that partners transferred. Allocators would be able to
decide how much of the tripled amount, if any, they wanted to transfer back to their partner.
Allocators in the Observed condition read that their partner would “decide whether to play the
game with you based on your decision in the dilemma.” Allocators in the Anonymous condition
read that their partner would “have no knowledge of your decision in the dilemma. Your choice
in the dilemma will NOT influence the outcome of the game.”

Allocators in both conditions then made identical allocation decisions. This decision
served as the primary dependent variable. Allocators read that three other mTurkers—
“receivers,” none of whom was the partner who would play in the trust game—had recently
completed a task. Allocators could award up to four bonus tokens worth $.20 each. They could
divvy the tokens in one of two ways. In the equal allocation they could award three tokens,
giving one token to each recipient. In the efficient allocation they could award all four tokens,
giving one token to two recipients and two tokens to a third. This efficient allocation was Pareto-
optimal in that it made one the receivers better off while making none of them worse off. The order in which these two allocations appeared varied according to Order condition. We predicted that allocators in the Observed condition would be more likely to choose the equal allocation than would allocators in the Anonymous condition.

After this choice, allocators then played the trust game by indicating what percentage, if any, they wanted to return to their partner.

We concluded the study with manipulation checks, tests for strategic reasoning, and demographic variables. As a manipulation check, allocators indicated the extent to which they agreed with the following statement: “When deciding between the equal and efficient allocation, I considered how the Sender [the Partner] would view my choice.”
Next, to test for strategic reasoning, we asked two sets of questions. First, to measure beliefs regarding the signal value of their allocation, we asked a question meant to clarify participant’s goals for their allocation decision and how important they believed the decision was to their partner’s play in the trust game. We asked allocators to “Imagine that your goal was to get the Sender [partner] to trust you as much as possible. Which allocation would you have chosen?” Half of allocators read a version of the question that substituted “transfer the greatest amount of money possible to you” in place of “trust you as much as possible.” Their decision options for this question mirrored those in the actual allocation decision (equal or efficient).

Second, we administered an individual difference measure of strategic reasoning (LeVeck et. al., 2014).

The study concluded with demographic questions on political ideology, education, whether religion was an important part of the individual’s life, gender, and age.

**Results**

**Preliminary Analyses.** Observed allocators were more likely to agree with the statement that they considered how their trust game partner would view their allocation choice (95% CI [1.62, 2.37], \( t = 10.46, p < .001 \)). We did not observe a main effect for Order on allocation decision (choice between equal or efficient allocation; log-odds 95% CI [.78, 1.70]; \( t = -0.71; p = .479 \)). Contrary to predictions, we observed an interaction between our Observation condition and the Order in which we presented the equal and efficient allocations (log-odds 95% CI [1.17, 5.70]; \( t = 2.34; p = .019 \)). We elaborate on this in supplementary material, though note here that a similar ordering variable fails to interact with our treatment Study 2 (or its direct replication).

**Primary analysis.** Our main hypothesis was that, even with real money on the line, allocators in the Observed condition would be more likely to make equal (yet inefficient)
allocations than those in the Anonymous condition. This was in fact the case. Allocators in the Observed condition chose the efficient allocation 47% of the time, while allocators in the Anonymous condition chose the efficient allocation 59% of the time, a statistically significant difference (log-odds 95% CI [.42, .92]; cohen’s d = .24, t = -2.38; p = .017). Indeed, allocators in the Observed condition were approximately 20% ((59-47)/59) less likely to choose the efficient allocation than were allocators in the Anonymous condition.

Secondary analysis—was the shift towards equality a strategic move?

Trustworthiness. Observed allocators were directionally, though not significantly, more likely to show reciprocity to their partners in the trust game (MObserved = 36.87, MAnonymous = 32.96, 95% CI [-0.51, 8.33], cohen’s d = .17, t = 1.74, p = .083). While tentative, this is initial evidence that observation led to more equal allocations because it raised the value of equality, not because allocators used the allocation strategically in order to maximize profits in the game.

Beliefs. We asked participants about their beliefs regarding which allocation decision would be most useful as a signal in the trust game. We phrased this question in two separate ways, with half of participants seeing a question that emphasized money and the other seeing a question that emphasized trust. When the question was worded in terms of money (which choice would convince the trust game partner to “transfer the greatest amount of money possible to you”?), most participants (51%) did not express a preference one way or the other (they believed that the two allocations would have similar effects). Those who did were nearly evenly divided over which allocation decision would be most effective (22% favored efficient, 27% favored equality; this latter division was indistinguishable from chance, χ² = .818, p = .366).

---

17 These results remain unchanged when we include all participants in the sample regardless of their answers to comprehension check question (log-odds 95% CI [.47, .99]; cohen’s d = .19, t = -2.01, p = .044).
When we phrased the question in terms of trust (which choice would convince the trust game partner to “trust you as much as possible”?), we observed a greater willingness to express an opinion. Only 30% chose not to express a preference, statistically less than the percentage who withheld their opinion when the question was framed in terms of money ($\chi^2 = 17.13, p < .001$). Among those who did express an opinion, a strong majority believed the equal allocation would work best (13% favored the efficient allocation, 56% favored the equal allocation, a split unlikely due to chance, $\chi^2 = 52.13, p < .001$). In terms of the economic game, the two questions asked the same thing. However, phrasing the question in terms of trust seemed to lead participants to think equality was favorable. This is an additional piece of evidence that allocators did not use the equal allocation strategically. They were unlikely to view the equal allocation as a means to earn more money in the game.

**Strategic Reasoning and Demographics.** Again contrary to the explanation that allocators would use the equal allocation strategically, we did not observe an interaction between our measure of strategic reasoning and our Observed condition (log-odds 95% CI [0.34, 1.96]; $t = -0.46; p = .646$). High strategic reasoners—individuals most likely to anticipate the behavior or judgment of others and adapt their own behavior accordingly (LeVeck et. al., 2014)—were no more likely to change their allocation in response to observation.

Last, we did not observe interactions between our Observed treatment and measurements of political ideology, education, religiosity, age, or gender.

**Discussion**

Study 1 tested whether the social context—specifically, observation by third-parties to an allocation—affects trade-offs between equality and efficiency in real allocation decisions. In an extension of the value pluralism model, the results showed that the social context can alter the
underlying value hierarchy of allocators. Observed allocators valued equality more than allocators operating in an anonymous setting. This was true despite the fact that efficient options were Pareto-optimal from the perspective of resource recipients. None were worse off as a result of efficiency.

The increased emphasis on equality did not appear to be the result of temporary strategizing in order to maximize one’s payoff in the economic game. Observed allocators were directionally more likely to reciprocally cooperate with their partners by returning money in the game, few allocators believed that the equal allocation would lead to more money in the game, and observed allocators who scored high on a measure of strategic reasoning were no more likely to make equal allocations than observed allocators who scored low in strategic reasoning.

**Study 2**

Study 1 provided initial evidence that the social context exerts a causal effect on choices involving equal-efficient trade-offs. Study 2 builds upon Study 1 in several ways. First, it directly primes observation rather than manipulating it through an economic game. It could be the case that the game context of Study 1 made salient a host of concerns beyond observation alone. Second, Study 2 introduces a situation of scarcity, where the need for resources exceeds the availability of resources. Scarcity is a common feature of real-world allocations, but also triggers decision making processes that are distinct from the processes that allocators use when resources exceed needs (Skitka & Tetlock, 1992). A scarce context leads allocators to consider factors such as the efficiency with which they can distribute resources (Skitka & Tetlock, 1992).

Third, Study 2 tests the generality of the effect of pressure from the social context by introducing a new type of equal-efficient trade-off: lives saved rather than money. Prior research shows that allocator preference for efficiency varies according to the resource being allocated.
Efficiency becomes more important when allocations affect life and death (Li, Colby, & Fernbach, 2018). Likewise, the effect of pressure from the social context on equality and efficiency tradeoffs may systematically vary across types of allocations. Decision-makers may anticipate that choosing equality over efficiency when lives are at stake would appear sufficiently reckless as to undermine their standing with others. We addressed this possibility by examining the effects of social pressure in an organ allocation decision. Last, this study allows us to further test our claim that the social context is altering an individual’s value hierarchy, rather than merely shifting final choice for instrumental reasons. In this study we rely on a writing task to prime observation, which allows us to examine the topics that individuals focus on prior to making their decision. This approach allows us to test whether an increased focus on equality relative to efficiency is indeed responsible for carrying the effect of observation, as we have hypothesized.

**Method**

**Participants**

Based on simulations we determined that a sample size of 425 participants would yield adequate power to test our primary hypothesis. We added 25 participants per cell to account for possible exclusions, leading to a target sample size of 475. We advertised an “Ethical Decision Making Survey” via Amazon mTurk, and collected 480 complete survey responses. Thirty participants did not follow instructions and were removed according to pre-registered criteria. This resulted in a final sample of 450 ($M_{age} = 37.32, 44.4\%$ female). As in Study 1, our main findings remain consistent regardless of exclusion criteria.

**Design and Procedure**
Design. All participants, whom we again call “allocators,” made hypothetical allocations of life-saving organs. Before making their allocation decision, we randomly assigned allocators either to consider trustworthiness in the eyes of the general public (Trustworthiness Salient condition) or to consider any factor that they believed most important to the decision (Control condition). As in Study 1, we counter-balanced the order in which the equal and efficient allocation options appeared to decision makers. The two factors resulted in a 2 (Trustworthiness: Salient, Control) X 2 (Order: Equal-first, Efficient-first), between-subjects design.

Procedure. Adapting a scenario from Ubel & Loewenstein (1996), we asked all allocators to make an organ allocation decision as the director of a regional allocation board in the United States. They read that 200 children were on a waiting list to receive a liver transplant, but that only 100 livers were available for distribution. The children were organized into two groups according to their likelihood of survival after a liver transplant. In one group of 100 children, each child had an 80% chance of survival; in the other group of 100 children, each child had a 50% chance of survival. The participant’s task was to “decide what percentage of the 100 livers should go to each group” (unlike Study 1, there are no Pareto-optimal allocations in this scenario, since giving an organ to one child necessarily deprives another child of the organ). A concern for efficiency in this scenario would lead allocators to distribute the livers to the group with the higher likelihood of survival, leading to a higher expected number of survivors. A concern for equality would lead allocators to distribute the livers equally to each group, giving all children an equal chance at receiving a liver but leading to a lower number of expected survivors (assuming a 50-50 split between the groups). Our primary dependent variable was the proportion of livers allocated to the group with the higher likelihood of survival.
Before making their choices, we asked allocators to consider and write about either “trustworthiness in the eyes of the general public,” (Trustworthiness Salient condition) or “the factor that you believe is most important” (Control condition). We required a minimum of 150 characters. After writing, allocators made their allocation decision by indicating the percentage of livers they would like to send to each group, using a sliding bar to indicate the percentage of livers they wished to allocate to each group (see Figure 2.2). This slider forced the participant’s allocations across the two groups to sum to 100.

![Figure 2.2: Allocation decision as seen by participants. Our primary dependent variable was the proportion of livers allocated to the group with the higher likelihood of survival, in this case Group 1 (the order in which the efficient and equal groups appeared was counter-balanced).](image)

We concluded the scenario by offering allocators the opportunity to provide any other rationale for their choice that they desired. We asked the same demographic questions as in Study 1.

**Results**

**Preliminary Analysis.** We coded participant writings to determine the extent to which allocators focused on public trustworthiness. Two coders who were blind to condition and outcome used a binary measure to indicate whether a participant expressed a concern for public trustworthiness (1 if yes, 0 if no). The two raters agreed 97% of the time and showed a Cohen’s kappa of .93. In the cases of disagreement, we averaged the two rater’s scores; possible variable
levels were thus 0, 0.5, and 1. Regressing this variable on our Trustworthiness Salient condition showed an effective manipulation (95% CI [0.45, 0.58], \( t = 15.98, p < .001 \)).

We observed neither a main effect of Order (95% CI [-5.02, 3.27], \( t = -.416, p = .67 \)), nor an interaction between Order and our treatment (95% CI [-6.57, 9.96], \( t = .403, p = .69 \)). Thus, we collapse across Order in all ensuing analyses.

**Primary analysis.** Our aim in this study was to determine whether priming concerns for trustworthiness leads to less efficient allocations when the decision involves life or death. We compare the proportion of allocations to the efficient group across conditions to test this question. Participants in the Control condition allocated 78.5% of their organs to the group with the highest chance of survival, while participants in the Trustworthiness Salient condition allocated 73.5% to the high-survival group. Even with lives on the line, concerns for trustworthiness led to less efficient allocations (95% CI [-9.11, -0.87]; cohen’s \( d = 0.23; t = -2.38; p = .018 \)).

**Mediation analysis.** The revised value pluralism model predicts that the social context can influence the ordinal ranking of values. To explore this, we conducted a mediation analysis of the written responses of allocators.

We coded participant writings to determine the extent to which participant writings focused on 1) lives saved and 2) equality. We used the same coding procedure that we used to code concerns for public trustworthiness, and observed Cohen’s kappas of .78 (91% agreement) and .86 (95% agreement) for lives saved and equality, respectively. We fit a structural equation

---

18 We conducted a direct replication of this study (N = 300; \( M_{age} = 36.32, 50\% \) female) and found nearly identical results. Those in the Trustworthiness Salient condition made less efficient allocations than those in the Control condition (95% CI [-10.32, -0.99]; \( t = -2.38; p = .018 \)). We conducted both studies because the first – the direct replication – was not pre-registered.
model using the Lavaan package in R (Rosseel, 2011). The results of the structural equation model are displayed in Figure 2.3.

In line with predictions from the revised value pluralism model, we find evidence that the allocators in the Trustworthiness Salient condition made less efficient allocations than did allocators in the Control condition as their focus shifted from lives saved to equality. Compared to allocators in the Control condition, allocators in the Trustworthiness Salient condition were more likely to focus on equality (95% CI [0.05, 0.20]; \( t = 3.31; p = .001 \) ) and less likely to focus on lives saved (95% CI [-0.33, -0.18]; \( t = -6.68; p < .001 \) ). While a focus on equality was associated with less efficient allocations (95% CI [-18.45, -9.83]; \( t = -6.43; p < .001 \) ), a focus on lives saved was associated with more efficient allocations (95% CI [17.05, 25.57]; \( t = 9.81; p < .001 \) ).

We observed a significant indirect effect of trustworthiness salience on equal (yet inefficient) decisions through both an increased focus on equality (95% CI [-2.95, -0.59]; \( t = -2.95; p = .003 \) ) and a decreased focus on lives saved (95% CI [-7.39, -3.51]; \( t = -5.52; p < .001 \) ).
Secondary analyses—demographics. Contrary to Study 1, we observed an interaction between our treatment and the reported importance of religion in a participant’s life, which we elaborate on in supplementary material. We observed no interactions between our treatment and each of our other demographic measures (political ideology, education, gender, and age).

Discussion

Study 2 builds upon Study 1 by showing that pressure from the social context outside the context of an economic game increases equal (yet inefficient) allocations. Furthermore, we document that this effect generalizes to a domain in which efficiency tends to be more valuable than equality: life-or-death allocations.
In line with hypotheses drawn from the social context postulate of the value pluralism model, Study 2 also provides evidence that observation can influence the ordinal ranking of values. Specifically, we find evidence that allocators in the Trustworthiness Salient condition made less efficient allocations because of a heightened concern for equality relative to efficiency.

**Study 3**

Our first two studies asked whether individuals are “intuitive politicians” (Tetlock, 1991) as they navigate equal-efficient trade-offs—do they respond to pressure from the social context by changing their allocation? Our third study builds on the first two by asking a different but related question: how good are allocators as intuitive politicians? Do observers judge allocators who prefer equality more favorably than allocators who prefer efficiency? To answer this question, we shift our focus in the present study from allocators to observers of allocators. Implicit in our focus on both allocators and observers is the recognition that individuals may be imperfect managers of their reputational interests. Allocators may attempt to respond to social pressures, but because mental models of others’ beliefs are often inaccurate (e.g., Epley, Keysar, van Boven, & Gilovich, 2004; Levine et al., 2018; Tamir & Mitchell, 2013), allocator responses may not successfully appease observing audiences.

To hypothesize how observers judge allocators who trade-off equality and efficiency, we draw on prior empirical work examining the moral character inferences that individuals draw from the value-laden decisions of others (even if not the values of equality and efficiency e.g., Everett, Faber, Savulescu, & Crockett, 2018; Everett, Pizarro, & Crockett, 2016; Uhlmann, Zhu, & Tannenbaum, 2013). In sacrificial dilemmas, where a smaller number of people must be harmed in order to save a greater number, observers infer that those who rely on deontological moral reasoning (i.e., based on duties and rights) are more morally trustworthy than those who
rely on consequentialist moral reasoning (i.e., based on benefit maximization; Everett, Pizarro, & Crockett, 2016; Everett, Faber, Savulescu, & Crockett, 2018; Sacco, Brown, Lustgraaf, & Hugenberg, 2017; Uhlmann, Zhu, & Tannenbaum, 2013). Similarly, costly punishment of selfishness can signal trustworthiness to observers when other, more direct signals are not available (Jordan, Hoffman, Bloom, & Rand, 2016). Observers also infer that those who give aid to others are trustworthy, especially if they do so without calculation (Jordan, Hoffman, Nowak, & Rand, 2016).

Overall, this body of research suggests that, contrary to the predictions of rational actor models (e.g., Berg, Dickhaut, & McCabe, 1995), observers tend to see calculating, benefit-maximizing individuals as less suitable cooperative partners than individuals who appear to operate on pro-social rules (Rand et al., 2014). While these findings do not translate directly to equal-versus-efficient dilemmas—there is no instance of harm as in sacrificial dilemmas and punishment decisions, for example—they do imply that observers may, on average, be skeptical of the cooperability of those who make payoff-maximizing efficient allocations.

The effect of allocator decision on observer perception may also depend on the observer’s own preferences. Individuals often judge the morality of certain actions egocentrically. For instance, judgments of dishonesty depend on whether the judge has a close relationship with the individual whose honesty is in question (Hildreth & Cameron, 2018), and individuals judge the morality of cheating based on their own views, especially what is in their self-interest (Bocian, & Wojciszke, 2014; Levine et al., 2018). In the equality-efficiency domain, while skepticism of efficiency may prevail on average, there is likely substantial variation at the level of individual observers. Those observers who prefer efficiency may egocentrically judge allocators.
Further, it may be the case that observers who themselves prefer equality will condemn allocators who prefer efficiency to a greater extent than observers who themselves prefer efficiency will condemn allocators who choose equality. “Treat others equally” is an injunction that aims directly at the management of relations with other humans. “Waste not” aims directly at the management of resources. Rules regarding relations with others are more likely to be deemed “sacred” than rules regarding the management of resources (Fiske & Tetlock, 1997). When thinking in terms of sacred values, “to compare is to destroy,” both the value itself and one’s standing with others (Tetlock et al., 2000). To the observer who values equality, an allocator’s equal-efficient trade-off can be an opportunity to observe adherence to a sacrosanct rule. If this is the case, observers who value equality may be more likely to draw character inferences from allocations than observers who value efficiency, with the former viewing allocation choice as a signal of one’s intrinsic sociability.

**Study 3 Hypotheses**

Based on the above work we pre-registered three overarching hypotheses. First, we predicted that, on average, observers will trust allocators who make equal allocations more than allocators who make efficient allocations, even when efficiency is Pareto-optimal (as in study 1, we return to a decision setting in which allocators can make some better off with efficiency without making others worse off). Second, we predicted an interaction such that the effect of allocator choice on observer trust depends on an observer’s own equal-efficient preference. Specifically, we hypothesized that while observers who prefer equality will draw negative inferences about allocators who disagree (i.e., choose efficiency), observers who prefer efficiency will trust equal and efficient allocators the same. Finally, we hypothesized that this asymmetric pattern of equal (but not efficient) observers mistrusting those who disagree would
be underpinned by an asymmetric pattern of perceptions of warmth but not competence. Specifically, observers who prefer equality will perceive allocators who also prefer equality as warmer than allocators who prefer efficiency. However, observers who prefer efficiency will see both types of allocators—those preferring equality and those preferring efficiency—as equally warm. We predicted that this asymmetry in perceptions of warmth will mediate the asymmetry in actual trusting behavior.

**Method**

As in Study 1, we utilized a two-stage task. First, allocators dispensed bonus payments in either an equal or Pareto-optimal efficient manner. Second, observers played a trust game with allocators in which they could condition their play on the allocator’s choice between equality and efficiency. We focus our attention here on the behavior of observers.\(^{19}\)

**Participants**

Simulations with pilot data showed that samples of 200 or more would yield adequate power for all hypotheses related to observer behavior. To account for possible exclusions, we added 50 to this number to for a sample goal of 250 observers for a “Survey on Judgment and Decision-Making” through mTurk (as well as the same number of allocators in order to avoid deception). We eventually recruited 252 observers. According to pre-registered exclusion criteria, we removed participants who missed at least three out of six basic comprehension questions. Eleven observers met this criterion, resulting in a final sample of 241 observers ($M_{age} = 39.65$, 49.8% female). As in Studies 1-2, the results remain consistent whether or not we enforce this exclusion criterion.

**Design and Procedure**

\(^{19}\) We report allocator behavior in supplementary material.
Design. Observers decided how much of their bonus, if any, they wanted to entrust to allocators who made the equal allocation or the efficient allocation. This “strategy method” approach asked individuals to make their decisions for all possible allocator choices before the game occurred (Camerer, 2003; Jordan et. al., 2016). We randomized the order in which equal and efficient allocator choices appeared to observers. This resulted in a 2 (Allocator Choice: Equal, Efficient) X 2 (Order: Equal-first, Efficient-first) mixed-factorial design, with Allocator Choice a within-subjects factor and Order a between-subjects factor.

Procedure. After informed consent we described the rules of the trust game, which we referred to only as a “game” in instructions to participants. We gave observers a $.30 bonus to begin, then told them that they would have the chance to send any portion of that bonus, including none at all, to the allocator. We would triple what they sent, and allocators would then have the chance to return any portion of what they received. We quizzed participants on the rules of the game.

After initial game instructions, we described the equal or efficient allocation decision using the same language as in Study 1. The end of the allocation decision instructions stated that observers would be able to base how much they sent on the allocator’s allocation (see Figure 2.4 for a depiction of the task). As with the trust game instructions, we quizzed participants on their understanding of the allocation decision and its application to the game.
Observers then decided how much money to transfer to an allocator making the efficient allocation and how much money to transfer to an allocator making the equal allocation (with order of decision varied by Order condition). Observers made their two trusting decisions on separate screens, using a sliding scale defaulted to zero. We hypothesized that observers would trust equal allocators more than efficient allocators.

We then asked observers which allocation—equal or efficient—they themselves would have chosen, and which allocation they thought would be easier to defend to others. As described above, we hypothesized that while observers who prefer equality will draw negative inferences about allocators who disagree (i.e., choose efficiency), observers who prefer efficiency will trust equal and efficient allocators the same.

---

Figure 2.4: Observers engaged in a two-stage task in Study 3. In the first task, observers learned of another participant’s allocation decision—allocators could either distribute a bonus payment equally but inefficiently, or efficiently but unequally (bottom left of the diagram). As in Study 1, the efficient allocation made one of the bonus-recipients better off while it did not make the other two worse off. In the second task, observers decided how much of their own bonus they wished to entrust to the allocator from the first task.

---

20 To assess the stability of these preferences, we contacted observers approximately two months after they completed the initial survey and asked them again for their equal-versus-efficient preferences. 193/252 responded, and 78.2% of these 193 kept the same preference.
We measured perceptions of warmth and competence to assess whether these social perceptions underpinned financial trust decisions. Observers completed twelve items from the warmth and competence subscales (Fiske, Cuddy, Glick, & Xu, 2002). We randomized the order of items and displayed four per screen; we phrased the items to compare an allocator who chose the equal allocation to an allocator who chose the efficient allocation. For example, we asked: “Who is more likely to be competent, a person who made the equal allocation or the efficient allocation?”. The seven-point scale ranged from “A person who chose equal would be much more competent – Neither would be more competent – A person who chose efficient would be much more competent.” We re-scaled the scores to range from -3 to +3. We also re-scored the warmth and competence measures to reflect perceptions of an allocator making the same allocation decision as the observer (i.e., if an observer preferred the efficient allocation, we reversed the scoring of the warmth and competence measures). In sum, these variables represent warmth and competence perceptions conditional on equal-versus-efficient allocation agreement. More positive numbers indicate a stronger preference for agreeing others, whereas more negative numbers indicate a stronger preference for disagreeing others. We hypothesized that observers who prefer equality will perceive allocators who also prefer equality as warmer than allocators who prefer efficiency. However, observers who prefer efficiency will see both types of allocators—those preferring equality and those preferring efficiency—as equally warm. In turn, we predicted that this asymmetry in perceptions of warmth will mediate the asymmetry in actual trusting behavior.

We also collected a variety of exploratory measures, the results of which we report in the supplementary materials. We asked, via an open-ended, optional response, what the observers considered as they made their trust decisions, and whether an allocator’s choosing an equal or
efficient allocation told the observers anything about the likelihood that an allocator would return money in the game. To assess the presence of any false consensus effect, we asked the participants to imagine 100 other mTurkers, and how many of those 100 would choose the same option as that preferred by the observer (we conditioned the wording on the observer’s response earlier in the survey). Observers then completed an indicator of risk tolerance adapted from Abrahao, Parigi, Gupta, & Cook (2017). To conclude, they answered whether religion was an important part of their life, along with gender and age demographic questions.

After collecting all observer and allocator data, we executed the trust game and paid bonuses based on random pairings of observers and allocators.

**Results**

**Preliminary analysis.** We found neither a main effect of Order (log-odds 95% CI [0.69, 1.20], \( z = -0.65; p = .514 \)) nor an interaction between allocator Choice and Order (log-odds 95% CI [0.74, 2.22], \( z = 0.92; p = .359 \)). Thus, we collapse across Order in subsequent analyses.

**Primary analysis—main effect of allocator choice.** Our first hypothesis regarded whether allocators inferred that equal allocators were, on average, more trustworthy. To test this, we compared the amount entrusted to allocators who favored equality with the amount entrusted to allocators who favored efficiency. Observers sent a higher proportion of their endowment to the former (\( M = 59.3\% \)) than the latter (\( M = 50.9\%; \ log-odds \ 95\% \ CI \ [1.02, 1.64], \ cohen's \ d = .21, \ z = 2.18; p = .029 \)). The effect of allocator choice on observer trust remained significant

---

21 According to our pre-registered analysis plan, we accounted for repeated measures in our data by using a multi-level regression with random intercepts for each subject (observers made trusting decisions for each equal-versus-efficient allocation the allocator could make). Further, we assumed a beta distribution to approximate the error function in our regression equations (see the appendix for a comparison of normal and beta distributions overlaid on the distribution of observer trust decisions). These analyses were conducted in R using the glmmADMB package (Fournier et al., 2012). This regression formulation also allows us to present log-odds ratios, which give an intuitive description of effect size. We use this regression set-up for observer trusting decisions only; for other measures in our data, we use standard linear regressions and t-tests.
when we controlled for observer gender, age, religiosity, risk preferences, and own preference in the equal-versus-efficient allocation in a simultaneous regression (log-odds 95% CI [1.03, 1.65], \(z = 2.21; p = .027\)).

**Moderation by observer preference.** Our second hypothesis regarded whether observer inferences depended on their (the observer’s) own preference. We predicted that, while observers who favored equality would punish allocators who disagreed, observers who favored efficiency would not (they would trust equal and efficient allocators alike). To test this, we examined an interaction term between Allocator Choice and the observer’s own preference for an equal or efficient allocation. This term was significant (log-odds 95% CI [1.28, 3.27], \(z = 2.98; p = .003\)).

To isolate the locus of the interaction, we compared the simple effects of Allocator Choice across different observer preferences. Observers who preferred the efficient allocation sent statistically equivalent amounts to equal allocators (\(M = 60.8\%\)) and efficient allocators (\(M = 64.5\%\); log-odds 95% CI [0.79, 1.54]; \(cohens\ d = .09, z = 0.59; p = .554\)).\(^{22}\) However, observers who preferred the equal decision sent significantly more to equal allocators (\(M = 58.1\%\)) than they did to efficient allocators (\(M = 39.5\%\); log-odds 95% CI [0.30, 0.98]; \(cohens\ d = .51, z = -2.02; p = .044\); See Figure 2.5).

\(^{22}\) 45.6% of participants preferred the efficient allocation; 54.4% of participants preferred the equal allocation.
Figure 2.5: Percentage of endowment transferred in trust game. The left side of the panel shows the average amounts sent by observers who favored the efficient allocation, while the right side of the panel shows the average amounts sent by observers who favored the equal allocation. The two colors depict the allocator’s possible allocation decisions. Light blue shows the amount transferred to allocators who chose the efficient allocation; dark blue shows the amount transferred to allocators who chose the equal allocation.

**Social perceptions of warmth and competence.** Our third hypothesis regarded whether asymmetric perceptions of warmth (but not competence) would statistically mediate asymmetric differences in trusting behavior. To analyze perceptions of warmth, we averaged the six items from the warmth subscale to form a composite warmth measure. As described above, these variables represent warmth and competence perceptions conditional on equal-versus-efficient allocation agreement, with more positive numbers indicating a stronger preference for an allocator who agreed with the observer.

**Perceptions of warmth.** An analysis of this composite measure showed that observers who preferred the equal allocation believed that an allocator who made the equal allocation would be warmer than an allocator who made the efficient allocation (95% CI [1.20, 1.61];
Observers who preferred the efficient allocation did not believe that an allocator who made the efficient allocation would be warmer than an allocator who made the equal allocation; if anything, they believed the opposite, that those who disagreed on the appropriate allocation would be warmer (95% CI [-.49, -.04]: t(109) = -2.31, p = .023).

**Perceptions of competence.** Our measures of competence largely mirrored, in reverse fashion, these measures of warmth. As with the warmth measure, we averaged the six items from the competence subscale to form a composite competence measure. Observers who preferred the efficient allocation perceived allocators who did the same as much more competent than allocators who made the equal allocation (95% CI [0.70, 1.10]: t(109) = 9.05, p < .001). Observers who preferred the equal allocation did not show the reverse preference; they perceived no difference in competence between the two types of allocators (statistically indistinguishable from zero: 95% CI [-0.09, .31]: t(130) = 1.11, p = .268).

**Mediation by social perceptions of warmth and competence.** To analyze the relationship between warmth, competence, and trust, we defined a new outcome measure: the difference between how much an observer transferred to an allocator with the same equal-versus-efficient preference as herself. Put another way, if an observer preferred efficiency, how much more (or less) did s/he send to an allocator who preferred the same allocation than s/he sent to an allocator who disagreed? See Figure 2.6 for a diagram of the complete model.

**Path 1—do perceptions of warmth mediate trust?** In line with our pre-registered analysis plan, we conducted a between-subjects mediation analysis, using a parallel multiple mediator model to simultaneously assess the mechanisms of warmth and competence. We find evidence for the hypothesized indirect effect of warmth. First, observers who preferred the equal allocation demonstrated a much stronger effect of agreement on perceptions of warmth than did
observers who preferred the efficient allocation (95% CI [1.46, 1.89], z = 15.31; p < .001). In turn, this difference in perceptions of warmth drove trusting behavior (95% CI [1.14, 2.49], z = 5.23; p < .001). Taken together, we find strong evidence for an indirect path: 95% CI [1.84, 4.24], z = 4.95; p < .001.

Path 2—do perceptions of competence mediate trust? We also find evidence for an opposing indirect path through competence. Specifically, observers who preferred the efficient allocation demonstrated a stronger effect of agreement on perceptions of competence than did observers who preferred the equal allocation (95% CI [-0.99, -0.59], z = -7.78; p < .001). In turn, this difference in perceptions of competence drove trusting behavior (95% CI [0.34, 1.81], z = 2.88; p = .004). Taken together, we find evidence for an indirect path: 95% CI [-1.46, -0.23], z = -2.70; p = .007. While we cannot make causal claims using these associative measures, we observed associations that indicate the possibility of oppositional mediation, in which two indirect paths have opposite effects on the dependent variable of interest.

Paths 1 and 2—overall test of the mediation model. While both indirect paths reach traditional levels of statistical significance, an examination of relative effect sizes makes clear that the indirect path through warmth outweighs the path through competence. A pairwise contrast between the two indirect effects revealed that the indirect path through warmth was approximately triple the size of the indirect path through competence (b = 3.04 vs. b = -0.85), and this difference is significantly different from zero; contrast between the two paths: 95% CI [0.65, 3.74], z = 2.78; p = .005). Thus, while analyses support the conclusion of oppositional mediation, perceptions of warmth carry far more explanatory power than do perceptions of competence in predicting the asymmetric effect of agreement on trust behavior.
We observed an oppositional mediation pattern in Study 3. Perceptions of warmth carried far more explanatory power than did perceptions of competence in predicting the asymmetric effect of agreement on trust behavior.

Secondary Analyses. For interested readers we note that observers who preferred the **efficient** allocation were more trusting overall, transferring an average of 63% of their endowment compared to an average of 49% for observers who preferred the equal allocation (log-odds 95% CI [1.03, 1.79], cohens d = .34, z = 2.20, p = .028). We attribute this to the asymmetric punishment pattern of observers who preferred equality. They viewed allocators who disagreed as lacking trustworthiness while efficient allocators viewed allocators who disagreed as trustworthy. Last, other measures (reported religiosity, gender, age, time spent on trust decision screen) had neither main nor interactive effects.

Discussion

Studies 1 and 2 showed that pressure from the social context led to more equal (but inefficient) allocations. This third study shifted the focus of the first two, asking what inferences
third-party observers draw from allocator choices. On average, observers drew harsher inferences about efficient allocators than they did about equal allocators. Crucially, however, observer’s inferences were qualified by their own views. Observers who preferred the efficient allocation trusted both types of allocators the same. Observers who preferred the equal allocation viewed only a preference for equality as a signal of an allocator’s social warmth, and in turn only trusted those allocators who chose equality.

**General Discussion**

Three pre-registered experiments (N = 1,095) extend the value pluralism model (Tetlock, 1986) to test the effect of social factors on equality-efficiency trade-offs. For allocators, pressure from the social context increased equal (yet inefficient) decisions. This was true with real money and Pareto-optimal efficiency (Study 1), and when decisions involved hypothetical life and death (Study 2). The inclusion of Pareto-optimal efficient allocations (Studies 1 and 3) created a conservative test of our hypothesis. If observation made individuals more likely to prefer equal allocations even when no resource recipients were made worse off by efficiency, then the effect of observation in more naturalistic settings, where Pareto optimality is rare (Denhardt, 2004), is likely stronger.

Further, a study of observers revealed that allocators’ reactions to social pressure were well placed. On average, observers were less likely to trust those who made Pareto-efficient allocations than those who made equal allocations. However, observer trust was conditional on their own views. Observers who preferred efficiency trusted allocators who agreed with them as well as allocators who disagreed with them. However, observers who preferred equality only trusted like-minded allocators, and did so because of underlying perceptions of warmth.
Advancing the value pluralism model. These findings extend the value pluralism model in at least two important ways. First, prior work with the value pluralism model largely focused on policy attitudes (e.g., Peterson, 1994; Tetlock, 1986; Sniderman, Tetlock, Piazza, & Kendrick, 1991). The present work extends the model to equal-versus-efficient trade-offs with real financial stakes.

Second, building on the social context postulate of the value pluralism model, we show that pressure from observers can change an allocator’s underlying value hierarchy, privileging equality relative to efficiency. Value theories propose that individual value hierarchies are fairly rigid, and that changes in values result from intrapsychic factors such as changing “definitions of the self” (Rokeach, 1979, p. 6; Leff, 1978). In contrast, we show that value hierarchies are responsive to interpersonal pressure.

Advancing research on equality-efficiency trade-offs. This research complements prior work on equality-efficiency trade-offs by showing a direct role for social observation by third-parties. Prior work has tended to focus on, broadly, how 1) individual-level characteristics and 2) characteristics of the decision setting affect equality-efficiency trade-offs. Research on individual-level characteristics, for example, has examined how preferences for equality develop in infants (e.g., Sloane, Baillargeon, & Premack, 2012), or whether equal allocations automatically attract attention over other types of allocations (e.g., Halevy & Chou, 2014). Research on characteristics of the decision setting has examined such factors as the role of reference points (Kahneman et al., 1986), different types of resources (Li, Colby, & Fernbach, 2018), and whether the decision setting allows allocators to signal impartiality to others (Choshen-Hillel, Shaw, & Caruso, 2015). To these two general categories we add a third,
characteristics of the social context, and show that observation from others biases individuals toward equal allocations.

**Advancing research on values as predictors of judgment and choice.** The findings reinvigorate value-oriented research more generally. Values have had a long and important history as a backstop of attitudes (Burroughs & Rindfleisch, 2002; Feather, 1975; Inglehart, Basanez, & Moreno, 1998; Kasser, 2016; Rokeach, 1973; Schwartz, 1992; Schwartz & Blisky, 1987). This predilection for attitudes may have led research on values to neglect implicit processes as a source of value change, focusing instead, for example, on long-term cultural determinants (e.g., Inglehart & Baker, 2000) or intrapsychic dissonance as motivators of change (e.g., Rokeach, 1979). The present work suggests that a social-contextual lens is necessary to fully understand values and how they shape consequential decisions.

The present era, or at least popular discourse about the present era, is animated by dissonance. Businesses proclaim pro-social and pro-environmental values to shareholders at the same time that they are obligated to maximize earnings on their behalf (Heryati, 2019). Voters espouse stringent conservative or liberal values at the same time that they demand compromise from their representatives (Laber-Warren, 2012). Parents demand the best for their own children at the same time that they say all children should receive an equal shot at success (Druckerman, 2019). A values-based research agenda can help navigate these and other tensions—the agenda at once recognizes that the world is a “very dissonant place,” full of trade-offs, but also that humans navigate trade-offs in systematic ways (Tetlock, Peterson, & Lerner, 1996, p. 25).

**Advancing practical ethics.** Our results also have important implications for practical ethics. Deliberative bodies regularly face decisions that include tradeoffs between efficiency and equality. Members of a field hospital may confront a choice over how to distribute precious
resources during an emergency. A school board may have to make a difficult budget decision about the distribution of educational resources. Since these choices, and plenty other similar choices, are made in a social context, not by a single individual, we can expect there to be a systematic increase in the emphasis on equality over efficiency.

To the extent we think that when facing such decisions, choosing equality over efficiency is the morally correct option, we may find nothing problematic with this. But often, it is precisely the more efficient outcome that seems, upon deliberation, to be the right choice—saving the most lives with the resources available, for instance. If pursuing efficiency is the more defensible moral outcome, deliberative bodies may find themselves mistakenly giving equality more weight than it deserves simply in virtue of the social context in which the decision is being made.

Moreover, our results suggest that deliberative bodies that end up making efficient choices could erode public trust by their decision in virtue of that choice. Effective altruists, and others committed to emphasizing best outcomes, need to take this into account with their recommendations. Public groups that choose to pursue the greater good may very well trust in them by doing so. This is a nontrivial concern, since public trust in many places is already at an all time low. And regardless, a loss of trust is rarely brought into a utilitarian calculation.

**Limitations and future directions.** The present work explored a handful of critical moderators—allocation type, for example (Study 2), and demographic differences (all studies). Future work could explore other important moderators of the effects of the social context. As one example, future work could explore the role of status, a fundamental human motive (Anderson, Hildreth, & Howland, 2015), by examining whether the effects of observation differ according to observer or allocator status. Our own pilot work suggests that allocator status is worth
exploring—in a convenience sample drawn from executive education programs in a public policy school in the American Northeast, we did not find a relationship between concerns for public trustworthiness and the efficiency of organ allocations, as we did for lay decision makers (Study 2). The difference across lay and executive decision makers seemed to result from a positive relationship between concerns for public trust and efficiency. Like lay allocators, executives showed the same pattern of devaluing efficiency relative to equality. Executives also, however, associated public trustworthiness with efficiency, in this case saving as many lives as possible. Future work could explore whether this moderation by allocator status is true across different types of backgrounds (i.e., business, public policy), different types of allocations (i.e., organs in a life-or-death allocation, money, time), and different types of observation (i.e., primes for concerns over public trust, direct observation).

The results could be expanded by studying the effect of observation in a wider variety of allocation decisions. For example, future studies could introduce variation in the characteristics of resource recipients. Prior work has shown that judgments of resource recipients’ role in causing their need for resources partially determine allocation decisions, with individuals perceived as responsible for their lack of resources less likely to receive aid than individuals whose lack is perceived to be due to factors outside their control (Skitka & Tetlock, 1992). In our studies, we held constant perceived agency in causing resource need. Given the possibility of interactions with third-party observation, future studies could vary perceived agency in causing resource need.

Other topics for exploration include the extent to which the effects of observation are durable over time, and whether interaction between allocators and observers moderates the effects of observation. The present work showed that value hierarchies shift in response to
observation—future work could explore how stable that shift is over time. Future work could also explore interaction—in the present studies, allocators could only signal to observers through the choices they made and were not able to signal in other ways, like providing a rationale. Finally, future work could explore whether the opportunity to give a rationale makes efficient allocations more likely by bolstering their moral reputations (Trivers, 1971), perhaps by allowing allocators to claim that their motives are rooted in improving human welfare, or whether rationales make equal allocations more likely, perhaps because equality is easier to justify (Adelberg & Batson, 1978).

**Summary.** When Vice President Joseph Biden insisted that budgets, not statements, reveal one’s true values, he recognized that candidates’ proposed budgets were a means for prospective voters to evaluate the character of the candidates vying for the highest office in the United States. In more mundane settings, allocator reactions to observers play a critical role in trade-off choice. As a vice presidential candidate or a regular citizen, value trade-offs rarely occur in a “social vacuum” (Tetlock, 1991, p. 454). Understanding how humans navigate trade-offs requires attending to the crucial role of the social context in shaping judgment and choice.
References


Essay 3

“I Was First and I Was Right”:
The Effects of Task Order on Evaluations of Peer Judgments
with Julia A. Minson

Research on quantitative estimation offers a gold standard for how individuals should approach group judgment and decision making. To maximize judgment accuracy, collaborators should begin by making independent assessments, and only later combine them with those of other group members, lest social influence cause estimates to assimilate toward each other and decrease the “wisdom of crowds” (Galton, 1907; Lorenz, Rauhut, Schweitzer, & Helbing, 2011; Minson, Mueller, & Larrick, 2017; Surowiecki, 2004). A tacit assumption behind this recommendation is that independent judgments can later be seamlessly aggregated to form a single product that reflects the consensus of the group.

Prior to aggregating, though, group members often evaluate each other’s judgments, and their evaluations may affect how much weight a judgment receives in the final group product. Evaluators may assign weights according to a number of factors, including work history (Barnes, Reb, & Ang, 2012), access to specific information (Budescu & Rantilla, 2000), and expertise (Goldsmith & Fitch, 1997). In the current research, we test the influence of an additional factor, one that should arguably play no role in evaluations of collaborative judgments: the sequential ordering of committing to one’s own judgment versus evaluating another’s.

Across seven experiments, we manipulated the order in which research participants committed to a judgment or decision of their own versus evaluated a judgment or decision produced by a peer. We consistently found that the order endorsed by prior research—offering one’s own estimate prior to evaluating peer input—led individuals to derogate the judgments and decisions of others, and assess them as being of lesser quality than participants who evaluated
peer input prior to offering their own judgment or decision. We observed this effect in a variety of estimation domains and tasks, with both lay and expert samples, and documented that the effect is driven by systematic differences in the amount of disagreement that emerges as a function of task order. Our research demonstrates that turning to the wisdom of crowds comes with an interpersonal catch—fellow crowd members appear less wise to those who have committed to an opinion of their own.

**Prior Literature on the Effects of Judgment Order**

Independent judgments constitute the gold standard for group accuracy, *provided that* the final group judgment represents an approximately equal weighting of the relevant inputs (Clemen & Winkler, 1986; Einhorn & Hogarth, 1975; Hogarth, 1978; Sniezek & Henry, 1990; Soll & Larrick, 2009). Equal weighting increases the likelihood that individual errors will cancel each other out (Larrick & Soll, 2006; Lorenz et. al., 2011; Soll & Larrick, 2009), and under most conditions outperforms attempts to identify and give priority to more accurate judgments. If group members are responsible for both generating the judgments and the weights assigned to those judgments, however, the benefits of independence rest on an assumption of unbiased evaluation. Yet, few studies of group judgment have addressed how members of groups evaluate each other’s contributions, instead treating the accuracy resulting from different aggregation approaches as the focal outcome (for a review, see Gigone & Hastie, 1997).

A related literature that indirectly addresses evaluation of peer judgment is work on the Judge Advisor System (JAS). The most common finding in this body of work is that judges underweight the advice of others relative to their own judgment, and thus relative to the normatively appropriate benchmark of equal weighting (Harvey & Fischer, 1997; Yaniv & Kleinberger, 2000; for a review, see Bonaccio & Dalal, 2006). Specifically, individuals typically
adjust 20%-30% of the distance between their own and others’ estimates, effectively treating their own judgments as more than twice as accurate as those of peers (Harvey & Fischer, 1997; Soll & Larrick, 2009). Explanations for this under-weighting of other’s inputs include differential access to reasons—judges have access to their own reasons for an estimate but not other’s reasons (Yaniv & Kleinberger, 2000)—and “egocentric bias”—individuals may believe that they are simply better judges than those from whom they are receiving advice (Krueger, 2003). The latter explanation is related to ones that invoke “epistemic vigilance” (Sperber et al., 2010; Trouche et al., 2018) in that both portray individuals as believing they have better judgment than others. Explanations invoking vigilance, however, distinguish themselves by arguing that discounting is not a bias, but a functional strategy because it helps individuals guard against being accidentally or intentionally misinformed (Sperber et al., 2010). While these explanations differ on mechanism and ultimate cause, together they paint a clear picture of individuals favoring their own judgments over those of peers.

Despite these robust findings, the advice taking literature cannot address the question of how evaluations of others respond to different orders of judgment and evaluation for two reasons. First, the JAS paradigm always requires an individual to make a tradeoff between own and other’s judgments, thus making it difficult to establish whether people are favoring the self or discounting the other. Second, task order is usually fixed. Individuals almost always begin by making their own estimates, after which they are exposed to those of a peer.

The few studies that have varied judgment order have produced mixed results. Rader, Soll, and Larrick (2015), for example, found that individuals who formed an independent judgment prior to receiving advice were more likely to accept advice than judges who saw advice before making judgments of their own. The judges who first saw advice “pushed away”
from it in forming their own judgments, suggesting that rendering one’s own judgment before evaluating that of a peer may lead to more favorable peer evaluations. In contrast, Yaniv & Chosen-Hillel (2012) found that judges who estimated first were less likely to accept advice than judges who did not first make an estimate of their own, implying the opposite effect on evaluations. Other findings from studies that manipulated order in an advice-taking paradigm are inconclusive. Sniezek & Buckley (1995), for example, found no difference in the rate of advice taking between those who made their own judgments before seeing advice and those who saw advice first.23

In the present work we experimentally manipulate the order in which individuals commit to their own judgments and decisions versus evaluate peer inputs. By varying task order while holding peer input constant, we address questions of theoretical and practical value. Theoretically, we examine how the mere formation of judgment affects interpersonal evaluations, an outcome ordinarily studied when individuals have already formed opinions of each other or the stimulus in question (e.g., Ross, Lepper, & Ward, 2010). Practically, task order represents a lever that decision makers and organizational leaders can easily manipulate in order to improve the effectiveness of collaboration.

**Anchoring, Disagreement, and Naïve Realism**

We generate our predictions regarding the effects of judgment order on evaluations of peer input by bridging two classic bodies of research in judgment and decision-making and social psychology. First, we draw on work on the phenomenon of anchoring and insufficient adjustment (Tversky & Kahneman, 1974) and its likely effects on disagreement in the context of

---

23 Sniezek & Buckley (1995) do find an effect of ordering on the use of what they call the “confidence utilization” strategy, or taking the advice of the advisor expressing the highest degree of confidence in her/his judgment. They find that individuals who first make judgments of their own are slightly less likely to use this strategy than individuals who first see advice. It is unclear how, if at all, this measure maps onto advisor evaluations.
collaborative judgment. Second, we draw on theory and research on “naïve realism” (Ross, Lepper, & Ward, 2010) in order to make predictions about how individuals are likely to interpret such disagreement.

Prior research has demonstrated that estimates under uncertainty are systematically influenced by seemingly irrelevant quantities (“anchors”), that are cognitively available at the time of making a judgment. A large literature has explored the underlying causes and boundary conditions of anchoring effects (Epley & Gilovich, 2001; Epley & Gilovich, 2006; Frederick, Kahneman, & Mochon, 2010; Frederick & Mochon, 2012; Janiszewski & Uy, 2008; Loschelder, Friese, Schaerer, & Galinsky, 2016; Mochon & Frederick, 2013; Simmons, LeBoeuf, & Nelson, 2010; Tversky & Kahneman, 1974), with the primary outcome of interest being the extent to which the focal judgment assimilates to the anchor. Although the vast majority of research on anchoring has focused on individual judgment, there are clear implications for judgments that are made collaboratively. Specifically, to the extent that anchoring leads judgments to be closer to each other, it must also be the case that truly independent judgments are going to be further apart. Thus, decision-makers following the now accepted wisdom of generating independent judgments, are likely to experience a greater level of disagreement between their judgments than decision-makers whose judgments assimilate toward each other.

Disagreement in group judgments is not in itself a negative (Janis, 1972). In fact disagreement in a particular form underpins the superior accuracy of group versus individual judgments: in most cases, estimates which diverge from each other are more likely to err on opposite sides of the truth and cancel each other out when aggregated (Larrick & Soll, 2006). This effect, however, depends on an approximately equal weighting of judgments, and the assumption that individuals will not systematically change their evaluations of others’ inputs as a
function of disagreement. In the present work, we draw on research on “naïve realism” (Robinson, Keltner, & Ward, 1995; Ross et al., 2010; Ross & Ward, 1996), to predict that judges will interpret disagreement in a self-serving manner, and consequently decrease their evaluations of others’ inputs.

Several research streams have demonstrated that individuals grant their own perceptions of reality a privileged place relative to the perceptions of others—i.e. they are “naïve realists” who assume that their subjective evaluations and judgments form an accurate representation of the world around them (Griffin & Ross, 1991; Ross et al., 2010). As a consequence, people derogate others who disagree with them on political and social views (Kunda, 1990; Pronin, Gilovich & Ross, 2004), and even matters of taste (Blackman, 2014). Similarly, individuals evaluate the merit of scientific findings as a function of whether they conform to their prior beliefs (Kahan, Peters, Dawson, & Slovic, 2017; Kahan et al., 2012; Lord, Ross, & Lepper, 1979). And in the domain of quantitative judgment, several studies have demonstrated that people take less advice after exposure to estimates that are very different from their own, attributing dissimilarity in estimates to the flawed judgment of others (Liberman, Minson, Bryan, & Ross, 2011; Minson, Liberman & Ross, 2011).

Hypotheses

The two research streams on anchoring and on naïve realism inform our hypotheses regarding the effect of task order on evaluations of peer judgments and the attributions we make about the peers offering those judgments. In line with the anchoring literature, we predict that participants who make their own estimates after evaluating the estimate of a peer will make estimates that are closer to that peer’s estimate than participants who made independent estimates before offering an evaluation of a peer’s estimate. Such assimilation toward the target
estimate will, in turn, result in participants observing lower levels of disagreement between their own estimates and the estimates of others.

By implication, participants who first committed to their own judgments and decisions (as recommended by the extant research literature), would observe greater disagreement between their own estimate and those of a peer. This difference in disagreement, in turn, leads to our second and focal prediction. Drawing on work on naïve realism, we predict that individuals will assess their own judgments as representing the most reasonable estimate to be made under the circumstances, and will assess peer judgments negatively to the extent that those judgments disagree with their own.

Our hypotheses extend the research literature in several ways. We contribute to the literature on collaborative judgment and decision-making by highlighting the potentially negative consequences of the classic advice to initiate group tasks with independent judgments. The prescriptive advice that the prior literature has offered based on concerns with judgment accuracy may be incomplete in light of the additional consequences of temporal ordering for judgment evaluation. Also, the key questions examined by research on naïve realism revolve around how individuals evaluate others’ positions relative to their own. As such, naïve realism research does not make predictions regarding situations when an individual has not yet formed their own opinion. By contrast, research on anchoring, while addressing how yet unformed judgments are affected by the presence of anchors, does not speak to how assimilation affects other down-stream consequences, such as evaluations of other actors and their inputs. The integration of these two bodies of research allows us to offer and test a novel prediction about an important set of social processes.
The Present Research

We present the results of seven studies that examined the effect of task order on evaluation of peer inputs in judgments and decision-making tasks across a range of domains. In Study 1 we tested our basic hypothesis and found that, in the domain of consumer judgments, individuals evaluated the accuracy of a peer’s estimate more negatively when they had first committed to their own estimate. In Studies 2A and 2B, we explicitly tested disagreement as a mediating mechanism and moderating condition for our effect. In Study 3 we explicitly tested our predictions based on naïve realism, and showed that the effect of ordering appears to be supported by a self-serving interpretation of disagreement. In Study 4 we expanded beyond the domain of quantitative estimation tasks and demonstrated our phenomenon in a complex medical decision-making scenario. Finally, in Study 5 we tested the effect of commitment to one’s own decision on the willingness of both novices and national security experts to follow a particular course of action in light of identical intelligence. In keeping with best practices for fully-reproducible science (Simmons, Nelson, & Simonsohn, 2012), we report all methodological decisions (e.g., determining sample size), manipulations, and measures.

Study 1

Study 1 provides an initial test of our phenomenon. Participants produced a simple consumer estimate (the lifetime cost of owning a dog), and evaluated the accuracy of an estimate ostensibly produced by a peer. Importantly, the target estimate we used was in fact an expert answer to the estimation task. We tested whether simply committing to one’s own estimate before evaluating another’s estimate changed participants’ assessments of the target estimate. In addition to examining task order, we varied the level of effort associated with producing an estimate. This comparison of effort allowed us to test whether our effect requires participants to
engage in involved, deliberate processing (as would happen in most professional contexts, or whether it would also emerge when people offer quick intuitive judgments.

**Method**

**Participants**

We determined our sample size by doubling the cell size from prior pilot studies (see Supplementary Materials) in order to detect any potential interaction between our variables. We recruited 808 volunteers ($M_{age} = 44$, 59% female) from the Harvard Digital Lab for the Social Sciences (DLABSS), a forum for unpaid volunteers who wish to contribute to social science research.\(^{24}\)

**Design & Procedure**

**Design.** Our study employed a 2 X 2 between-subjects design. We manipulated task order such that participants did or did not commit to their own estimates prior to evaluating the focal estimate. We crossed these two levels of having committed to one’s own estimate (Committed, Uncommitted), with two methods of generating the estimate (following a structured 7-step process vs. making a single intuitive estimate).

**Procedure.** In this and all subsequent studies we obtained informed consent at the start of the study procedure. After informed consent, participants answered a demographic questionnaire (a precondition of volunteering in the DLABBS subject pool). We then randomly assigned participants either to give their own estimate of the lifetime cost of an average dog and then to evaluate another estimate (Committed treatment), or to evaluate the target estimate without committing to their own judgment (Uncommitted treatment). We also randomly assigned participants either to use (and/or evaluate the result of) a systematic seven-step process for

\(^{24}\) [http://dlabss.harvard.edu/about/]
making the estimate, or simply make and/or evaluate an intuitive estimate. The process used to make the estimates by the participants in the Process conditions is presented in Appendix A.

After making their estimates and/or evaluations, we then asked participants whether they were dog owners (which we defined as owning a dog in the past 5 years), along with a series of questions designed to measure how important this domain of knowledge was to them (i.e., “how important is knowledge about dog ownership to you?”). We recorded responses to these domain-importance questions using a 5-point Likert scale (1: “Not at all” — 5: “Very much”).

**Results**

**Commitment.** In line with our predictions, participants who committed to their own estimate prior to evaluating the target estimate judged that estimate as less likely to be accurate ($M_{\text{committed}} = 39.04$, $SD = 26.98$), than participants who did not commit to their own estimate ($M_{\text{uncommitted}} = 45.34$, $SD = 26.67$), 95% CI[-10.01, -2.60], $t(806) = -3.34$, $p < .001$). The effect of Commitment was negative whether participants evaluated a peer who used a structured 7-step process (95% CI[-12.25, -1.62], $t(391) = -2.57$, $p = .011$), or made an intuitive judgment (95% CI[-10.80, -0.47], $t(413) = -2.15$, $p = .033$).

**Process vs. Intuition.** There was a positive main effect of using a process. Participants rated the accuracy of estimates derived via the 7-step process as being higher than the accuracy of estimate derived via intuition ($M_{\text{process}} = 44.10$, $SD = 26.98$; $M_{\text{intuition}} = 40.41$, $SD = 26.92$), 95% CI[0.02, 7.46], $t(806) = 1.97$, $p = .049$. When we simultaneously regressed our outcome variable on both factors and their interaction, we observed a significant negative effect of committing to one’s own estimate on evaluations of the target estimate (95% CI[-10.80, -0.47], $t(804) = -2.14$, $p = .032$), and a marginally significant positive effect of following a process (95% CI[-0.88, 9.52], $t(804) = 1.63$, $p = .104$). Interestingly, the magnitude of the effect size for Commitment
was nearly 70% larger than the effect size for Process ($d_{\text{committed}} = -0.24, 95\% \text{ CI}[-0.37, -0.10]$, $d_{\text{process}} = 0.14, 95\% \text{ CI}[-0.005, 0.29]$). The mere fact of committing to one’s own estimate led to a larger change in participants’ evaluations of a peer’s contribution than whether that peer used a thoughtful process or simply guessed the answer. There was no interaction between the two variables ($95\% \text{ CI}[-8.70, 6.10], t(804) = 0.35, p = .730$).

![Figure 3.1](image1.png)

*Figure 3.1.* The vertical axis represents the judged likelihood that the target response was within 10% of the correct answer. Committing to a judgment of one’s own led to lower evaluations of a peer’s estimate (left panel). The effect of commitment held whether a participant engaged in (or observed the use of) an in-depth seven-step process to generate a judgment. Bars represent standard errors.

We were also able to examine the effect of disagreement on participants’ evaluation of the target estimate in the two Committed conditions. Participants evaluated the target estimate more negatively as a function of the extent to which that target estimate deviated from their own estimate ($95\% \text{ CI}[-19.70, -13.93], t(396) = 11.45, p < .001$), lending some initial support to our theorizing.

**Discussion**

In Study 1 participants who had committed to their own estimate rated a target estimate less positively than participants who rated the exact same target estimate, produced in the same

---

25 A robust regression (Rousseeuw et al., 2015) also showed a significant result ($p < .001$).
manner, but without having committed to an estimate themselves. This effect held for both intuitive “snap” judgments as well as judgments derived through a step-by-step process. A comparison of effect sizes showed how important task order can be to evaluations of others’ judgments. The mere act of committing to a judgment had a 70% greater effect on evaluations than the difference between following a seven-step process relative to following one’s intuition.

We did not find evidence that domain importance or expertise moderated the effects of commitment, suggesting that explanations related to a motivation to defend one’s skill in personally important domains do not account for the effects of ordering. Suggestively, the amount of disagreement with the target estimate in the two committed conditions strongly predicted target evaluations. We systematically test the role of disagreement in Studies 2A and 2B.

**Studies 2A & 2B**

We next examine the role of disagreement in the evaluation of peer judgments by randomly assigning participants to judgment tasks where high levels of disagreement are more or less likely. We also broaden our investigation to include different stimuli: a new consumer estimate, as well as an estimate based on visual input.

**Method**

**Participants**

We recruited 424 adult participants for Study 2A \(M_{age} = 35, 53\% \text{ female} \) via mTurk. We offered participants $.70 for survey completion. The stimuli in this survey dealt with the costs of childrearing. We directed the study advertisement at parents in order to recruit participants with some domain expertise. At the end of the survey, we asked participants whether they were in fact parents, making it clear they would not be penalized for answering truthfully if they were not.
participants stated that they were not parents, leaving us with a final sample of 381 ($M_{age} = 35$, 55% female). In Study 2B we recruited 1,211 participants ($M_{age} = 38$, 49% female) via mTurk for a five-minute study on estimation. Participants were offered $.50 with no bonus opportunities.

**Design & Procedure**

**Design.** Our next two experiments featured a 2 (Committed, Uncommitted) X 2 (Agreement, Disagreement) factorial design. We report them as Studies 2A and 2B because they followed the same design but required participants to make judgments in different domains.

**Procedure.** Participants in both studies first stated how much they “trusted their own judgment” in estimating either the costs associated with childcare or estimating numerical quantities. For Study 2A, we also asked how knowledgeable participants believed they were about the costs associated with childcare.

Participants in both studies read that fellow survey takers made the same estimates that participants would see on subsequent screens. Participants in the committed condition read: “After making your own estimates, we would like you to evaluate the likelihood that another participant's estimate is correct” (emphasis not present in survey instructions). Participants in the uncommitted condition were told that they would evaluate the likelihood that another participant’s estimate was correct “Before making your own estimate.” Participants in both conditions then went on to make their own estimates and evaluate a target estimate, with the order of the tasks determined by condition assignment.

We varied the level of disagreement by asking participants to make judgments of relatively small or large quantities, with judgments of large quantities meant to create more disagreement between participant’s and the targets they evaluated. In the Agreement condition of
Study 2A, participants estimated the average *monthly* cost of raising a child, while participants in the disagreement condition estimated the *total* cost of raising a child from birth to age 18. For the target response, we used an estimate calculated by CNN Money ($1,145 monthly; or a total of $233,610; Vasel, 2017). These two versions of the question ensured that, on average, participants making estimates about monthly costs observed smaller discrepancies between their own estimates and the target estimate than participants making the total cost estimate.

We elicited evaluations of the target estimate in the Disagreement (Agreement) condition by asking participants how likely they believed it was that the target estimate was within $20,000 ($100) of the correct answer. We recorded these evaluations on a five-point Likert scale, ranging from 1: “Not at all likely” to 5: “Very likely.”

In Study 2B, the two conditions respectively asked participants to estimate the number of M&Ms in a container holding 22 M&Ms or 3,791 M&Ms. After making their own estimate, committed participants saw the target estimate and then stated how likely they believed it was that the target estimate was within 10% of the right answer.

*Figure 3.2. Agreement (left) and disagreement stimuli for Study 2B.*
Participants in both studies then stated how much they agreed with the target estimate, which we again recorded on a five-point Likert scale (1: “Did not agree at all” – 5: “Agreed completely”). We ended the study by collecting basic demographic data.

**Results**

**Study 2A.** In line with our prior results, participants who estimated lifetime childrearing costs (Disagreement condition) rated their partner’s estimate less positively after having made their own estimate ($M_{uncommitted} = 2.88, M_{committed} = 2.53; 95\% CI[-0.68, -0.03], t(210) = -2.15, p = .032$). By contrast, when participants estimated the monthly costs of childrearing (Agreement condition), the effect of committing to one’s own estimate was no longer significant, and in fact in the opposite direction ($M_{uncommitted} = 2.49, M_{committed} = 2.66; 95\% CI[-0.18, 0.53], t(167) = 0.935, p = .351$). As a result, we observed a statistically significant interaction between commitment and level of agreement ($95\% CI[-1.01, -0.04], t(377) = -2.138, p = .033$), suggesting that likelihood of agreement moderated the effect of commitment.

Unlike in Study 1, participants in all conditions of Studies 2A and 2B gave a judgment of their own, with uncommitted participants giving their judgment after evaluating the target. With judgments for both conditions, we were able to test whether objective agreement—defined as the absolute difference between a participant’s judgment and the target judgment—mediated the effect of commitment on evaluations of the target. Using the “mediation” package in R (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014), we simulated the average causal mediating effect (ACME) of disagreement. We observed a significant, negative mediating effect for distance between one’s judgment and the judgment of the target ($95\% CI[-0.11, 0.00], p = .044$).

**Study 2B.** We observed the same moderation and mediation pattern in Study 2B. As in Study 2A, the effects of Commitment were significant in the Disagreement condition ($M_{uncommitted}$
= 2.17, $M_{\text{committed}} = 1.98$; 95% CI[-0.35, -0.05], $t(619) = 2.57, p = .011$), but not the Agreement condition ($M_{\text{uncommitted}} = 3.28, M_{\text{committed}} = 3.29$; 95% CI[-0.16, 0.17], $t(588) = -0.069, p = .945$).

The interaction between the two treatments was marginally significant (95% CI[-0.42, 0.02], $t(1207) = 1.767, p = .078$). A mediation analysis again showed a significant, negative mediation effect for objective disagreement (95% CI[0.17, 0.36], $p < .001$).

![Figure 3.3](image)

**Figure 3.3.** Judged likelihood that the target response was within 10% of the correct answer, measured using a five-point Likert scale (vertical axis). Each panel shows different effects for commitment depending on level of agreement. When agreement was likely, committed participants were statistically indistinguishable from uncommitted participants, whether the question domain was the costs of raising a child (left panel) or the number of M&Ms in a container (right panel). When agreement was unlikely, committed participants evaluated peers lower than did uncommitted participants. Bars represent standard errors.

**Discussion**

Studies 2A and 2B extended our investigation by manipulating the likelihood of disagreement orthogonally to commitment. In line with our theorizing, the effect of commitment only emerged on estimates where participants were likely to disagree with the estimate they were evaluating. In cases where disagreement was unlikely, commitment had no effect on evaluations.

In order to examine how much disagreement is required to obtain the effect of commitment we re-ran Study 2B with several levels of likelihood of disagreement by asking participants to estimate the number of M&M’s in jars of seven different sizes. We report the full methods and results of this study in the supplementary materials. Importantly, the effect of
commitment was significant at six of the seven levels of disagreement, implying that the negative effects of commitment require minimal discrepancy between own and other’s judgments.

To briefly review the findings so far, Study 1 showed that an ordering of judgment before evaluation systematically leads to lower evaluations than an ordering of evaluation before judgment, and that the effect seems to emerge irrespective of level of domain expertise or effort invested in making the judgment. Studies 2A & 2B showed that the lower evaluations result from the increased levels of disagreement that occur when judges make estimates of their own before evaluating, and in turn that the effect of commitment is moderated by the likelihood of disagreement.

**Study 3**

In Study 3 we investigate whether participants’ treatment of disagreement as a negative signal of the quality of the target estimate is normatively appropriate. Given that the only information participants possess at the time of making the evaluation are the two estimates, it may seem reasonable to evaluate the target as a function of how far it deviates from one’s own judgment. However, if that is the case, the same standard of accuracy should apply to both self and other: to the extent that disagreement points to the difficulty of the judgment and the potential presence of error, individuals should evaluate both their own judgments as well as the judgments of peers more negatively. In the words of Dawes and Hastie (2001), “interjudge disagreements tell us someone is wrong and undermine our confidence in all judgments” (emphasis added, p. 54). If, instead, only peer evaluations suffer while evaluations of own judgments remain intact, we can conclude that individuals are privileging their own judgments, in line with predictions of naïve realism.
Our studies to this point are open to two different interpretations of how participants are reacting disagreement. Committed judges may have reported lower evaluations of others’ judgments while also mentally lowering evaluations of their own judgments, and thus interpreting disagreement as a signal that either judgment may be off the mark. Or, they may have reported lower evaluations because they interpreted disagreement to mean that the other judgment is less likely to be accurate than their own. We designed our third study to differentiate between these two approaches by having judges rate themselves and others after encountering varying levels of disagreement.

**Method**

**Participants**

We pre-registered a sample size of 400 prior to exclusions. We recruited 401 participants through mTurk (\(M \text{ age} = 38\), 50% female). Participants received $.30 for completing the study and had the possibility of receiving a $.50 bonus for if their estimate fell within 10% of the correct answer. According to our pre-registered criteria, we first excluded participants who did not correctly answer all comprehension questions (\(n = 28\)). We then excluded participants who gave extreme estimates, pre-defined as the bottom and top 10% of the distribution of judgments in the current study (\(n = 71\)). This left a final sample of 302 (\(M \text{ age} = 38, 53\% \text{ female}\)).

**Design & Procedure**

**Design.** Our design employed three independent variables. Within-subjects, participants evaluated their own judgment and the other’s judgment (Focus of Judgment: Self, Other ). We counterbalanced the order in which participants evaluated their own and another’s judgment. To these factors we added a third, approximately continuous treatment for disagreement. We created this variable by randomly selecting an estimate from a previous pool of estimates and calculating
the absolute value of the difference between this estimate and the participant’s own estimate
(more details follow in the Procedure section).

Procedure. After informed consent we told participants that they would make one
estimate of the number of M&Ms in a jar, a picture of which they would see on the next screen
(see the large container of M&Ms in Figure 3.2), and that they would receive a $.50 bonus if
their estimate was within 10% of the correct number. We administered three basic
comprehension questions based on these instructions (i.e., “How many estimates will you
make?”). Following these questions, participants read that they would see the estimate of another
participant. Participants then read: “After you see the estimate of the other participant, you will
evaluate the likelihood that your [own estimate/the other’s estimate] AND [the other’s
estimate/your own estimate] is correct.” We counterbalanced the order in which participants
evaluated their own versus the other participant’s estimate. Participants then estimated the
number of M&M’s in a container.

On a separate screen, participants saw a reminder of their own estimate and the estimate
of another participant (the order of presentation was again counterbalanced). We built the pool of
other’s estimates (664 in total) by culling the middle 80% of estimates from previous studies that
had used the same stimuli and asked participants to make judgments of their own before seeing
the judgment of another. This allowed us to expose each participant to a randomly selected level
of disagreement. Participants then rated their own judgment and the judgment of the other, or the
reverse, by indicating how likely it was that the estimate was within 10% of the correct answer.
We recorded responses on a five-point Likert scale with options ranging from “Not at all likely”
(1) to “Very likely” (5). The next screen concluded the study with questions regarding
participants’ gender and age.
Results

Overall, participants evaluated their own estimates more positively than they evaluated those offered by others (95% CI[0.39, 0.67], \( t(301) = 7.37, p < .001 \)). We tested our question of interest by comparing the relationship between disagreement and evaluation across self- and other-evaluations. If judges penalized themselves at the same rate as they penalized others, we would expect to see approximately the same (negative) relationship between disagreement and evaluation for both self- and other-judgments. If, however, judges interpreted disagreement in a self-serving way, we would expect to see a negative relationship between disagreement and evaluation for other-evaluations but not self-evaluations, leading to a significant interaction between focus of judgment and amount of disagreement.

In line with our predictions, we observed a clear negative relationship between the level of disagreement and evaluations of others’ judgments (95% CI[-7.0*10^-4, -3.4*10^-4], \( t(300) = -5.64, p < .001 \)), but no relationship between the amount of disagreement and one’s own judgments (95% CI[-2.4*10^-4, 0.89*10^-4], \( t(300) = -.89, p = .376 \), and a significant interaction between level of disagreement and focus of judgment (95% CI[2.3*10^-4, 6.5*10^-4], \( t(300) = 4.16, p < .001 \); see Figure 3.4). It seems that individuals interpreted disagreement to mean that the other individual whose judgment they were evaluating was incorrect, and did not apply this level of scrutiny to their own judgments.

26 We used a hierarchical linear model to account for the lack of independence in our repeated-measures design (self, other). The order in which participants evaluated their own versus another’s judgment had no main effect on evaluations, nor did it interact with our other variables.
Participants judged the likelihood that either their own (blue) or another’s (gold) estimate was correct on a scale that ranged from 1 - 5 (vertical axis; points above and below 1 and 5 are due to jittering the display). As disagreement increased (horizontal axis), the differing slopes suggest that participants interpreted disagreement as a sign that the other’s estimate was incorrect, rather than as a sign that both own and other’s estimates were less likely to be correct.

Participants still may have judged normatively if their penalties only applied to inaccurate others. If ratings were based on accuracy, we would observe a negative correlation between one’s evaluation of others and the other’s distance from the truth. Ratings would decrease as inaccuracy increased. Instead, we observed the opposite. Evaluations of other’s judgments were positively correlated with the distance the other judgment deviated from the truth ($r$ 95% CI [.022, .244], $t = 2.37$, $p = .019$).

**Discussion**

Self-serving interpretations of disagreement underlie the effects of order on evaluations of peer judgments. In line with research on naïve realism, participants viewed higher levels of disagreement as an indicator that the other’s judgment was more likely to be wrong, rather than
as an indication that the other’s judgment and their own judgment were both more likely to be wrong.

**Study 4**

Study 4 examines our phenomenon in a different context, a medical domain laden with ethical considerations. Specifically, Study 4 asked participants to make and evaluate choices and supporting reasons for which of several deserving individuals should receive a life-saving kidney. As with many real-world decisions, there was no clearly correct decision. Because there is no way to evaluate the “accuracy” of such a choice, our dependent variable changed to an evaluation of the overall quality of a particular course of action, as well as evaluation of the quality of reasoning behind it.

Although in many real-world situations individual decision-makers have no way of knowing a priori which choice or which judgment is best, groups must still choose and execute a plan. In such cases, the individual who proposes the plan that is ultimately chosen often accrues reputational credit. To simulate this dynamic, we told Study 4 participants that a new set of participants, whom we referred to as “supervisors,” would make a choice of their own in the scenario. If the supervisor made the same recommendation as the participant, the participant would receive a bonus.

**Method**

**Participants**

We aimed for a sample size of 400 and successfully recruited 399 participants through mTurk (\( M \text{age} = 35, \text{46\% female} \)). Our recruitment message told participants that they would “make decisions about ethical dilemmas,” and that the survey would involve reading and writing.
Compensation for the study was $1.00. We offered bonus payments of $.50 if the supervisor’s favored option matched that of the participant.

**Design & Procedure**

**Design.** We employed a single-factor, two-level (Committed, Uncommitted) design.

**Procedure.** We adapted materials from “The Kidney Case” (Austen-Smith, Feddersen, Galinsky, & Liljenquist, 2014), a simulation designed for teaching about biases in ethical decision-making. Participants took on the role of members of a Kidney Transplant Review Board. Their task was to decide the allocation of one kidney among four deserving candidates (we simplified the task from the eight candidates presented in the original exercise). Each description of the four transplant candidates offered a compelling reason for being selected as the kidney recipient (e.g. one candidate was a veteran, another a single parent, another a philanthropist, etc.). The complete descriptions of the candidates are presented in Appendix B.

We told all participants that their recommendations would be paired with another participant in the survey, and that, in a separate survey, a fellow mTurk worker would play the role of “supervisor” and evaluate the pair’s recommendations. After evaluating the recommendations, the supervisor would make a recipient recommendation of her/his own. If the supervisor made the same recommendation as the participant, the participant would receive a bonus. Importantly, the bonus did not depend on whether the supervisor agreed with the other mTurker. We specified that the supervisor would not see the participant’s evaluations of the target, but only the recommendations provided by both.

We again randomly assigned participants to conditions (Committed versus Uncommitted), which determined whether or not they made their own kidney allocation decision prior to evaluating the decision of another participant. In the Committed condition, participants
read the four candidate profiles, and then selected a single candidate to receive the kidney. Participants did not rank the remaining candidates. After making their selection, we asked them to write a few sentences to explain their choice.

Committed participants then saw the choice ostensibly made by another mTurker, along with a brief explanation for that decision. In reality, participants were randomly assigned to see one of the four transplant candidates. This ensured that by chance, 25% of participants evaluated a target who chose the same candidate as they did, and 75% evaluated a target who made a different choice.

We asked participants to consider how similar the target’s answer and reasoning were to their own using a 7-point Likert scale anchored at 1: “Not at all” and 7: “Extremely.” We then asked participants whether they would support the target’s choice for the kidney using a 7-point Likert scale anchored at 1: “Strongly disagree” and 7: “Strongly agree.” We also asked participants a series of questions evaluating the target’s choice in terms of whether it was the best from the perspective of being intelligent, thoughtful, ethical, and moral. These four questions were again elicited on 7-point Likert scales, anchored at 1: “Strongly disagree” and 7: “Strongly agree.” The four items showed high reliability (Cronbach’s $\alpha = .93$), and we combined them into a composite rating representing the participants’ overall evaluation of the target’s choice.

Participants in the Uncommitted condition engaged in the exact same tasks, though in a different order. They viewed the four candidate profiles, and then, rather than choosing a candidate of their own, saw the other participant’s choice and justification. After viewing the target response, they answered the same questions with regard to similarity of this choice and reasoning to their own, their willingness to support the target choice, and the morality, ethicality,
thoughtfulness and intelligence of that response. Only after making these evaluations did they select a kidney recipient and provide an explanation for their own decision.

**Results**

In line with our previous studies, Committed participants were less likely to believe that a target’s preferred transplant candidate was the best possible recipient ($M = 4.21$, $SD = 2.22$) than were Uncommitted participants ($M = 4.71$, $SD = 2.15$; 95% CI[0.08, 0.94], $t(397) = 2.31$, $p = 0.021$). Using our composite measure comprised of how moral, ethical, intelligent, and thoughtful the choice was, we also observed lower evaluations in the Committed ($M = 4.11$, $SD = 1.87$), than the Uncommitted condition ($M = 4.51$, $SD = 1.75$; 95% CI[0.04, 0.75], $t(397) = 2.17$, $p = 0.031$). Within the composite measure, the effect of Commitment was negative and statistically significant on three of four scales: Committed participants thought that targets were less intelligent (95% CI[0.06, 0.86], $t(397) = -2.26$, $p = .025$), less reasonable (95% CI[0.04, 0.81], $t(397) = -2.15$, $p = .033$), and less moral (95% CI[0.04, 0.82], $t(397) = -2.14$, $p = .033$). The negative direction held when we asked participants to rate how ethical the target was, though the results did not reach standard levels of significance (95% CI[-0.13, 0.65], $t(397) = -1.31$, $p = .19$).

**The role of disagreement.** Because participants chose between four potential kidney recipients and the target’s choice was randomly selected from these four options, Committed participants had a 25% chance of evaluating a target choice that agreed with their own, and a 75% chance of evaluating one that disagreed. We refer to this variable as “objective agreement.” Among all participants who agreed with the target choice ($n = 152$), Committed participants were more likely to support the target’s recommendation ($M = 6.74$, $SD = 0.96$) than were Uncommitted participants ($M = 6.33$, $SD = 0.97$; 95% CI[0.08, 0.73], $t(150) = 2.47$, $p = .015$).
There was no effect of commitment on willingness to support the target choice among participants who objectively disagreed with their targets ($M_{committed} = 3.31, SD = 1.74$; $M_{uncommitted} = 3.36, SD = 1.83$; $95\% CI[-0.44, 0.46]$, $t(245) = 0.02, p = .99$). The interaction between commitment and objective agreement did not reach significance ($95\% CI[-0.23, 1.04]$, $t(395) = 1.23, p = .22$).27

Our simple effects for commitment at each level of objective agreement is seemingly at odds with our finding of a negative effect for commitment overall: we observed a significant, positive effect for commitment when participants agreed with their targets, and no effect for commitment when participants disagreed with their targets. Commitment had a negative effect overall because Committed participants were less likely to find themselves in agreement with their targets. Committed participants agreed with their targets 26.3% of the time, a proportion indistinguishable from chance, given random selection from four possible target responses ($p = .686$).28 The proportion of Uncommitted participants who agreed with their targets was nearly twice as high as the level of agreement in the Committed condition: 50.5% of Uncommitted participants agreed with their targets, a number highly unlikely due to chance given random selection from four possible options ($p < .001$).

As in Study 3, we examined whether these different levels of objective agreement mediated the effect of commitment on participants’ likelihood of endorsing the target’s choice. We did in fact observe a significant, negative mediated effect ($95\% CI[-0.96, -0.35], p < .001$).

27 Unlike the quantitative judgment of Studies 1 and 2, the answer set in this task was discrete, which allowed for exact agreement between participants and targets. 107 of 335 participants chose the same answer as the target. Agreement was more likely in the Uncommitted group ($log-odds 95\% CI[1.11, 1.34], t(397) = 4.19, p < .001$).

28 The $p$ values we report here and later are based on a binomial test, giving the likelihood that our observed proportion of agreement among participants in each condition was generated by an underlying binomial distribution in which the true probability of success was 0.25.
The same simulation showed a non-significant direct effect (95% CI[-0.15, 0.44], p = .326), and a significant total effect (in agreement with the OLS regression above; 95% CI[-0.93, -0.08], p = .032). A significant mediator without a significant direct effect suggests that the mediating variable is completely responsible for the outcome (Tingley et al., 2014). In our case, disagreement was responsible for the lower rating of the target. In line with our theorizing, commitment had an effect to the extent that it made agreement less likely.

Discussion

Study 4 documented the effect of committing to one’s own judgment in the domain of ethical decision-making where participants considered life and death scenarios with no obvious correct answer. In line with our prior results, participants who had made their own choice first were less willing to endorse the course of action chosen by another participant and evaluated that course of action less positively. The effect of commitment was driven by the likelihood that a participant would agree with the target. Uncommitted participants were twice as likely to agree with the target they were evaluating as committed participants. It was as if uncommitted participants were able to adopt the target’s response as their own. Having done so, they evaluated it more positively.

In our final study, we continued examining important judgment domains with no clear correct answers. We studied the effects of commitment in a national security decision with both a lay and expert sample.

Studies 5A & 5B

We tested the effect of committing to one’s own decision in the domain of national security by developing our own scenario, loosely based on the Obama administration’s decision
to raid the suspected compound of Osama Bin Laden in Abbottabad, Pakistan in 2011.\textsuperscript{29} In terms of structure, the scenario represents a conceptual replication of Study 4: decision-makers faced a set of options that were each appealing in their own way, making for a difficult decision. In terms of content, it represents a realistic situation in which elite national security decision-makers might find themselves. We developed the scenario in consultation with members of the National Security Fellows program at the Harvard Kennedy School of Government, a program reserved for individuals at high levels of military command or civilian leadership in national security (i.e., Colonels in the Army, Marine Corps, and Air Force; Captains in the Navy or Coast Guard). Within the constraints of an embedded survey experiment, the scenario is consistent with the limited information, uncertainty, and high stakes inherent in many national security decisions (Snyder, Bruck, & Sapin, 1962).

We first conducted this experiment on a lay sample recruited from mTurk. We then ran the same experiment with a sample of professionals with extensive experience in the relevant decision domain. This sample broadens the external validity of our findings and provides a more stringent test of our hypothesis. Experts should be less susceptible to the effects of a simple manipulation like the ordering of tasks in a judgment and decision sequence. Below, we report the two studies in parallel, noting only where they significantly diverged in method or result.

**Method**

**Participants**

We aimed for a sample size of 400 for the lay sample and pre-registered that we would collect a sample of 500 experts or recruit for one month, whichever came first.\textsuperscript{30} We recruited

\textsuperscript{29} We based our scenario on public reporting of the raid, not official details (e.g., Mahler, 2015).
\textsuperscript{30} In the pre-registration, we selected aspredicted.org’s “it’s complicated” option for whether data had already been collected. We selected this option because we had collected data on the lay sample, but not the expert sample. No experts had taken our survey when we pre-registered.
402 participants for the lay sample ($M_{age} = 38$, 44% female). For the expert sample, we were able to collect data from 164 participants after one month ($M_{age} = 36$; 20% female). The sample began with contacts of the National Security Fellows program, described above, and expanded to include current and former members of the US Department of Defense (military and civilian), members of the Department of State, Congressional staff members, academics with research interests in national security, and staff members of the White House National Security Council. 78% of the sample reported having military experience, with ranks ranging from junior enlisted to brigadier general. Civilians in government included GS-13s, -14s, and -15s, which are individuals at the upper end of the civilian rank scale equivalent to mid- through senior-officer military ranks.

Our recruitment message to lay participants stated that they would make decisions in a national security context; our message to the expert population—delivered via e-mail by leveraging the personal and professional contacts of the authors—stated that we were conducting “a research project on decision-making in national security environments.” In our expert recruitment message, we stressed that we were looking for individuals who had national security experience, and that completing the survey was strictly voluntary. Compensation for the mTurk study was $1.00 with bonus payments of $.50. For the expert sample we offered the possibility of a bonus (a $100 Amazon e-gift card), but specified that the bonus was not guaranteed for completing the survey.

**Design & Procedure**

**Design.** As in Study 4, we employed a single-factor, two-level (Committed, Uncommitted) design.
**Procedure.** For both samples, participants gave informed consent and then assumed the role of an operations staff member for the commander of U.S. military forces in Africa. They read that the commander had recently received intelligence on the location of a threatening terrorist, whom we referred to as “Combatant X.” Participants read that their task consisted of four steps: 1) reviewing background information on Combatant X; 2) considering possible courses of action; 3) recommending and explaining a course of action; and 4) evaluating the course of action proposed by another member of the staff. We reversed the third and fourth steps in this process based on condition.

Similar to Study 4, participant recommendations were reviewed by a supervisor, in this case the “Commander” in the scenario. The commander, who in actuality was a retired national security professional, reviewed all participant recommendations before making a decision of his own. If participants made the same recommendation as the commander, we entered them into a drawing for the bonus.

After reviewing these initial instructions, participants read the main body of the scenario, which was identical across conditions (Appendix C). The scenario stated that the commander had recently received intelligence on the possible location of Combatant X. The scenario stressed that Combatant X was considered one of America’s deadliest enemies. It also stressed, however, that Combatant X’s suspected compound was in a heavily populated area, which posed the risk of civilian casualties if U.S. forces were to attack. It was unclear whether the local government was aware of and supporting Combatant X’s shelter. This uncertainty posed a difficult diplomatic problem for the U.S., which had interests in maintaining good relations with the local government as well as in capturing the terrorist.
At the request of the commander, participants reviewed four decision options. The options, which we presented in randomized order, included “embedding a conspirator,” “waiting for movement,” “assisting the host nation,” and “independently attacking” (see Appendix C for a complete description of each option). Each option offered compelling reasons for being selected, as well as clear risks in terms of loss of life or diplomatic tensions.

Our treatment occurred after participants reviewed the options. Committed participants selected their top option to propose to the commander, and then provided an explanation of their choice. Committed participants then reviewed the recommendation of their purported partner, which consisted of one randomly selected option and a corresponding explanation.

As in Study 4, Committed participants stated how similar they perceived the partner’s recommendation to be to their own, and how intelligent, thoughtful, ethical, and moral the partner’s recommendation was. Also as in Study 4, these assessments of the partner’s recommended option were highly correlated in both samples (lay sample Cronbach’s $\alpha = .92$; expert sample: $\alpha = .89$), so we combined them to form a global measure of a participant’s evaluation of the partner’s recommendation. Participants also stated whether they believed that the partner’s chosen option was the “best overall” option on a 7-point Likert scale. Uncommitted participants completed the exact same tasks, only with the order of selecting an option of their own and evaluating the recommendation of another reversed.

We concluded the study with demographic data. With the lay sample we collected age, gender, and political preference. With the expert sample, we asked participants whether and how much experience they had in national security areas, what rank they had attained in their most recent national security job, their highest level of education, and their age, gender, and political orientation.
Results

In line with our prior results, the participants in the lay sample evaluated the choice of another participant more negatively after having made a choice of their own ($M_{\text{committed}} = 4.33$, $SD = 1.70$; $M_{\text{uncommited}} = 5.01$, $SD = 1.63$; 95% CI[0.35, 1.00], $t(400) = 4.05$, $p < .001$). Importantly, the same pattern emerged among the experts, who seemed similarly prone to change their evaluation of a decision based on a simple ordering manipulation (composite rating $M_{\text{committed}} = 3.59$, $SD = 1.64$ $M_{\text{uncommited}} = 4.13$, $SD = 1.62$; 95% CI[0.03, 1.05], $t(162) = -2.11$, $p = .037$).

Furthermore, in the lay sample, Committed participants were less likely to believe that the proposed option of a partner was the best possible option ($M = 4.07$, $SD = 2.07$) than were Uncommitted participants ($M = 4.91$, $SD = 1.85$; 95% CI[0.46, 1.22], $t(400) = 4.30$, $p < .001$). In the expert sample, the relationship held directionally, though the difference was not statistically significant ($M = 3.97$, $SD = 2.07$, vs. $M = 4.26$, $SD = 1.96$; 95% CI[-0.34, 0.91], $t(162) = -0.91$, $p = .362$). We suspect that this dependent variable—whether the target’s response was the best possible option—was especially conservative among experts, who would be better able to imagine other possibilities.

The role of disagreement. As in Study 4, we examined the differential effects of commitment to one’s own decision in cases where a participant agreed or disagreed with their target. In the lay sample, we observed a significant interaction between our treatment and agreement on the composite evaluation of the target (95% CI[-1.50, -0.38], $t(398) = -3.31$, $p = .001$). Both simple effects of this interaction were significant. Committed participants gave higher composite ratings to targets in cases of agreement ($M = 6.59$, $SD = 0.60$) than did Uncommitted participants, ($M = 6.01$, $SD = 0.90$; 95% CI[-0.87, -0.27], $t(138) = -3.76$, $p < .001$).

31 In the lay sample, 140 of 402 participants chose the same answer as the target. In the expert sample, 41 of 164 participants chose the same answer as the target.
The opposite pattern emerged in cases of disagreement: Committed participants gave lower ratings to targets \( (M = 3.72, SD = 1.35) \) than did Uncommitted participants \( (M = 4.09, SD = 1.61; 95\% CI[0.01, 0.73], t(260) = 2.02, p = .044) \).

In the expert sample, the pattern was similar. Committed participants gave slightly higher composite ratings to targets who agreed with them \( (M = 5.98, SD = 1.21) \) than did Uncommitted participants \( (M = 5.84, SD = 0.93) \), though this difference did not reach significance \( (95\% CI[-0.55, 0.84], t(39) = -.41, p = .68) \). The opposite pattern emerged when experts disagreed with their targets, which again mimicked the pattern in the lay sample \( (M_{\text{committed}} = 3.06, SD = 1.19; M_{\text{uncommitted}} = 3.38, SD = 1.25; 95\% CI[-0.76, 0.11], t(121) = 1.44, p = .151) \). The interaction between Commitment and agreement was not significant in the expert sample \( (95\% CI[-0.42, 1.35], t(160) = 1.03, p = .306) \).

![Figure 3.5](image)

**Figure 3.5.** Group means (in text) of participant’s evaluations of the target’s morality and intelligence, using a composite measure of participant’s rating of the target of the peer options intelligence, thoughtfulness, ethicality, and morality. The two panels show the results for the two types of samples collected in Study 5—the left panel shows the effects of commitment for lay participants who either agreed or disagreed with their targets (horizontal axis). The right panel shows the same thing for a sample of national security experts. Bars represent standard errors, and the larger error bars in the right panel indicate the smaller sample of experts.

The likelihood of agreement with the target appeared to mediate the effect of Commitment for both the lay and expert sample. First, we observed that agreement was more likely in the Uncommitted condition (lay sample: log-odds 95\% CI[1.19, 1.42], t(401) = 5.71, p
< .001; expert sample: log-odds 95% CI[0.99, 1.29], t(163) = 1.82, p = .070). Simulations suggested that this agreement carried the effect of Commitment. In both samples, we observed a negative mediating effect of agreement (lay sample: 95% CI[-0.81, -0.38], p < .001; expert sample: 95% CI[-0.66, 0.04], p = .080), a non-significant direct effect (lay sample: 95% CI[-0.34, 0.18], p = .59; expert sample: 95% CI[-0.58, 0.15], p = .26), and a significant total effect (lay sample: 95% CI[-0.99, -0.35], p < .001; expert sample: 95% CI[-1.03, -0.03], p = .038). Again, this pattern of results suggests that Commitment had an effect to the extent that it made agreement less likely. In line with our theorizing, it was the disagreement caused by the act of committing that led to lower ratings of the target.

**Discussion**

Study 5 examined the effects of commitment using a realistic scenario, in an important setting, with both lay and expert participants. For both types of participants, the act of committing to a course of action in a national security setting lowered evaluations of the intelligence and morality of the options offered by peers. In both samples, this effect was driven by the greater likelihood of disagreement in the committed condition.

This study provides a replication of our effect from Study 4 both in a different context as well as with individuals experienced in the decision domain. Although our expert sample was considerably smaller due to the challenges of recruiting experienced professionals with a security background, we observed a pattern that was quite similar to that produced by lay participants.

**General Discussion**

In seven experiments we demonstrate that people’s evaluations of the judgments and decisions made by others are driven in part by the temporal ordering of the task. Individuals who offered their own judgment prior to evaluating that of another participant consistently assessed
the target judgment more negatively than those who evaluated an identical judgment without first committing to their own stance. We documented this phenomenon across different domains, including a variety of quantitative judgments (Studies 1, 2A & 2B), and complex moral choices (Studies 4 & 5). Furthermore, the phenomenon emerged whether participants believed that the judgment they were evaluating was the result of a simple guess (Studies 2A and 2B) or the result of a structured judgment process (Study 1).

In line with our theorizing, participants’ evaluations of the target judgments were largely driven by the extent to which those judgments diverged from the participants’ own views. On quantitative estimates, when participants evaluated a peer’s input prior to committing to their own assessment, their own judgments assimilated toward the target. On the decision tasks used in Studies 4 & 5, participants who had not made their own decision prior to evaluating that of a peer were more likely to make the same decision as the one they had evaluated. As a consequence, participants who did make their own judgments or decisions before evaluating those of others observed a greater amount of disagreement between themselves and their peers. This disagreement ultimately accounted for the different evaluations produced by our order manipulation.

Although one could argue that questioning the merit of a judgment that deviates from one’s own makes logical sense, Study 3 demonstrates that people go beyond the dictates of logic in punishing judgments that disagree with their own. When we asked participants to evaluate their own judgments and those of randomly selected partners, we observed that disagreement had a negative effect only on evaluations of peer judgments, not on evaluations of own judgments. The negative evaluations that follow from making independent judgments result from biased interpretations of disagreement.
Importantly, we also observed our effect with both lay and expert samples. In our final study, national security professionals engaged in a decision-task that was likely highly familiar to many of them. Although the task featured a high degree of uncertainty (similar to analogous real-world situations), it would have been reasonable to expect that the evaluations of experts would remain immune to our order manipulation. Instead, similar to our lay samples, the experts assimilated their own decisions to those of an unknown “peer” after seeing the peer’s choice, resulting in different degrees of disagreement between conditions. The perception of greater disagreement led experts who had first offered their own decision to evaluate those of a fellow national security professional more negatively.

**Theoretical implications.** Our research extends work on anchoring by demonstrating that this phenomenon can have additional consequences for complex judgment and decision-making processes beyond the well-documented assimilation effects. Specifically, we show that, because of assimilation, anchoring can affect both actual and perceived disagreement between the judgments of group members, a process that ultimately affects the group members’ assessments of each other’s contributions.

Furthermore, our work extends research on the phenomenon of “naïve realism” and the manner in which individuals assess the merit of judgments, decisions, and viewpoints espoused by others. Prior work has demonstrated that people disparage ideas and viewpoints to the extent that they differ from their own. However, in many contexts people are confronted with the ideas of others when they have not yet had the chance to formulate their own stance. Our data suggest that naïve realism continues to operate in this context, via the assimilation process referenced above. When individuals have to assess the judgments and decisions of others without first committing to their own view, those target judgments *appear* to be more similar to one’s own, as
of yet unformed, judgments. As a result, people evaluate an identical decision, made through an identical process, and justified with identical reasoning, more positively.

Our work is also related to prior research on advice-taking, or the “Judge Advisor System” (JAS) in quantitative domains. As we previously mentioned, studies using the JAS paradigm examine the extent to which participants adjust their own estimates of an unknown quantity in response to seeing the “advice” of another participant (Bonaccio & Dalal, 2006). This measure, usually operationalized as the proportion of the total distance between a participant’s and an advisor’s estimate that a participant adjusts over the course of the task, is taken to be a proxy for how much trust or confidence a participant places in the advisor’s estimate. This body of work has consistently found that individuals systematically give greater weight to their own judgments than those of others.

Different studies using the JAS paradigm have reached divergent conclusions regarding how disagreement impacts the process of advice taking, with some studies concluding that weight of advice (WOA) shows a negative linear relationship with disagreement, and other studies concluding that WOA shows a curvilinear relationship with disagreement (low WOA at high and low disagreement, with WOA peaking at mid-range levels of disagreement). These divergent conclusions may result in part from the adjustment measure: when two estimates in a JAS experiment are close together, any small adjustment by the participant in the direction of the advisor constitutes a large proportion of the distance between the two estimates. Our work advances this discussion by showing that when participants are free to use an easily interpretable self-report measure to evaluate the quality of others’ estimates, we see a clear linear relationship between evaluation and disagreement. Evaluations of other’s judgments begin to suffer at the
first sign of disagreement (see supplementary studies), and increasingly suffer as disagreement grows.

Relatedly, work by Rader et al. (2015) examined the impact of estimation order on the utilization of modal advice. Intriguingly, those authors found that participants who observed modal advice before making their own judgment “pushed away” from the advice by offering estimates that were further away from the advice than participants who made an estimate on their own, then saw advice, and then revised their estimates. The “push away effect” was in turn mediated by the lower confidence that participants expressed in the advice when they saw it at the beginning of the task, prior to making their own estimate.

Although we do not find evidence of lower confidence in the estimates of a peer when that estimate is seen by uncommitted participants, our Study 4 provides a conceptual replication of Rader et al. (2015). Importantly, we observed this effect in a decision task instead of an estimation task. Participants who were in agreement with their partner’s kidney recipient choice actually evaluated that choice more positively when they made the choice first and then evaluated the choice of a partner who agree with them compared to when they saw the partner’s choice first and then chose to agree with it.

**Practical implications.** Our work has important implications for the structure of collaborative judgment and decision-making processes. The prior literature on collaborative judgment has advised group-members to arrive at judgments and decisions independently in order to minimize the assimilation of estimates toward one another and preserve “the wisdom of crowds” (Gigone & Hastie, 1997; Minson, Mueller & Larrick, 2017; Sunstein & Hastie, 2015). Our current work demonstrates that while this approach is undoubtedly correct from the perspective of increasing the diversity of estimates, it might introduce a hidden cost in terms of
intra-group conflict. Lower evaluations of group member contributions and the reasoning behind them are likely to increase task and relationship conflict in groups, both of which have been shown to be deleterious to performance (De Dreu & Weingart, 2003).

Future research should examine alternative approaches to structuring group processes that preserve the independence of members’ inputs while avoiding potential relational pitfalls associated with group members assessing each other’s contributions after committing to their own views. In the case of quantitative estimation, such a process might involve simple mathematical aggregation of independent estimates. In the case of more complex decisions, one could imagine appointing a group leader who is not committed to a course of action to evaluate and aggregate the views of group members.

Summary. Our research brings together two prior theories in judgment and decision-making and social psychology to make a novel set of predictions regarding the effect of the structure of the decision process on the group outcomes. Across a variety of stimuli and populations, we find that participants’ evaluations of their peers’ judgment vary systematically as a function of an ostensibly irrelevant factor, namely whether the evaluator had or had not previously committed to their own viewpoint prior to considering the input of a peer. Future research should examine additional factors that may further influence this pattern—for example, group cohesion, organizational culture, and group composition. However, our current results clearly suggest that decision-makers should closely consider the impact of the judgment process on all relevant group outcomes. Factors that positively impact one outcome might have a harmful effect on another.
References


Appendix, Essay 1

Supplementary Materials

Portrait of the politically-savvy decision maker
Ambiguity aversion reverses when ambiguity seeking is justifiable
with Jennifer S. Lerner

Study 1 Supplementary Materials

Individual Difference Measures. Last, we conducted exploratory analysis of accountability’s interactions with the individual difference measures we administered (numeracy, honor orientation, and self-monitoring). We observed a significant interaction between accountability, knowledge, and numeracy (log-odds 95% CI [0.44, 0.82], z = -3.14, p = .002). For ease of explanation, we performed a median split of numeracy scores (the interaction analysis used a continuous measure of numeracy). Accountable individuals who were highly numerate were marginally more likely to bet on their own judgment for high knowledge predictions (log-odds 95% CI [0.97, 5.54], z = 1.89, p = .058). Accountable individuals scoring low in numeracy were less likely to bet on their own judgment for low knowledge predictions (log-odds 95% CI [.02, .29], z = -3.81, p < .001). Other contrasts between levels of accountability were not significant. Separately, we did not observe any significant interactions between accountability, knowledge, and the remaining individual differences in the survey, self-monitoring and honor orientation.

Confidence Measures. We examined the effects of accountability across confidence levels (i.e., judged probability), within a given level of knowledge. This allows us to compare whether decision-makers who expressed high confidence on predictions in a high-knowledge domain reacted differently than accountable decision-makers who expressed low confidence in the same domain. We tested the effects of accountability at different levels of confidence by
regressing one’s ambiguity choices onto accountability, reported confidence, and their interaction. We did this separately for high- and low-knowledge domains.

We observed a significant interaction between accountability and confidence in the high-knowledge domain ($\log$-odds 95% CI [0.95, 0.99], $z = -2.31$, $p = .021$), but not in the low-knowledge domain ($\log$-odds 95% CI [0.98, 1.05], $z = 0.89$, $p = .372$).

An alternative explanation to social admissibility as a driver of accountability’s effects on ambiguity tolerance is that accountability leads decision-makers to focus on how much they know of a given domain. Accountable participants would think more about how much they knew in high-knowledge domain and how little they knew in low-knowledge domains. To test this, we analyzed whether a focus on knowledge mediated the interaction we observed above. To do so, we first encoded an individual’s relative knowledge differential between knowledge domains. Specifically, we encoded the difference between an individual’s self-reported knowledge of football and self-reported knowledge of stocks (1-5 Likert scales for each). We also calculated an individual’s difference in ambiguity tolerance across knowledge domains. Next, we tested whether a focus on differential knowledge across accountability conditions mediates the difference between ambiguity tolerance in high and low-knowledge domains. We found no support for an indirect effect of accountability through knowledge. Accountable participants were no more likely to report a greater knowledge of football than stocks than were unaccountable participants ($95\%$ CI [-0.17, 0.22], $t(460) = .276$, $p = .783$).

### Study 2 (Supervisors)

Below is a more detailed write-up of the second study than what appears in the main text.

#### Design and Measurements

[32] We subtract their preference for ambiguity in low-knowledge predictions from their preference for ambiguity in high-knowledge predictions.
To pay bonuses without deception, we randomly paired the choices and justifications of accountable individuals, and had participants (N = 285) in a separate survey on rewards – supervisors – decide which of the pair should receive a bonus. This survey, however, provided an overly noisy test of the effect of tolerance or intolerance of ambiguity across levels of knowledge. Because the pairing was random, for example, participants who made largely similar choices (e.g., ambiguity tolerant at high knowledge and intolerant at low knowledge) could be pitted against one another. Even in instances where the pairing provided a clear contrast of choices under ambiguity, supervisors could have based their bonus payment on content in the participant’s justification that was not pertinent to the ambiguity choice (e.g., number of words written, grammar, whether the writing was polite).

Given the limitations of a random pairing, we conducted a derivative of Study 1 in which we isolated the effects of particular ambiguity choices across high and low knowledge. Specifically, we wanted to test whether supervisors would consistently reward some approaches to ambiguity while avoiding others. To do so, we defined the four strategies of ambiguity attitudes:

1. **Strategy 1**: An individual is tolerant of ambiguity at high and low knowledge, betting on her prediction at both levels of knowledge.

2. **Strategy 2**: An individual is tolerant of ambiguity at high knowledge and intolerant at low knowledge, betting on her prediction at high-knowledge and taking the sure thing at low-knowledge (this is the strategy we observed in the aggregate in Study 1).

3. **Strategy 3**: An individual is intolerant of ambiguity at high and low knowledge, taking the sure thing at both knowledge levels.
4. **Strategy 4**: An individual is intolerant of ambiguity at high knowledge and tolerant at low knowledge, taking the sure thing at high knowledge and betting on their prediction at low knowledge.

To test the effect of choosing one strategy over another, we took each possible permutation of the two strategies (e.g., 1-versus-2, 2-versus-1, 1-versus-3, etc.) and presented it to a unique supervisor (these pairings did not include a written justification), then allowed the supervisor to select the strategy that she believed deserved the bonus. Specifically, we employed a 6 (strategy combination) X 2 (strategy ordering) between-subjects design. The outcome measure was a supervisor’s binary choice of one of the strategies. These strategy-permutations held constant all other factors that could influence a supervisor’s decision, such as the judged probability or justification an individual may have offered.

**Participants**

We recruited 361 participants via mTurk and TurkPrime (48% female, $M_{age} = 36$). We offered participants $.50 for a 5-minute survey, with the possibility of a $1 bonus for carefully reading instructions. Two participants dropped out of the survey after receiving treatment (seeing a permutation), leaving us with a final sample of 359.

**Procedure**

We made clear that participants would evaluate the decisions of individuals who had completed a different survey. We told participants that we would refer to them as supervisors, and that we had referred to them by that title in the earlier study (Study 1A) as well. The supervisor’s read that their task was to evaluate a pair of participants and select one for a reward. Our instructions to them emphasized that the participants in the previous survey had made different types of predictions – football and stocks – and that they had claimed more knowledge
about football than stock predictions. The instructions also discussed how the previous participants gave their judged probability estimates, and how we used those estimates to offer certain amounts at equal subjective expected value. On the same page as the instructions we gave participants basic comprehension checks.

After all instructions, supervisors saw the choices of a pair of hypothetically-named individuals. The choices showed the same judged probability for all high- and low-knowledge predictions (80% and 60% respectively). The certain alternative, then, was also constant across the two knowledge levels ($0.40 and $0.30 respectively). A supervisor’s strategy combination condition determined which pair of strategies a supervisor saw, and the supervisor’s strategy-ordering condition determined which strategy appeared first on the screen. With strategy combination and ordering as between-subjects factors, a supervisor evaluated only a single pair of strategies. Supervisors, using any criteria they deemed appropriate, selected a strategy to receive a double bonus. After their selection, supervisors completed basic demographic questions and then concluded the survey. Within a week of survey completion, we bonused a randomly selected 10% of supervisors who correctly answered all comprehension questions.

**Results**

Supervisors consistently rewarded individuals who employed Strategy 2, tolerating ambiguity at high knowledge and avoiding it at low knowledge. Supervisors also preferred Strategy 1, choosing to base the bonus on their own judgment, even in low-knowledge domains. This support, though, was not as strong as the support for Strategy 2.

We began our analysis by defining a binary outcome variable for each choice among the two pairings. In each pairing, a strategy received a score of 1 if the supervisor chose that strategy, and a 0 otherwise. First, we tested for ordering effects by regressing supervisor choice
on a dummy variable for ordering; we did not find a significant result for any of the six strategy combinations. Given this, we collapsed across ordering for all subsequent analyses (put another way, we examined only combinations, not permutations).

A series of binomial tests allowed us to examine whether a strategy was, on average, a winning strategy in the eyes of supervisors. Setting the probability within the binomial test to 0.5, we conducted two-sided tests for each strategy. A strategy that won (or lost) on average against the other three strategies would show a statistically significant probability of success above (or below) 0.5. We observed such a result for Strategy 2 (95% CI [.64, .77], \( p < .001 \)), and a similar result for Strategy 1, though of less magnitude (95% CI [.51, .65], \( p = .037 \)). Strategies 3 and 4 showed the opposite pattern, losing on average (Strategy 3: 95% CI [.32, .47], \( p = .005 \); Strategy 4: (95% CI [.23, .37], \( p < .001 \)).

We observed that Strategy 2, the aggregate strategy in our initial study, won against Strategy 3 (consistent intolerance of ambiguity; 95% CI [.70, .90], \( p < .001 \)) and Strategy 4 (intolerance of ambiguity only at high knowledge; 95% CI [.64, .87], \( p < .001 \)). It was directionally likely to win against the strategy of pure tolerance of ambiguity (Strategy 1), though not statistically significantly (95% CI [.43, .69], \( p = .36 \)).

Figure A1.1 shows how all the strategies fared when faced against each other. For ease of presentation, we converted the binary outcome variable from (1, 0) to (1, -1). Doing so gave us an intuitive presentation of whether a strategy won or lost against another strategy. If, on average, a strategy won against another strategy, it scored greater than zero; if it lost on average, it scored less than zero.
Figure A1.1: The vertical axes depict a strategy’s winning or losing against another strategy; the horizontal axes show the specific strategy against which a strategy was paired. Positive numbers indicate a winning strategy, negative numbers a losing strategy (error bars represent standard errors). Strategy 2 was the most consistent winner; Strategy 1 also won on average, as long as it was not paired against Strategy 2.

Discussion

Study 1B suggested that the social context – at least a social context of peers – rewards when judges of ambiguity stand behind their predictions rather than opt for a guaranteed amount, especially when decision-makers claim high competence in a given domain.

Study 3

Reported Confidence Measures. In Study 3 we did not observe a significant interaction between accountability and confidence in the high-knowledge domain (log-odds 95% CI [0.95, 1.02], z = -0.88, p = .378). We also did not observe an interaction in the low-knowledge domain (log-odds 95% CI [0.94, 1.04], z = -0.48, p = .628). When we examine the effects of accountability across different levels of reported knowledge (using the difference between one’s
reported knowledge in football and stocks as a measure of knowledge differential between the two levels of knowledge), we did not see similar effects as in Study 1. Specifically, we did not observe a significant interaction between accountability and one’s reported knowledge differential \((\text{log-odds } 95\% \ CI [0.67, .3.27], z = 0.98, p = .329)\).

**Individual Difference Measures.** We analyzed scale data for the data set as a whole and then within each of the choice conditions. We did not observe an interaction between accountability, knowledge, and an individual’s numeracy score within the entire data set \((\text{log-odds } 95\% \ CI [0.78, 1.20], z = -0.41, p = .682)\), nor within the certainty condition \((\text{log-odds } 95\% \ CI [0.62, 1.38], z = -0.36, p = .719)\) or lottery condition alone \((\text{log-odds } 95\% \ CI [0.72, 1.30], z = -0.22, p = .824)\).

For the remaining scales, we used the factor scores from a two-factor analysis using a varimax rotation. With concern for honor, we did not observe a significant interaction among accountability, knowledge, and honor across both choice conditions \((\text{log-odds } 95\% \ CI [0.59, 1.28], z = -0.71, p = .479)\), nor within the lottery condition \((\text{log-odds } 95\% \ CI [0.88, 2.43], z = 1.47, p = .141)\). However, we did observe an interaction among those participants facing certain alternatives \((\text{log-odds } 95\% \ CI [0.21, 0.75], z = -2.85, p = .004)\). Among those participants, individuals high in concern for honor were less likely to react to accountability – put another way, the interaction we observe between accountability and knowledge is carried by individuals low in concern for honor. At high knowledge, accountability is significant among participants low in concern for honor \((\text{log-odds } 95\% \ CI [1.27, 10.24], z = 2.41, p = .016)\), but not those high in concern for honor \((\text{log-odds } 95\% \ CI [0.60, 4.00], z = 0.91, p = .363)\). The same pattern holds, at low knowledge, though the data does not reach standard levels of significance (low concern
for honor (log-odds 95% CI [0.13, 1.20], z = -1.63, p = .102; high concern for honor: log-odds 95% CI [0.21, 1.49], z = -1.15, p = .250).

We did not observe a significant interaction between accountability, knowledge, and social risk in the data as a whole (log-odds 95% CI [0.61, 1.29], z = -0.63, p = .530), the certainty condition (log-odds 95% CI [0.73, 2.29], z = 0.89, p = .369), and marginally in the lottery condition (log-odds 95% CI [0.37, 1.04], z = -1.80, p = .072). We observed no interaction with financial risk in the data overall (log-odds 95% CI [0.77, 1.63], z = 0.60, p = .547), the certainty condition (log-odds 95% CI [0.60, 1.92], z = 0.25, p = .806), or the lottery condition (log-odds 95% CI [0.67, 1.83], z = 0.42, p = .676). We did not observe an interaction with need for cognition in the entire dataset (log-odds 95% CI [0.51, 1.13], z = -1.37, p = .171), the certainty condition (log-odds 95% CI [0.37, 1.23], z = -1.27, p = .204), or the lottery condition (log-odds 95% CI [0.57, 1.70], z = -0.07, p = .946).

**Study 4 (Supervisors)**

The pattern of results in Study 2 suggests that tolerance of ambiguity has reputational significance only when decision-makers appear to back down to safe-space alternatives. Study 2B allows us to test whether supervisors agree. We pair the same strategies as in Study 1B (printed below again for convenience), but we add an additional factor for the alternative choice that a decision-maker faced.

1. **Strategy 1**: An individual is tolerant of ambiguity at high and low knowledge, betting on her prediction at both levels of knowledge.

2. **Strategy 2**: An individual is tolerant of ambiguity at high knowledge and intolerant at low knowledge, betting on her prediction at high-knowledge and taking the sure thing at low-knowledge (this is the strategy we observed in the aggregate in Study 1).
3. **Strategy 3**: An individual is intolerant of ambiguity at high and low knowledge, taking the sure thing at both knowledge levels.

4. **Strategy 4**: An individual is intolerant of ambiguity at high knowledge and tolerant at low knowledge, taking the sure thing at high knowledge and betting on their prediction at low knowledge.

Supervisors evaluated decision-makers who chose between their prediction and a sure thing of lesser value, or between their prediction and a lottery matched to an individual’s judged probability. In terms of expected value, the alternatives were the same across choice conditions. Though, just as participants in Study 2A appeared to see a reputational signal in turning to a certain alternative but no signal in turning to a matched lottery, Supervisors may also interpret different reputational signals from the two choices as well.

**Method**

**Design and Measurements**

We employed a similar design to Study 1B, with the addition of a treatment condition for alternative choice. As with the permutations of different strategies, choice was a between-subjects factor, resulting in a 6 (strategy combination) X 2 (strategy ordering) X 2 (alternative choice) between-subjects design. One additional change from Study 1A was an open-ended response asking participants why they made the choice they did.

**Participants**

In Study 1B we recruited 30 participants per cell (60 per strategy combination). We employed a similar strategy here, recruiting N = 780 via mTurk and TurkPrime. We again offered participants $.50 for a 5-minute survey, with the possibility of a $1 bonus for carefully

---

33 In Study 1, 4.5% of predictions offered 100% confidence; 5.5% did so in Study 2. This percentage of participants chose between their prediction and a sure thing of equal value.
reading instructions. Somewhat surprisingly, 55 participants dropped after receiving treatment. We attribute this difference from Study 1B to the additional writing question (though writing was optional, participants may have decided to leave the study upon seeing an open-ended response). Regardless of the reason, and despite the large number of variables in our design, we observed no differential attrition across any of our three treatment variables or their interactions (at uncorrected levels of significance). Our final sample totaled 725 (46% female, $M_{age} = 37$).

**Materials and Procedures**

Materials were unchanged from Study 1B except for the matched lottery condition. Supervisors in that condition read that decision-makers chose between their own prediction and a “lottery matched to her/his confidence level.” We showed the choice alternatives on the supervisor’s screen, reminding them of a decision-maker’s confidence level and odds of winning the matched lottery: “I prefer $0.50$ if my prediction is correct (80% confident)” vs. “I prefer the matched lottery (80% chance of winning $0.50$).” The same was true in the certain condition: “I prefer $0.50$ if my prediction is correct (80% confident)” vs. “I prefer $0.40$ for certain.” Within a week of completion we bonused a randomly selected 10% of supervisors who correctly answered all comprehension questions within the survey.

**Results**

Supervisors rewarded ambiguity tolerance at high knowledge and intolerance at low knowledge only when decision-makers faced a certain alternative to their prediction. Supervisors showed no systematic preferences when decision-makers faced a matched lottery alternative to their prediction.

As in Study 1B, our analyses focus on Strategy 2 (ambiguity tolerant at high knowledge, intolerant at low knowledge), though Figure A1.2 shows how each strategy fared against each of
the other strategies. When decision-makers faced certain alternatives to their own payoffs, the pattern of results is unchanged from Study 1B. Strategy 2 was a winning strategy overall (probabilities greater than .5 signify a winning strategy, on average: 95% CI [.59, .73], p < .001), as was Strategy 1 (95% CI [.52, .67], p = .015).

Examining head-to-head match-ups, Strategy 2 won against Strategy 3 (ambiguity intolerance; 95% CI [.53, .78], p = .011) and 4 (ambiguity tolerant at low knowledge; 95% CI [.61, .82], p < .001). Strategy 2’s likelihood of winning was not greater than 50% against Strategy 1, though it won on average more than it lost (95% CI [.43, .71], p = .35).

When decision-makers faced matched lotteries, however, supervisors showed no systematic preferences. They did not consistently reward or avoid Strategy 2 against Strategy 1 (95% CI [.37, .62], p = 1), Strategy 3 (95% CI [.44, .71], p = .289), or Strategy 4 (95% CI [.37, .67], p = .885). A similar pattern of results occurred for Strategy 1 (ambiguity tolerant in both high and low knowledge). When supervisors evaluated decision-makers who faced certain alternatives, they (marginally) rewarded Strategy 1 against Strategy 3 (95% CI [.48, .75], p = .105) and Strategy 4 (95% CI [.59, .82], p < .001). When the alternative was a matched lottery, though, Supervisors showed no statistically significant preferences for Strategy 1 against both Strategy 3 (95% CI [.44, .70], p = .306) and Strategy 4 (95% CI [.43, .70], p = .350).
Figure A1.2: Strategy 2 was the most consistent winner only when supervisors observed decision-makers who faced certain alternatives. When decision-makers faced match-lottery alternatives, supervisors showed no consistent preference for one strategy over another.

Supplementary Study

This supplementary study tests an assumption made in Study 4 of the main essay: despite a matched lottery being a technically less ambiguous decision option than one’s own prediction, participants nevertheless perceive the matched lottery bet alternatives as subjecting their initial judgments of ambiguity to evaluation by others. They would, however, view the guaranteed alternative as the best means for avoiding evaluation on their initial judgment of ambiguity.

To test this assumption we asked participants to imagine that their goal was to avoid being evaluated on their initial confidence level (a judgment of ambiguity), then offered participants a choice between betting on their own prediction at a confidence level they previously stated, a guaranteed amount defined by that confidence level, or a lottery matched to
the participant’s stated confidence level. Results supported our assumption. When a decision-maker’s goal was to avoid being evaluated for their judgment of ambiguity, they were most likely to choose the guaranteed alternative and least likely to choose the matched lottery bet alternative.

**Method**

**Participants**

We pre-registered a sample size of 110 participants, and recruited 111 via Amazon’s mTurk and the TurkPrime interface. Also pre-registered, we excluded individuals who were not able to answer basic comprehension questions. Twenty-three individuals met this criteria, leaving a final sample size of 88 individuals (25% female, $M_{age} = 36, SD_{age} = 10.73$). We described the study as a survey for basketball fans, stated that it would take approximately five minutes, paid $0.50 as a flat rate to all participants, and paid $1.00 bonuses to a randomly selected 10% of participants who correctly answered all comprehension questions.

**Design and Procedure**

To ensure that the likelihood of choosing a given bet alternative was not an artifact of the order in which we presented alternatives, we randomly assigned participants to an order condition (guaranteed-first, matched-first) in which they saw the guaranteed alternative before the matched lottery alternative, or vice versa. The experiment took place on an online survey platform. Participants completed an informed consent page, then stated whether they were more comfortable making predictions about NBA basketball games – our high-knowledge domain of interest – or movements in the stock market. We asked this question to maintain parallelism with previous study designs (see main manuscript). Participants also stated how confident they were in their ability to predict NBA games.
We trained participants on how to think of confidence ratings in a frequentist format. This training was the same as the training given in Studies 1 and 2. It asked participants to imagine making a prediction 100 times – the number of times they felt that they would be correct out of 100 was the confidence level they should report for a given prediction. Participants answered two comprehension questions on this training and read that they had to answer all comprehension questions correctly in order to qualify for the bonus drawing.

Next, participants made a prediction about an upcoming NBA basketball game (New Orleans vs. Golden State, May 8th, 2018). They stated which team they believed would win, and how confident they were in their prediction. The confidence scale ranged from 50 – 100 in increments of 5. After making this prediction, participants saw a screen describing three possible choice alternatives for a hypothetical bet. Participants read that they could bet on their own prediction (at a given confidence level), a guaranteed alternative based on their confidence level, or a lottery matched to their confidence level. Participants in the guaranteed-first condition saw instructions that listed the options in the following order: one’s own prediction, guaranteed alternative, matched lottery. Participants in the matched-first condition saw instructions that reversed the ordering of the latter two options: one’s own prediction, matched lottery, guaranteed alternative. We altered the ordering of the latter two options only because they were our primary focus – we were most interested in the likelihood that participants would choose the guaranteed alternative or the matched lottery. We gave an example which showed the details of each alternative for a given confidence level. Participants then answered a single comprehension question on the bet alternative instructions.

We then reminded participants of the prediction (New Orleans vs. Golden State), what they predicted, and how confident they were in that prediction. They then answered the
following question: “For the hypothetical bet, imagine that your goal was to avoid being evaluated based on your stated confidence level. Based on that goal alone – avoiding being evaluated based on your confidence level – which option would you prefer?”. We then presented the options according to ordering condition. The options reminded participants of their confidence level, guaranteed amount, or likelihood of winning the lottery, respectively. Figure A1.3 shows a view of the question and associated choice options from the perspective of a participant in the guaranteed-first condition who predicted that Golden State would win with 80% confidence.

![Figure A1.3: Strategy View of the dependent variable screen for a participant who predicted that the Golden State Warriors would win with 80% confidence.](image)

Participants then gave basic demographic information – age and gender – before receiving their completion codes for the survey.

**Results**

The majority of participants believed that the guaranteed alternative offered the best option for avoiding evaluation of their initial judgment of ambiguity. Also, participants were less
likely to choose the matched lottery than their own prediction if their goal was to avoid evaluation.

**Preliminary analysis: order effects.** We found no evidence that the order in which options were presented affected participant’s choice of option ($\chi^2(2) = 1.01, p = .604$).

**Primary analysis.** We assumed that participants would perceive the certain alternative as more likely to allow them to avoid being evaluated for their initial judgment. The results supported this assumption. Out of 88 participants, 50 said that the certain alternative would be the best option for avoiding evaluation, 27 said their own prediction would be the best way to avoid evaluation, and 11 chose the matched lottery as the best way to avoid evaluation.

As an initial test of the statistical significance of these differences, we conducted a Chi-squared test, which showed the observed frequency of choices in each bin was highly unlikely due to chance if participants were equally likely to choose each option ($\chi^2(2) = 26.11, p < .001$). To quantify the uncertainty around each individual option being chosen, we created dummy variables indicating whether one’s own prediction, guaranteed alternative, or matched lottery was chosen. For example, if a participant perceived that choosing her own prediction was the best means of avoiding evaluation, we scored the dummy variable for prediction a 1 and the dummy variables for the guaranteed alternative and matched lottery 0. We then ran binomial tests on each of these dummy variables.

The certain alternative’s likelihood of being chosen ranged from just under ½ to 2/3 (95% CI[.46, .67]). This was followed by a participant’s own prediction (95% CI[.21, .41]) and the matched lottery (95% CI[.06, .21]). The observed frequency of choice for the guaranteed alternative was highly unlikely due to chance if the true proportion were equal to the observed proportion of participants choosing their own prediction ($p < .001$) or the matched lottery ($p <
.001). We conducted the binomial test for the guaranteed alternative with the true proportion/null hypothesis equal to the observed proportion for the other two options. Further, it was unlikely that the observed proportion for the matched lottery was due to chance if the true proportion were equal to that observed for one’s own prediction ($p < .001$).

**Discussion**

Results supported the assumption that participants viewed the guaranteed alternative as the option most likely to allow them to avoid being evaluated for their initial judgments of ambiguity. Further, participants viewed the matched lottery as least likely to subject their judgments of ambiguity to evaluation.
Appendix, Essay 2

Supplementary Materials

When waste pays:
Equality versus efficiency in social context
with Charles A. Dorison, Zoe Rahwan, Chris Robichaud, and Jennifer S. Lerner

Study 1 Additional Design & Analyses

Design and Procedure

We asked comprehension questions that reinforced instructions during the procedure. For the trust game comprehension questions, one question differed by condition. In the Observed condition the first comprehension check question read: “You can only be made better off if the sender chooses to join the game and send money;” in the Anonymous condition the first question read: “You can only be made better off if the sender chooses to send money.” For both questions the correct answer was “True”. Correct answer rates were nearly perfect and virtually identical (observed $M = .990$; unobserved $M = .995$).

For the allocation decision instructions, the last comprehension question read: “Your decision in the dilemma will determine whether [your partner] chooses to join the game” (True/False). For Unobserved participants the last question read: “Your decision in the dilemma will remain strictly confidential” (True/False). Unfortunately, we observed different correct response rates to this question across conditions. Among all participants (not just those who met our exclusion criteria), Observed participants were less likely to answer correctly ($M = .943$) than unobserved participants ($M = .996$), a statistically significant difference ($\log$-odds 95% CI [.009, .564]; $t = -2.51; p = .012$). However, we note that our results do not change as we relax or tighten the exclusion criteria – we still observe our main effect if we include participants who answered this question incorrectly. Our results strengthen if we remove participants who
answered this question incorrectly; in effect, the better an individual understood the structure of the game and the role of the allocation decision in that game, the more likely they were to choose the equal allocation.

**Additional Analyses – Allocators**

*Order.* Contrary to predictions, we observed an interaction between our Observed condition and the Order in which we presented the equal and efficient allocations (*log-odds 95% CI [1.17, 5.70]; t = 2.34; p = .019*). An analysis of simple effects revealed that our Observation treatment decreased the proportion of efficient allocations among those who saw the efficient allocation first (*log-odds 95% CI [0.22, 0.68]; cohen’s d = .48, t = -3.32; p < .001*), but not among allocators who saw the equal allocation first (*log-odds 95% CI [0.57, 1.73]; cohen’s d = .005, t = -0.04; p = .969*). While we did not anticipate this effect, we note that a similar ordering variable fails to show an effect in Study 2 (or its direct replication).

*False Consensus.* Allocators were evenly split in their preference for equal-versus-efficient allocations – overall, 52.7% of Allocators chose the efficient allocation. When we asked Allocators what percentage of people they thought would make the same allocation as the Allocator herself, the Allocators were remarkably accurate. Their average guess of the number of people who would make the efficient allocation was 51.5%, which is statistically indistinguishable from the actual percentage who made the efficient allocation in our sample (95% CI [49.3, 53.7]; t(403) = -1.07; p = .286). This accuracy, however, results from both sides erroneously believing that they were in the majority. Allocators who preferred equality thought that preferring efficiency was the minority position (*estimated M 95% CI [37.6, 43.5]; t(190) = -8.09; p < .001*), while Allocators who preferred efficiency thought that efficiency was more of a majority position than it actually was (*estimated M 95% CI [58.8, 63.9]; t(212) = 6.65; p < .001*).
An Allocator’s observation condition also influenced her/his estimate of what proportion of others preferred efficiency. Allocator’s in the Observed condition believed that fewer people made efficient allocations than did Allocators in the Unobserved condition (95% CI [-10.5, -1.88]; t(402) = -2.82; p = .005). Put another way, Allocators did not believe that most people would make the equal allocation, and only express a preference for that allocation if they were in the observed condition, but rather, randomization into the observed condition changed their report of how prevalent equal views were.

Within observation conditions, observed Allocators accurately estimated the proportion of people who preferred efficiency (true M = 46.7; estimated M 95% CI [45.1, 51.7]; t(200) = 1.02; p = .310), while unobserved Allocators slightly underestimated the proportion choosing efficiency (true M = 58.6; estimated M 95% CI [51.7, 57.5]; t(202) = -2.75; p = .007). As with the sample overall, however, high and low estimates by Allocators expressing different preference led to these levels of agreement. Observed participants who preferred efficiency believed that a majority would agree (true M = 46.7; estimated M 95% CI [56.6, 64.4]; t(93) = 7.00; p < .001). Observed Allocators who preferred equality believed a minority would prefer efficiency (true M = 46.7; estimated M 95% CI [33.6, 42.0]; t(106) = -4.21; p < .001).

Unobserved Allocators showed the same pattern of over- and under-estimation, though the differences were less pronounced (among those who preferred efficiency: true M = 58.6; estimated M 95% CI [58.6, 65.5]; t(118) = 1.80; p = .051; among those who preferred equality: estimated M 95% CI [40.1, 48.1]; t(83) = -7.14; p < .001).

**Additional Analyses – Observers**
In order to avoid deception, we recruited a similar number of Observers to play in the trust game opposite the Allocators. As these participants were not pertinent to our hypotheses, we describe their choices here in supplementary material.

We did not offer any hypotheses for observer trusting decisions. However, we report key findings from this stage of the experiment. In order to parallel the observed and unobserved conditions of the Allocator stage, we collected equal-versus-efficient partner choices for only half of the Observers. Within this observed condition, the only other randomized variable was the order in which equal and efficient allocations were submitted, for which we did not observe an effect (likelihood of choosing an efficient partner, log-odds 95% CI [.49, 1.48]; \( t = -0.56; p = .57 \)).

Measuring the proportion of an endowment transferred in the game, we do not observe a significant effect for being in the observation condition (i.e., for being able to choose one’s partner in the trust game), nor do we observe a significant effect for an Observer’s own equal-versus-efficient preference.

Observer’s equal-versus-efficient decisions correlated with their general tendencies to prefer fairness. We observed a negative correlation between Observer’s scores on the fairness sub-scale of the Moral Foundations Questionnaire (MFQ; Graham et al., 2011), and their preference for the efficient allocation (95% CI [-.35, -.17]; \( t(423) = -5.59; p < 001 \)).

We also observed a negative correlation between an Observer’s composite scores on the harm and authority sub-scales and their preference for efficiency (harm: 95% CI [-.25, -.07]; \( t(423) = -3.46; p < 001 \); authority: 95% CI [-.19, -.00]; \( t(423) = -2.02; p = .044 \)).
Study 2 Additional Analyses

Additional Analyses

Interaction between religion and treatment. We observed an interaction between our
treatment and the reported importance of religion in a participant’s life (95% CI [1.78, 18.25], t =
2.39, p = .017). Religious individuals in the control condition made less efficient allocations than
non-religious individuals in the control condition. (95% CI [-13.43, -2.20], t = -2.74, p = .007).
However, in the treatment condition, religious individuals were indistinguishable from non-
religious individuals (95% CI [-3.87, 8.27], t = .715, p = .476).

Study 3 Additional Analyses

Additional Analyses – Observers

Risk. Though the observer’s decisions in the trust game entailed a level of risk (would
the allocator return anything?), we did not observe an association between reported risk tolerance
and trusting in the game. After log-transforming our heavily right-skewed measure of risk
tolerance, we entered it into a regression with cents transferred using the same regression
parameters for previously reported observer trusting behavior. In line with behavioral
explanations of trusting behavior (i.e., Dunning, Fetchenhauer, & Schlösser, 2012), but not
rational-actor models of trusting behavior (e.g., Berg, Dickhaut, & McCabe, 1995), risk tolerance
did not predict trusting behavior (log-odds 95% CI [0.94, 1.07]; z = 0.06; p = .95). Other
measures (reported religiosity, gender, age, time spent on trust decision screen) had neither main
nor interactive effects.

5 of 241 participants indicated that a coin flip game costing $100 to play, and paying off 50% of the time would
need to pay $0 in order for them to play (risk measure adapted from Abrahao et al., 2017). While this is a
nonsensical answer, we retain them in the data. We set their log-transformed risk values equal to zero (the log of 0
equals negative infinity). Results remain unchanged if we remove them from the data. Results also remain
unchanged if we remove all participants who indicated that they were risk seeking, a procedure adopted by Abrahao
et al. (2017).
**False Consensus.** Both sides over-estimated how popular their allocation would be. Observers who preferred equality believed that 37.9 percent of people would prefer the efficient allocation. In our sample, the actual total who preferred efficiency was 45.6; the two estimates were statistically different (95% CI [34.7, 41.2]; $t(130) = -4.65; p < .001$). Observers who preferred efficiency over-estimated the percent of people who would also prefer efficiency (95% CI [57.3, 64.8]; $t(109) = 8.16; p < .001$). In short, both sides showed a false consensus effect.

Despite a false consensus among Observers of both allocation preferences, we observed that most Observers believed the equal allocation would be easier to defend: 70.9 percent of Observers held this view. This belief, though, was closely related to an Observer’s own equal-versus-efficient allocation preference. 94.65 percent of Observers who preferred equality believed that equality would be easier to defend; by comparison, 42.72 percent of Observers believed equality was easier to defend (difference in means 95% CI [41.7, 62.1]; $t(146) = 10.12$, $p < .001$). Put another way, about half of Observers who preferred efficiency indicated that the other decision would be easier to defend, while almost none of the Observers who preferred equality thought that the other decision would be easier to defend.

**Expected Payoffs for Game Strategies.** We reported average bonus payments at the conclusion of the Methods section. Those averages, however, were an artifact of a single randomized pairing of Observers and Allocators, and not necessarily reflective of a given strategy’s expected payoff. To better measure expected payoffs for Observers and Allocators, we ran as many simulations as we had participants in our sample.* The average payoff across the simulations constitutes an expected payoff for an individual’s strategy in our sample’s trust game.

---

* We randomly selected a single Allocator to form a Observer-Allocator pair; we then calculated the Observer’s bonus based on that pair’s decisions. We repeated this 241 times for Observers and 244 times for Allocators.
We observed that trusting was a payoff-increasing strategy. When we regressed average trusting – the mean amount of transfers sent to Allocators with the two allocation preferences – we found a positive relationship (95% CI [0.10, 0.12]; t(239) = 26.46; p < .001†). Expected payoff was lowest for Observers who sent nothing (we should note, however, that actual trusting decisions entailed risk while simulated payoffs do not – for any individual pairing, a trusting Observer could still be matched with an Allocator who returned nothing). Given that trusting paid off by our simulated measure, we could only detect a difference between Observers who preferred equality and Observers who preferred efficiency among those who were not fully trusting of both types of Allocators. Figure A2.1 depicts the expected payoffs for this group, showing that Observers who transferred less than their complete endowment and who themselves preferred equality could expect a higher bonus than Observers who transferred similar amounts and themselves preferred efficiency (according to previous analyses, Observers who favored equality were more likely to discriminate between the two Allocator player types).

† The statistical significance of the findings associated with these simulations should be interpreted with caution. The large number of simulations artificially reduced the standard error of expected payoff of a given strategy. In actual game play, payoffs were highly variable (see standard deviations of bonuses paid, reported at the conclusion of the procedure section).
Figure A2.1: Simulated bonus payoffs for Observers, charted against average money sent to Allocators. Dark blue lines represent Observers who chose the equal allocation and light blue represent Observers who chose the efficient allocation.

Additional Analyses – Allocators

Allocator Trustworthiness Decisions. Allocators who chose the equal allocation opted to return a slightly larger share in the game ($M = 40.25$) than Receivers who chose the efficient option ($M = 34.00$). This difference, however, does not reach standard levels of statistical significance ($95\% CI [-12.91, 0.40]$; $t(242) = -1.85, p = .066$).
Appendix, Essay 3

Supplementary Materials

“I was first and I was right”:  
*The effects of task order on evaluations of peer judgments*  
with Julia A. Minson

Answer Process in Study 1

<table>
<thead>
<tr>
<th>Cost of dog ownership seven-step estimation process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1: The person estimated how much dog food costs every year.</td>
</tr>
<tr>
<td>Step 2: The person estimated how much veterinarian bills cost every year.</td>
</tr>
<tr>
<td>Step 3: The person then added estimates from Steps 1 &amp; 2 to get an estimate of annual costs (they were allowed to use a calculator).</td>
</tr>
<tr>
<td>Step 4: The person estimated how many years the dog would live.</td>
</tr>
<tr>
<td>Step 5: The person multiplied the estimate of annual costs (Step 3) by the estimate of the dog’s longevity (Step 4). They were again allowed to use a calculator.</td>
</tr>
<tr>
<td>Step 6: The person estimated the one-time costs of owning the dog, such as collar, bedding, leash, etc.</td>
</tr>
<tr>
<td>Step 7: The person then added together answers from Step 5 (continuing costs) and Step 6 (one-time costs) to calculate a final estimate of the lifetime cost of owning an average dog.</td>
</tr>
</tbody>
</table>

*Table A3.1:* Seven-step process for estimating the lifetime cost of owning a dog. Participants in the uncommitted condition saw the language above; participants in the committed condition saw the same seven-step process, only directed at the participant him or herself (i.e. “Estimate how much dog food costs every year”).

161
## Kidney Allocation Scenarios

### Candidate Descriptions, Kidney Case

| **Candidate A Description:** “Candidate A is a 34-year-old army veteran and the recipient of the Decoration of Valor. As part of the lead company in a successful and pivotal ground attack, he was exposed to a chemical warfare agent, which is known to be nephrotoxic (i.e., directly damages the kidneys). Since returning from the war, Candidate A has experienced a slow but steady decline in renal function and has now reached end-stage renal disease.

He suffers from post-traumatic stress disorder/depression and therefore may be less likely than others to comply with the rigorous post-transplant regimen. However, he has shown improvement with counseling. His psychiatrist argues that much of his depression stems from his poor health and may be partially or completely resolved when he is restored to health after transplant.” |
| **Candidate A Target Justification (i.e., justification given by target):** "Candidate A needs a kidney because he sacrificed himself on behalf of his fellow soldiers. He is relatively young, and even though his depression may complicate his compliance, the transplant may help with the depression. We could give him back the life that his military service has taken away from him. If we do not give it to him, what sort of message would that send to others about the value of service?"

---

| **Candidate B Description:** “Candidate B is a 44-year old woman who has only been on the wait list for 5 months. She has suffered from high blood pressure for several years, but as a busy, single mother with four children, she has been inconsistent in taking her medication. As a result, her high blood pressure has definitely hastened her renal failure.

However, the available kidney is an exact match and therefore Candidate B has the best chance for a successful transplant outcome. Because the kidney is a good match, she won't have to take as many immuno-suppressant medications as other potential recipients. This means she may be able to avoid many of the health complications that the average kidney recipient experiences.” |
| **Candidate B Target Justification:** "Candidate B is the best match for the kidney, so we know it will be put to good use. Avoiding any health problems from a kidney transplant is rare, and we should take advantage of it. A successful transplant means that she will go off the list and stay off the list. More importantly, she is a single mother, and there is no replacing a single mother of four children. Her children depend on her and we are in a place to make sure that they can continue to depend on her."

---

| **Candidate C Description:** “When he was in his twenties, Candidate C donated a kidney to his brother. At the time, doctors informed him that there was a small risk |

---
that he might someday develop the same disease as his brother, but as his kidney was an excellent match, he was willing to do whatever he could to aid his brother's health.

Now, fifteen years later, Candidate C has unfortunately developed the same disease as his brother, and although his case is less severe, his only remaining kidney is overwhelmed and beginning to fail.”

**Candidate C Target Justification:** “Candidate C is in this situation because he volunteered a kidney earlier in life. His family has seen enough pain due to kidney-related health issues; we can help stop that pain by giving him a kidney now. It's because of people like him that we have kidneys to donate in the first place -- that sort of sacrifice should be rewarded. If we do not give it to him, what sort of message would it send to prospective donors?"

**Candidate D Description:** “Candidate D developed renal failure due to adult polycystic kidney disease, which is usually an inherited disease. She watched her father die of the same disease and is concerned about her younger siblings who may develop the same condition.

After graduating from college, she joined a commercial real estate firm. She subsequently founded Capital Realty, currently one of the country's highest grossing privately owned real estate firms. 51 years old, she has already donated millions of dollars to support polycystic kidney disease research. Should she survive her disease through a kidney transplant, she is willing to convert her company to a non-profit entity and donate all future profits to kidney research.”

**Candidate D Target Justification:** "Candidate D, who inherited her disease, is in no way responsible for the position she is in. She and the rest of her family have suffered more than enough as a result of this disease, and we are in a position to put a stop to it. Further, we should consider the effect that giving her kidney will have on others. If her life is saved, she will continue to give to kidney research, which could save the lives of many more people."

*Table A3.2:* Options and justifications provided to participants in Study 4 of Essay 3.
National Security Scenario

In today’s scenario you will be a member of the operations staff for the commander of U.S. troops in Africa.

You, along with several other members of the staff, are meeting to discuss an urgent and important national security topic. Specifically, you have recently received intelligence indicating the possible location of a highly-valued military target, a person we will call Combatant X.

Your steps for today's task [committed condition]:
1. Review background information on Combatant X
2. Consider possible courses of action
3. Recommend a course of action and explain your recommendation
4. Evaluate the course of action proposed by another member of the staff (your “Partner”).

Earlier this week the Commander received an intelligence briefing that a high-value military target, Combatant X, may be hiding outside of Kampala, Uganda.

The government of Uganda is not aware (as far as U.S. intelligence knows) that Combatant X may be sheltering in their country. Though, if they were aware, it is unclear whether they would assist in capturing him, and they would be upset if the U.S. captured him without their permission.

This presents a difficult diplomatic and tactical problem, as the U.S. wants to maintain positive relations with the Ugandan government – their support in the fight against terrorism is very important over the long-term – but the U.S. also wants to capture Combatant X.
Intelligence analysts believe Combatant X may be hiding in a compound, pictured below. The compound is in the middle of a heavily populated area and is adjacent to a building that serves as a home and school for early-grade children (approximately an elementary school, though grades are not clearly defined). The possible target is too close to the school to allow for an airstrike of any kind, as that would entail unacceptably high civilian casualties.

Combatant X is considered one of the United States’ deadliest enemies. The intelligence community has documented a long history of his leadership role in terrorist plots against the U.S. For over 10 years he has raised funds, recruited fighters, and planned attacks against U.S. personnel. Intelligence sources believe he travels with well-armed guards and that he has a well-established network of informants living in the area, making on-the-ground intelligence collection difficult.

At the request of the Commander, staff members have prepared a range of viable options. Please proceed to the next screen to review the options.
**National Security Scenario Decision Options**

**“Embed a Conspirator” Option:** The intelligence community has a highly-trained human intelligence agent in the area who could attempt to embed within Combatant X’s private network. If successful, the agent could orchestrate an event to capture Combatant X.

The advantages to this option are that it would limit human casualties and runs only a slight risk of upsetting the host nation (the U.S. can deny the existence of the intelligence agent). The disadvantages to this option are that it runs a strong risk of compromising the agent and alerting Combatant X, who, again, is believed to have several informants of his own in the area.

**Embed a Conspirator Partner Justification (i.e., justification given by partner):** "This should limit casualties, and probably won't upset Uganda's government. Both are really important to us. The other options have too many downsides like U.S. casualties in the independent option, upsetting Uganda in the wait option, and losing Combatant X in the assist option."

**“Wait for Movement” Option:** The U.S. has aerial intelligence assets stationed above the compound at all times. While constantly monitoring the feeds from those sources, the intelligence community could wait and see if Combatant X leaves the compound and moves to an area where an airstrike could target him without the risk of civilian casualties.

The advantages to this option are that it limits the risk of civilian casualties, and it also limits the risk to U.S. casualties, since personnel would not have to be sent to the compound for a raid. The disadvantages to this option are that it runs the greatest risk of losing track of Combatant X – if he does move, his best time to elude intelligence would be in the highly-populated area of the city (where an airstrike would not target him). This option may also upset the host nation if the U.S. conducts an airstrike without their permission.

**“Wait for Movement Partner Justification:** "This should limit civilian and U.S. casualties. Both are really important to us. The other options have too many downsides like U.S. casualties in the independent option, sacrificing the agent in the embed option, and losing Combatant X in the assist option."

**“Assist Host Nation” Option:** The U.S. could take what information they have to the host nation, and prompt them to conduct a raid of their own to capture Combatant X.

The advantages to this option are that it runs the least risk of upsetting the host nation – it might actually strengthen ties – and it poses the least risk to U.S. personnel, since they would not have to raid the compound themselves. The disadvantages to this
option are that it runs the greatest risk of losing Combatant X. U.S. intelligence believes that Combatant X has informants within the host nation’s government. There is also no guarantee of protection against civilian casualties if the host nation does launch a raid of its own.

**“Assist Host Nation” Partner Justification:** "This won't upset the Ugandan government and will limit U.S. casualties. Both are really important to us. The other options have too many downsides like U.S. casualties in the independent option, upsetting Uganda in the wait option, and sacrificing the agent in the embed option."

**“Independent Action” Option:** The U.S. could send military personnel to raid the compound and attempt to capture Combatant X. A military team is prepared to conduct the mission if desired.

The advantages to this option are that it gives the greatest chance of capturing Combatant X if he is, in fact, in the compound. The disadvantages to this option are that it puts the lives of U.S. personnel at risk, and will also very likely upset the host nation, who would not be aware of the attack beforehand. It also carries a moderate risk of civilian casualties – while military raids pose less risk than an airstrike, it is possible that civilians may be caught in the crossfire of a raid.

**“Independent Action” Partner Justification:** "This gives us the best chance at Combatant X and likely minimizes civilian casualties. Both are really important to us. The other options have too many downsides like sacrificing the agent in the embed option, upsetting Uganda in the wait option, and losing Combatant X in the assist option."

Table A3.3: Options and justifications provided to participants in Study 4 of Essay 3.

**Supplementary Study 1 – Pilot Test**

**Method**

**Participants**

We aimed to recruit 300 participants, a sample size sufficient to detect a small-to-medium effect with .80 probability. Our final sample totaled 290 adult participants, recruited through Amazon Mechanical Turk (mTurk) (M age = 37, 44% female). Differences from our goal of 300 were due to mTurk and were outside of our control. We compensated participants $0.75 for completing the survey, plus offered an additional $0.75 bonus payment to 10% of participants who followed all instructions. To discourage participants from looking up
information online during the survey, we made clear that the accuracy of their answers would not affect their eligibility for the bonus.

**Procedure**

After consenting to participate, participants read an attention check and were prompted to pay greater attention if they gave an incorrect answer (we retained participants who failed the attention check). As our primary treatment, we randomly assigned participants to one of two conditions that varied whether the participant had committed to an estimate prior to evaluating the peer’s estimate. We describe the “uncommitted” condition first.

Participants in this condition began the study by evaluating a target’s answer before we asked them to generate an answer of their own. As such, these participants first read about the target and the process he or she used to generate an answer:

“A fellow mTurker was asked to estimate the lifetime cost of owning an average dog. This person was instructed to imagine that he or she received the dog itself at no cost, and that he or she owned the dog from an early age until it died. We have designed a process that helps participants make the most accurate estimates, which this person used to come up with their answer.”

On the next screen participants saw the process mentioned above. The process consisted of seven-steps including estimates of an average dog’s life expectancy, one-time expenses, and annual expenses. The process called for rudimentary arithmetic to combine the estimates into a single number. Participants in the uncommitted condition then saw the target’s estimate of the lifetime cost of an average, medium-sized dog – $23,410 – which was based on a 2016 estimate by students at the Veterinary School at the University of Pennsylvania (Giffear & Scott, 2016). Next, participants estimated the likelihood that their fellow mTurker’s estimate was within 10%
of the correct answer, using a scale that ranged from 0% - 100% in five-point increments. We labeled 0% as “no chance”, 100% as “absolute certainty”, and 50% as “coin flip, even chances.” This rating of likelihood of being within 10% of the correct answer was our main dependent variable.

We reversed the process of evaluating and answering for participants in the “committed” condition. These participants made their own estimate of the lifetime costs of owning a dog (following exactly the same seven-step process). After completing the process and arriving at their own estimates, participants in this condition next evaluated the $23,410 estimate of a fellow mTurker. Thus, the conditions were identical in the target of evaluation, and the description of the process that gave rise to that target estimate.

Finally, we asked participants whether they were dog owners (dog owners were defined as anyone who had owned a dog in the past 5 years), and collected basic demographic information.

**Results**

In line with our predictions, participants who had committed to their own estimate of the cost of owning a dog rated the target estimate as being significantly less accurate ($M = 36.13, SD = 27.6$) than participants who saw exactly the same target estimate but did not previously commit to their own judgment ($M = 48.6, SD = 27.6$), $t(288) = 3.84, p < .001$.

A closer examination of this finding revealed that the effect was primarily driven by the responses of the participants who reported owning a dog. Among the 174 dog owners included in our sample, participants in the not-involved condition rated the target estimate as being more accurate ($M = 51.5, SD = 29.2$) than participants in the involved condition ($M = 34.7, SD = 28.4$), $t(172) = 3.84, p < .001$. By contrast, participants who were not dog owners reported similar
levels of accuracy for the target estimate in both the not-involved ($M = 45.1$, $SD = 25.7$) and the involved ($M = 38.9$, $SD = 26.1$) conditions, $t(114) = 1.3$, ns. However, the interaction term did not reach traditional levels of statistical significance $t(286) = 1.57$, $p = .12$.

In the committed condition we can evaluate the role that the discrepancy between the participant’s own estimate and the target estimate played in the evaluation of the target. (We cannot do this comparison in the uncommitted condition because participants did not make their own estimates). A plot of participant estimates against evaluation of the target response suggests that participants base their rating, in part, on how close the target’s response is to their own – that is, evaluations increase as a participant’s estimate approaches the target’s estimate, then decreases as participant estimates move away from the target’s estimate.

![Figure A3.1](image.png)

*Figure A3.1:* The chart depicts evaluation of a target’s estimate against disagreement (in the committed condition only). Disagreement decreases as the values on the horizontal axis approach the correct answer (dashed line). As this happens, evaluations increase.
To formally test this relationship, we regressed a participant’s evaluation of the target on a log-transformation of the difference between a participant’s answer and the target answer.\(^{35}\) (We used a log transformation because the estimates were strongly right-skewed – a few participants gave estimates an order of magnitude above the cost we presented them). There was a strong negative relationship between the log-transformed difference and the rating of another \((b = -17.147, se = 2.784, t(139) = 6.16, p < .001, Adjusted R^2 = 0.21)\). A one percent increase in answer discrepancy is associated with a 0.17 decrease in rating of the target’s likelihood of being within 10% of the correct answer. It appears as though, participants treated their own estimate as a standard of accuracy and evaluated the target estimate relative to that standard.

**Discussion**

In Study 1 participants who had previously committed to their own estimate rated a target estimate less positively than participants who rated exactly the same target, produced in the same manner, but who had not previously produced an estimate themselves. Our effect was primarily driven by dog owners, who may be both more knowledgeable about dogs and more invested in their identity as pet owners. However, it is also possible that dog owners were more critical of another’s estimate because they may have assumed that the other participant was not a dog owner. We addressed these possibilities in Study 2.

**Supplementary Study 2 – Testing Domain Knowledge**

**Method**

\(^{35}\) We also used an additional robust regression method to perform this test in the statistical programming language R: Rousseeuw et al., (2015)’s the “lmrob” function from the “robustbase” package. This test also led to a \(b\) estimate for the difference between a participant’s answer and the target answer that was significant at similar levels to the logarithmic transformation reported above \((p < .001)\).
We recruited dog-owners only for this follow-up study, and made it clear that the estimate they evaluated was also generated by a dog owner. Furthermore, we collected data on the extent to which participants felt knowledgeable regarding dog ownership and considered knowledge about dog ownership important to their identity as pet owners.

**Participants**

We maintained a similar sample size as in Supplementary Study 1, recruiting adult participants \( N = 287 \) through mTurk \( (M \text{ age} = 37, 44\% \text{ female}) \). Compensation was again $0.75 for completing the survey, with an additional bonus of $0.75 for 10% of participants who followed instructions.

**Procedure**

Two participants failed an initial attention check, but we retained their data, as in Study 1. Although the study explicitly recruited dog owners, five of the participants reported that they had not owned a dog in the last five years (we posed this question at the conclusion of the study). We eliminated their data, though this did not change the direction or the magnitude of the results (final \( N = 282 \)).

We made two changes from our Supplementary Study 1 procedure. First, we made it clear to participants that the target estimate they evaluated was made by a fellow dog owner on mTurk. Second, we asked participants about their level of knowledge regarding dog ownership and how important this knowledge was to them. Specifically, we asked, “How knowledgeable are you about the costs of dog ownership?” and “How important is knowledge about dog ownership to you?” We recorded answers to both questions using 5-point Likert scales (1: “Not at all knowledgeable” — 5: “Very knowledgeable”; and 1: “Not at all important” — 5: “Very important”, respectively).
Results

As in Supplementary Study 1, participants who evaluated a target estimate after making the estimate themselves offered a more negative evaluation of the accuracy of that estimate (\(M = 32.3\) \(SD = 26.16\)) than participants who did not previously offer their own estimate (\(M = 43.3\) \(SD = 25.7\)), \(t(280) = 3.56, p < .001\). Also, similarly to Supplementary Study 1, in the involved condition we observed a negative relationship between the participants’ ratings of the target estimate and the extent to which their own estimate differed from the target (we again used a log transformation of the difference between a participant’s estimate and the target estimate: \(b = -11.676, se = 2.696, t(139) = 4.33, p < .001\)).

In examining the effect of self-reported knowledge about the costs of dog ownership and one’s caring about being knowledgeable on this subject, we saw no relationship between participants’ level of knowledge or caring about dog ownership and their rating of the accuracy of the target estimate (knowledge: \(b = -2.54, t(280) = 1.46, p = .15\); importance: \(b = -0.03, t = -0.02, p = .99\)). This was true in both conditions.

Discussion

Supplementary Study 2 tested whether lower ratings of a target estimate may have resulted from a participant’s assumption that the target was not a dog-owner and was thus less knowledgeable than the participant. With the effect of involvement replicating from Supplementary Study 1, and a null effect for the importance of participants’ level of knowledge, we did not find evidence supporting this assumption.

\[\text{As before, we used a robust regression model (Rousseeuw et al., 2015) and found a similar significance level (} p = .001).\]
Supplementary Study 3 – Continuous Disagreement in Quantitative Judgment

Method

In this supplement we test boundary conditions of agreement – how much disagreement is necessary before we observe the disparaging for the effects of commitment? We test this with six different treatment levels of agreement, attempting to approximate a continuous treatment for likelihood of agreement. Initial treatment levels make agreement very likely, but this likelihood gradually decreases across treatment levels. The final study design was a fully crossed 2 (committed, uncommitted) X 6 (likelihood of disagreement levels) design.

Participants

We collected a sample of 1,009 adult participants via Amazon’s mTurk (M age = 36, 47% female) with a description of the study that read: “Study on Estimation – 5 minutes.” Participants were paid $.50 for the study, and the consent page informed them that a bonus was possible.

Procedure

After providing consent participants were told that their task would be to estimate the number of items in a container and to explain the process they used to make the estimate. We also told them that they would see the estimate of another mTurk participant, someone we labeled “Partner” throughout the survey. As a means of incentivizing participants to take the estimate of the Partner seriously, we made the bonus for the survey contingent on evaluation by a third mTurk worker, or “Supervisor”. Specifically, participants read that the “Supervisor,” in a separate survey, would see the estimate of the mTurk worker and her/his Partner; after viewing both estimates, the Supervisor would make an estimate of her/his own. If the participant’s
estimate was within 10% of the Supervisor’s final estimate, the participant would receive a $.50 bonus. The rationale behind this procedure was that a Supervisor would be more likely to make an estimate within 10% of a participant’s answer if another individual independently reached a similar answer.

Participants answered three comprehension questions about this bonus procedure – the correct answers to these comprehension questions drew verbatim from the instructions, and the questions were presented on the same screen as the instructions. 894 of 1,009 participants answered all three questions correctly; 102 missed one question; 12 participants missed two questions, and one participant missed all three questions.

After these instructions, we randomly assigned participants to one of six different levels of agreement, where level of agreement was operationalized as the number of M&Ms in a container. The containers featured 24, 49, 67, 95, 126, or 154 M&Ms, with agreement highest with fewer number of M&Ms in the container. We also assigned participants to either the committed or uncommitted condition. In the committed condition, participants saw a picture of the container of M&Ms, then made an estimate of their own. The subsequent screen showed them the estimate of a target, which was set to the right number of M&Ms in the container (i.e., participants in the highest agreement condition – 24 M&Ms – saw a target who estimated that 24 M&Ms were in the container. Participants then recorded how likely it was that they believed the target’s (“Partner’s”) response was within 10% of the correct answer. They did so on a 10-point scale that ranged from “Extremely Unlikely” to “Extremely Likely”, with no descriptive points in between. For participants in the uncommitted condition this process was reversed – they saw the picture of the container alongside the target’s estimate; evaluated the target, then made an estimate of their own on the subsequent screen.
We concluded by asking how confident a participant was in her own answer of the number of M&Ms in a container, and how much she agreed with the Partner’s estimate of the number of M&Ms.

Results

We first examined whether our operationalization of agreement was effective. Disagreement, defined as the absolute value of the difference between a participant’s estimate and the target’s, increased with the number of M&Ms in the container \( (b = 12.44, t(1007) = 29.877, p < .001) \). Next, we observe a significant, negative effect for commitment when we collapse across all levels of agreement \( (b = -.349, t(1007) = -2.099, p = .036) \). As the figure below shows, commitment led to lower target evaluations for all but the container in which agreement was nearly guaranteed. Across all levels of agreement, we do not observe an interaction, whether we examine the agreement treatment as a continuous variable or as a factor with six levels.
At treatment levels of agreement – number of M&Ms – actual disagreement was higher among committed than uncommitted participants. We observed a significant interaction between agreement and commitment ($b_{interaction} = 3.51$, $t(1005) = 4.353$, $p < .001$). See Figure below.

*Figure A3.2:* Evaluation of the target decreases with commitment; the effect is similar for all levels of likelihood of agreement with the target.
Figure A3.2: At all levels of likelihood of agreement, committed participants disagreed with the target more than uncommitted participants.