"I was first, and I was right"

The effects of order on evaluation of peer judgment

Bradley R. DeWees

Harvard University and United States Air Force

Julia A. Minson

Harvard University

Word count: 10,597

Data, analysis code, and stimulus materials publicly available here

Authors' note. The research reflects the work of the authors and not necessarily the opinions of the U.S. Government or any part thereof. Correspondence concerning this paper should be addressed to Julia A. Minson, Harvard Kennedy School, Harvard University, 79 JFK Street, Cambridge, MA 02138. E-mail: julia_minson@hks.harvard.edu.

Abstract

Across seven experiments (combined N = 3,775), we test whether committing to one's own point of view influences evaluations of peer judgments and decisions. We find that such commitment, relative to simply considering peers' input on its own, leads evaluators to derogate others' contributions. Studies 1 & 2 demonstrate this effect and examine several alternative explanations. Committing to one's own judgment prior to evaluating a peer's judgment creates a greater level of disagreement between the two estimates. Greater disagreement, in turn, drives the derogation of others' inputs (Studies 3A & 3B). In Studies 4 & 5 we go beyond quantitative estimates and test our effect in two complex decision-making scenarios. We find 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. We discuss the implications of our findings for collaborative decision-making. (147 words)

Keywords:

decision making, groups and teams, naïve realism, anchoring, evaluation

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" (Lorenz, Rauhut, Schweitzer, & Helbing, 2011; Minson, Mueller, & Larrick, 2017). 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 can affect how much weight a judgment receives in the final group product. The evaluation a group member assigns to the input of another can vary for many reasons, including work history, expertise, access to specific information, and status. 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 rendering one's own judgment versus evaluating another's.

Across seven experiments, we manipulated whether research participants did or did not commit to their own judgment or decision prior to evaluating a judgment or decision produced by a peer. We consistently found that offering one's own estimate prior to evaluating peer input led individuals to derogate the judgments and decisions of others, assessing them as being of lesser quality than participants who had not yet committed to their own judgments and decisions. We observed this effect in a variety of estimation domains and tasks, and with both lay and expert samples.

Prior literature on evaluating others' judgments and decisions

The evaluation of others' ideas, decisions, and contributions is a basic social activity with a wide range of literatures addressing the factors that drive outcomes. Most prior research suggests that individuals use reasonable criteria to evaluate others: they base evaluations on the accuracy of another's judgment (Budescu, Rantilla, Yu, & Karelitz, 2003), the quality of the information underlying the judgment (Budescu & Rantilla, 2000; Burgoon, Henderson, & Wakslak, 2013), and the reputation of the individual making the judgment (Feng & MacGeorge, 2006; Goldsmith & Fitch, 1997).

When accuracy is difficult to evaluate or when outcomes are ambiguous, evaluators rely on cues such as the general trend of an individual's performance (Barnes, Reb, & Ang, 2012), whether the individual met a minimum threshold of performance (Luan & Reb, 2017), or how confident the individual was in making their judgment (Price & Stone, 2004; Sniezek & van Swol, 2001). Our current work adds to these findings by examining the effect of a novel factor, one that arguably should not have a systematic impact: the temporal ordering of the evaluation relative to producing one's own judgment.

Effects of task order on judgment outcomes

Prior research has shown that changing the order of interacting with others in groups can alter an individual's level of confidence in their own judgments (Heath & Gonzalez, 1995). Relatedly, research on advice-taking in quantitative domains has shown that the order of making a judgment before or after considering advice affects the extent of advice utilization (Rader, Soll, & Larrick, 2015). However, the largest and most relevant body of work examining the effect of

temporal ordering on judgment outcomes, is the extensive literature on "anchoring and insufficient adjustment" (Tversky & Kahneman, 1974).

Research on anchoring demonstrates 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 relevant judgment assimilates to the anchor. Importantly, however, this body of work has not explored many potential "downstream" effects that such assimilation might have on related group judgment and decision-making outcomes.

One consequence of anchoring effects that has been extensively examined in group settings has been judgment accuracy. Classic work on social influence demonstrates that group members' judgments and opinions are often shaped by the judgments and opinions of those around them (Asch, 1955). Indeed, scholars have argued that such behavior can be normative, since the consensus of the group possesses informational value (e.g. Deutsch & Gerard, 1955). However, assimilation toward group-member judgments also emerges in dyads where neither member possesses unique insight, or represents a "majority" opinion (Koehler & Beauregard, 2006; Minson & Mueller, 2013; Minson, Mueller, & Larrick, 2017). In such contexts, assimilation decreases the accuracy of collaborative judgments because it makes team members less likely to "bracket" the truth (Larrick & Soll, 2006; Lorenz et. al., 2011; Soll & Larrick,

2009). Thus, the recommendation of the quantitative estimation literature is to make judgments independently and then combine them for a final estimate (Gigone & Hastie, 1997). This recommendation, however, focuses on the accuracy of judgments and does not consider how independent estimation might affect group members' evaluations of each other's inputs.

In the current work we suggest that, in addition to the accuracy effects noted above, the temproral ordering of estimates also impacts the evaluations of peers and the attributions we make about the bases of their judgments. Thus, 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. 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.

Hypotheses

In the present investigation, we hypothesize that task order (specifically, whether an individual does or does not commit to one's own estimate before evaluating someone else's), systematically affects evaluation of others' judgments and decisions. We make this prediction based on an integration of research on anchoring with research on "naïve realism" (Robinson, Keltner, & Ward, 1995; Ross, Lepper, & Ward, 2010; Ross & Ward, 1996), discussed below.

First, in line with the anchoring literature, we predicted that participants who made their own estimates *after* evaluating the estimate of a peer would make estimates that were 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 would 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. Several research streams have demonstrated that individuals grant their own perceptions of reality a privileged place relative to the perceptions of others – 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 are less willing to adjust their estimates 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).

Based on the above research, we predict that individuals who make a judgment of their own prior to being exposed to the judgment of a peer, will evaluate the peer's judgment more negatively than individuals who make the evaluation without committing to their own estimate first. Individuals who first make a judgment of their own will observe a higher level of disagreement between their own estimate and the estimate of their peer. To the extent that individuals assess their own judgments as representing the most reasonable estimate to be made

under the circumstances, they will assess peer judgments more negatively the more those judgments disagree with their own.

Our hypothesis extends prior theorizing in several ways. First, 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.

Research overview

We present the results of seven studies that examined the effect of task order on evaluation of peer judgment in judgment and decision-making tasks. 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 Study 2, we replicated our effect and ruled out several alternative explanations. In Studies 3A and 3B, we explicitly tested the role 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. Our stimuli, data and analysis scripts are available here.

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 the simple ordering of giving one's own estimate versus evaluating another's estimate changed participants' assessments of the target estimate. Additionally, we examined the role of domain knowledge on our main effect.

Method

Design and Participants. Our sample totaled 415 adult participants, whom we recruited through Amazon Mechanical Turk (mTurk) (*M age* = 36, 47% female).[1] We recruited dog owners and non-dog owners for a 2 (Commitment: Committed, Uncommitted) X 2 (Dog Ownership: Yes, No) design to examine the effect of commitment to one's own estimate on evaluation of peer judgments, as well as whether domain specific knowledge would moderate our primary effect of interest.

We compensated participants $0.50 for completing the survey, plus offered an additional $0.50 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 then randomly assigned

[1] Initial pilot studies (see supplementary material) showed a small to medium effect size for the phenomenon of interest – a cell size of approximately 100 yielded statistical power of 0.80. We decided to recruit an additional five percent per cell – our goal was 420 participants. Differences from our goal were due to mTurk and were outside of our control. Four participants failed an initial attention check but were retained in the data.

participants to one of two conditions that varied whether the participant would (or would not) commit to their own estimate prior to evaluating a peer's estimate.

Participants in the Uncommitted condition began the study by reading about their purported peer and the process they had 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 that we had described. It consisted of seven-steps meant to decompose the estimation task into a series of simpler estimates. The process called for rudimentary arithmetic to combine the estimates into a single number (see Appendix A).

After viewing the process, Uncommitted participants saw the target estimate of the lifetime cost of an average, medium-sized dog: $23,410. We based this estimate on a 2016 report by the Veterinary School at the University of Pennsylvania (Loy, 2017). Next, these participants estimated the likelihood that this 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 judged likelihood of the correctness of the target estimate was our main dependent variable.

By contrast, participants in the Committed condition made their own estimates before they evaluated the target estimate. They followed the seven-step process described above to

make their estimate, and afterwards used the same scale to report their evaluation of the $23,410 estimate, ostensibly produced by a fellow mTurker, who had used the same process.

For both conditions, we then asked participants whether they were dog owners (which we defined as anyone who had owned a dog in the past 5 years), along with a series of questions designed to measure how important the domain in question 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"). We ended the study with basic demographic questions.

Results

In line with our predictions, participants who had committed to their own estimate first rated the target estimate as less likely to be correct (M = 38.80, SD = 28.7) than participants who saw exactly the same target estimate but did not previously commit to their own judgment (M = 46.45, SD = 28.26), b = -7.65, $t(413) = -2.73$, p = .007. We observed no significant interaction between our treatment and whether an individual was a dog owner, $b_{interaction}$ = 0.95, $t(411)$ = 0.17, p = .866. The effect of Commitment was slightly (but not significantly) more pronounced for dog owners (Dog owners: M_{diff} = 8.60, SE = 3.97, $t(222) = 2.16$, p = .031; Non-owners: M_{diff} = 7.65, SE = 3.97, $t(189) = 1.92$, p = .055). Likewise, we did not observe an interaction between our treatment and self-reported caring about dog ownership ($b_{interaction}$ = -2.21, $t(411) = -1.02$, p = .30).[2]

[2] Our three self-reported measures of the importance one placed on dog knowledge were highly correlated (Cronbach's α = .92). We combined them into a composite measure of caring about the domain.

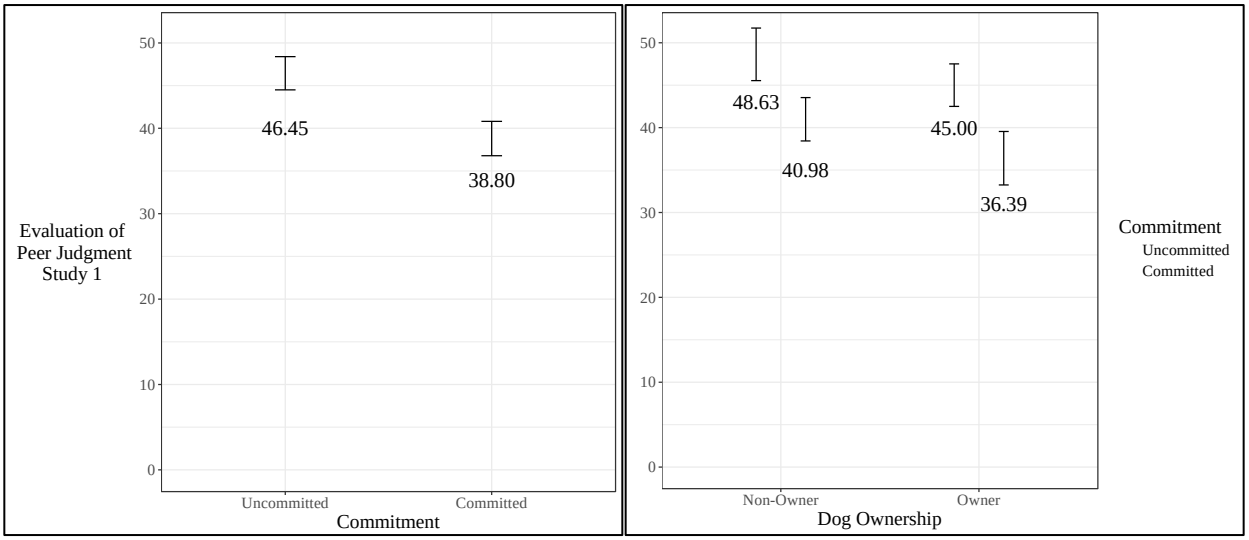

Figure 1. Participants rated the likelihood that a peer's judgment was correct on a scale from 0 – 100 (vertical axis). Committing to a judgment of one's own led to lower evaluations of a peer (left panel). The effects of commitment held across levels of expertise (right panel). Individuals with knowledge of a domain–dog owners– lowered their evaluations of others after committing to a judgment to a similar degree, as did individuals with less knowledge of the domain. Bars show standard errors.

To begin exploring the psychological mechanism underpinning our effect, we examined whether the level of disagreement between a Committed participant's estimate and the target estimate was associated with lower evaluations. When we regressed a participant's evaluation of the target estimate on a log-transformation of the difference between the two estimates,[3] we observed a strong negative relationship (b = -12.41, se = 1.96, $t(202) = 6.34$, $p < .001$). A one percent increase in disagreement was associated with a 0.12 decrease in the evaluation of the target estimate's accuracy. At least within the Committed condition, it appears as though participants treated their own estimate as a standard of accuracy and evaluated the target estimate relative to that standard.

[3] 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. We also used the robust regression method (Rousseeuw et al., 2015) to test the relationship between disagreement and evaluation. This test led to a similar, negative effect of disagreement's effects at highly significant levels (p = .009).

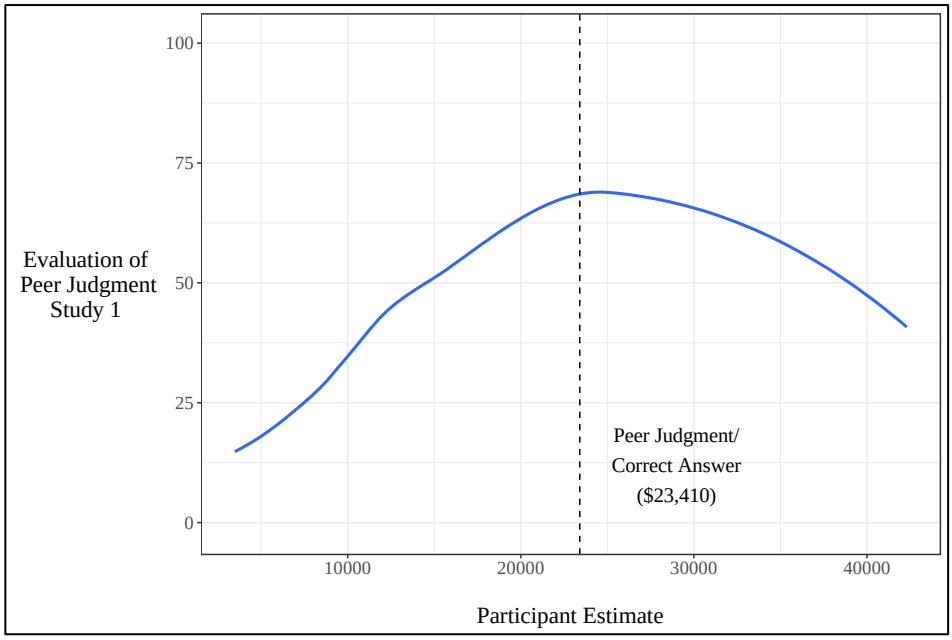

Figure 2. Judged likelihood that the target estimate was within 10% of the correct answer is shown on the vertical axis; the participant's own estimate is on the horizontal axis. The closer a participant's own estimate was to that of the purported peer, the more favorable their evaluation of the peer, suggesting that evaluation of another was in part a function of disagreement. Gray area represents standard error.

Discussion

In Study 1 participants who had previously committed to their own estimate rated a target estimate less positively than participants who rated the exact same target estimate, produced in the same manner, but without having previously committed to an estimate themselves. We did not find evidence that domain importance or expertise moderated the effects of commitment. Suggestively, the amount of disagreement with the target estimate in the committed condition strongly predicted target evaluations.

Study 2

Study 2 extends our investigation by systematically varying the process that participants used to make their estimates. It is possible that Committed participants in Study 1 rated the target estimate lower because they followed an in-depth process – the 7-steps that we provided them – to develop their answer. It may have been their use of the process, rather than a simple commitment to their own judgment, that led them to lower their evaluations of the target if they believed the target had not engaged in a similar process or had executed it less diligently. In Study 2, we directly test the role of using a systematic process (versus making an intuitive estimate) in producing our effect.

Method

Design and Participants. Our study employed a 2 X 2 between-subjects design, with two levels of committing to an estimate (Committed, Uncommitted) crossed with two levels of process complexity (7-step Process, Single Intuitive Estimate). We intended to double our sample size per cell from Study 1 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 volunteers who wish to contribute to social science research.[4] DLABBS volunteers do not receive compensation for their participation in surveys. Given that dog ownership did not affect our results in Study 1, we recruited both dog owners and non-owners for this study.

Procedure. As a condition of volunteering in the DLABBS subject pool, participants had to first answer a demographic questionnaire. As before, we randomly assigned participants either

[4] http://dlabss.harvard.edu/about/

to give their own estimate of the lifetime cost of an average dog and then to evaluate another estimate, or to evaluate the target estimate without committing to their own judgment. In this study, we also randomly assigned participants to two conditions that featured the seven-step process used in Study 1, or two conditions that simply asked them to make/evaluate an intuitive estimate. Thus, in the two Process conditions participants believed that their peer had followed a 7-step process. By contrast, in the two Intuitive Judgment conditions we told participants simply that their peer had made an estimate and made no mention of a more complicated process. Lastly, we asked participants how much they cared about being knowledgeable about dog ownership.

Results

Participants who committed to their own estimate prior to evaluating the target estimate again judged that estimate as less likely to be accurate ($M_{committed}$ = 39.04, *SD* = 26.98), than participants who did not commit to their own estimate ($M_{uncommitted}$ = 45.34, *SD* = 26.67), $t(806) = 3.34$, $p < .001$. The effect of Commitment was negative whether participants evaluated a peer who used a 7-step process ($b_{committed}$ = -6.94, $t(391) = 2.57$, $p = .011$), or made an intuitive judgment ($b_{committed}$ = -5.64, $t(413) = -2.15$, $p = .033$. There was a positive main effect of using a process – participants rated peers who used the 7-step method higher than they rated peers who used intuition alone ($M_{process}$ = 44.10, *SD* = 26.98; $M_{no\text{-}process}$ = 40.41, *SD* = 26.92), $t(806) = 1.97$, $p = .049$. When we simultaneously regressed our outcome variable on both factors and their interaction, we observed a significant negative effect for committing to one's own estimate ($b_{committed}$ = -5.64, $t(804) = -2.14$, $p = .032$), and a marginally significant positive effect for process ($b_{process}$ = 4.32, $t(804) = 1.63$, $p = .104$). Interestingly, the magnitude of the effect size for

Commitment was nearly 70% larger than the effect size for process ($d_{committed}$ = -0.24, $d_{process}$ = 0.14). 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 ($b_{interaction}$ = -1.30, $t(804) = 0.35$, $p = .730$).

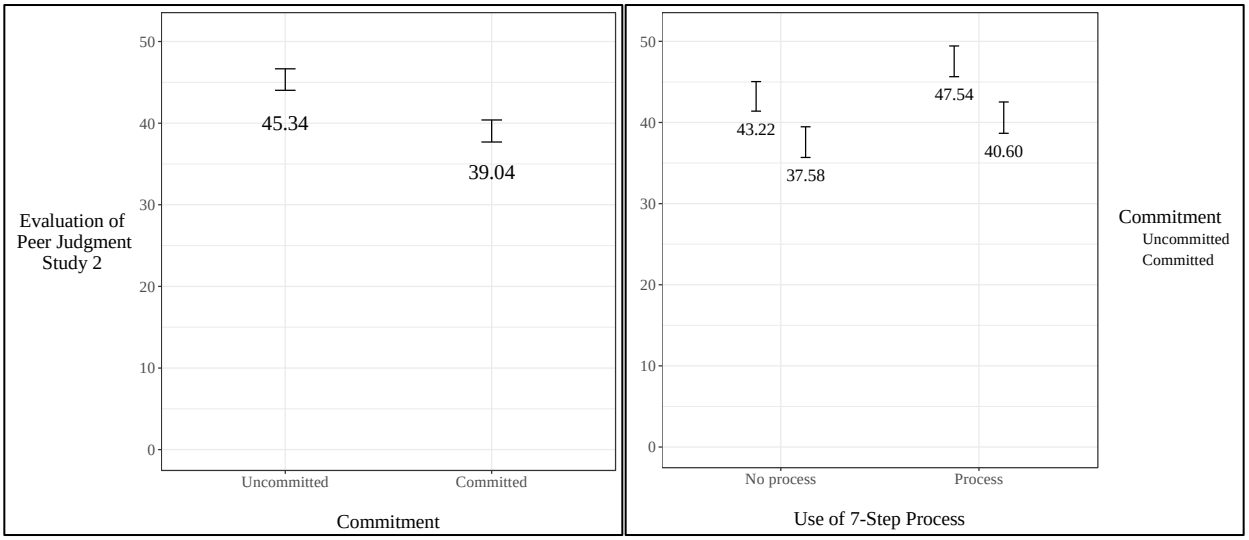

Figure 3. Judged likelihood that the target response was within 10% of the correct answer is shown on the vertical axis. Again, committing to a judgment of one's own led to lower evaluations of a peer (left panel). Commitment's effects held whether a participant engaged in (or observed the use of) an in-depth seven-step process to generate a judgment. In terms of lowering one's evaluation, the effect size of merely committing to an answer is larger than the effect size of engaging in a seven-step process. Bars represent standard errors.

As before, we did not observe a main effect of caring about dog ownership-related knowledge on the evaluations of the target estimate, or interactions between caring and our treatment variables. Finally, we again tested the effect of disagreement on a participant's evaluation of the target estimate. The results followed the same pattern as in Study 1 (b = -16.800, se = 1.467, $t(396) = 11.45$, $p < .001$).[5] Participants evaluated the target estimate more negatively as a function of the extent to which that target estimate deviated from their own.

[5] A robust regression (Rousseeuw et al., 2015) also showed a significant result ($p < .001$).

Discussion

We conducted Study 2 to determine whether the negative evaluation of another's judgment that we observed in Study 1 was the result of using a seven-step process or merely forming an answer of one's own before evaluating the target answer. The results support the latter interpretation. Committing to a judgment consistently lowered participant evaluations of a target estimate, and 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 ordering can be to evaluations of another – the mere act of committing to a judgment was 70% more important to peer evaluations than the difference between following a seven-step process relative to simply making a snap judgment.

Studies 3A & 3B

In Study 3 we begin to explicitly examine the role of disagreement in the evaluation of a peer's judgment 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

Design and Participants. Our next two experiments featured a 2 (Committed, Uncommitted) X 2 (Agreement, Disagreement) factorial design. We report them as Studies 3A and 3B, which followed the same design but required participants to make judgments in different domains.

We recruited 424 adult participants for Study 3A (M_{age} = 35, 53% 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. 43 participants stated that they were not parents, leaving us with a final sample of 381 (M_{age} = 35, 55% female). In Study 3B 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.

Procedure. Participants in both studies first stated how much they "trusted their own judgment" in estimating either the costs associated with childrearing or estimating numerical quantities. For Study 3A, which asked about childrearing, we also asked how knowledgeable participants believed they were about the costs associated with childrearing.

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 read the same instructions, but were told that they would evaluate the likelihood that another participant's estimate was correct "*Before* making your own estimate." Committed participants then made their estimates.

In the Agreement condition of Study 3A, 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 17. 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 questions 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 ratings 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 ratings on a five-point Likert scale, ranging from 1: “Not at all likely” to 5: “Very likely.” Participants in the uncommitted condition reversed this ordering – they saw the target estimate, rated its likely accuracy, and then made an estimate of their own.

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

Figure 4. Agreement (left) and disagreement stimuli for Study 3B.

Participants in both studies concluded by stating 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 with age and gender questions.

Results

Study 3A. 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; $t(210) = 2.15$, $p = .032$). 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; $t(167) = -0.935$, $p = .351$). As a result, we observed a statistically significant interaction between commitment and level of agreement ($b_{interaction}$ = 0.52, $t(377) = 2.138$, $p = .033$).

Unlike in Studies 1 and 2, participants in all conditions of Studies 3A and 3B 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 evaluation. 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], $p = .044$). This effect was stronger in the Disagreement (lifetime costs) condition (*95% CI* [-0.28, -0.01], $p = .040$) than it was in the Agreement (monthly costs) condition (*95% CI* [-0.02, 0.14], $p = .26$).

Study 3B. As in Study 3A, the effects of Commitment were significant in the Disagreement condition ($M_{uncommitted}$ = 2.17, $M_{committed}$ = 1.98; $t(619) = 2.57$, $p = .011$), but not the Agreement condition ($M_{uncommitted}$ = 3.28, $M_{committed}$ = 3.29; $t(588) = -0.069$, $p = .945$). The interaction between the two treatments was marginally significant ($b_{interaction} = 0.20$, $t(1207) = 1.767$, $p = .078$). A mediation analysis again showed a significant, negative mediation effect for objective disagreement (*95% CI* [0.17, 0.36], $p < .001$). The mediation was again stronger in the Disagreement condition (*95% CI* [0.18, 0.54], $p < .001$) than it was in the Agreement condition (*95% CI* [-0.08, 0.07], $p = .87$).

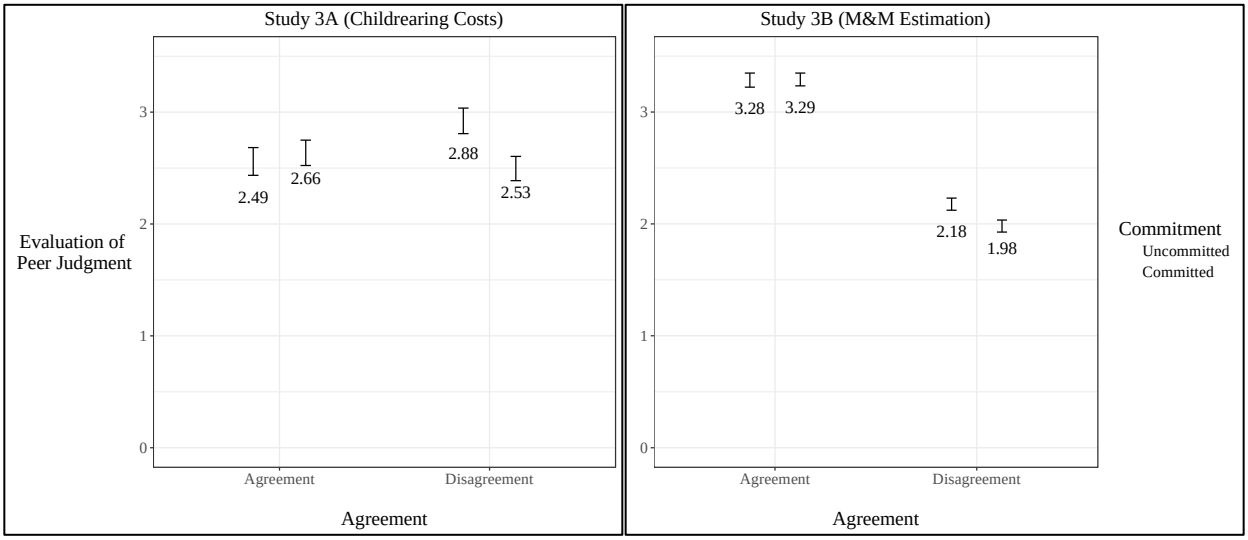

Figure 5. Judged likelihood that the target response was within 10% of the correct answer, measured using a five-point Likert scale (vertical axis). Each panel shows different effects for commitment depending on level of agreement. When agreement was likely, committed participants were statistically 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 3A and 3B 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 a peer estimate. 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 3B 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.

In sum, Studies 1-3 support the hypothesis that the mere ordering of a judgment task has a predictable effect on the evaluation of peer inputs. The effect seems to emerge irrespective of level of domain expertise or effort invested in making the judgment. In line with research on the phenomenon of naïve realism, participants derogated the judgments of others to the extent that those judgments differed from their own.

Study 4

Study 4 examines our phenomenon is a different context, a medical domain laden with ethical concerns. Specifically, Study 4 asked participants to make and evaluate choices about which of several deserving individuals should receive a life-saving kidney. As with many real-world decisions, there was no clear right answer to this question. Because there is no way to evaluate the "correctness" of such a choice, our dependent variable changed to be the participants' reported evaluation of the overall quality of a particular course of action, and evaluation of the quality of reasoning behind that course of action.

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, individuals who propose the plan that is ultimately chosen often accrue 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

Design and Participants. We employed a single-factor, two-level (Committed, Uncommitted) design. We aimed for a sample size of 400 and successfully recruited 399 participants through mTurk (*M age* = 35, 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.

Procedure. We adapted materials for Study 4 from "The Kidney Case" (Austen-Smith, Feddersen, Galinsky, & Liljenquist, 2014), a simulation designed for teaching about a variety of 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 each candidate 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 Commitment 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 their 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 on7-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 for overall evaluation of the target's choice, our dependent variable.

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 ($M = 4.71$, $SD = 2.15$; $t(397) = 2.31$, $p = .021$). 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; t(397) = 2.17, p = .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 (t(397) = -2.26, p = .025), less reasonable (t(397) = -2.15, p = .033), and less moral (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 (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; 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; t(245) = -0.02, p = .99). The interaction between commitment and objective agreement did not reach significance ($b_{interaction}$ = 0.40, t(395) = -1.23, p = .22).[6]

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

[6] 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).

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).[7] 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], $p < .001$). 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

[7] 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.

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.[8] 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 an elite national security decision-maker 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. Within the constraints of an embedded survey experiment, the scenario is consistent

[8] We based our scenario on public reporting of the raid, not official details (e.g., Mahler, 2015).

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

Design and Participants. As in Study 4, we employed a single-factor, two-level (Committed, Uncommitted) design. 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.[9] We recruited 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 included current and former members of the US Department of Defense, 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

[9] 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.

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.

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 a "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 the U.S.'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 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 judgments 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 ($M_{committed}$= 4.33, SD = 1.70; $M_{uncommitted}$ = 5.01, SD = 1.63; $t(400) = 4.05$, $p < .001$). 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 = 3.59, SD = 1.64), than were Uncommitted experts (M = 4.13, SD = 1.62; $t(162) = -2.11$, $p = .037$).

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; $t(400) = 4.30$, $p < .001$). 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; $t(162) = -0.91$, $p = .362$). 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 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 ($b_{interaction} = 0.94$, $t(398) = -3.31$, $p = .001$).[10] 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; $t(138) = -3.76$, $p < .001$). 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; $t(260) = 2.02$, $p = .044$).

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 who chose to agree with their targets (M = 5.84, SD = 0.93), though this difference did not reach significance ($t(39) = -.41$, $p = .68$). The opposite pattern emerged when experts disagreed with their targets, which again mimicked the pattern in the lay sample ($M_{committed}$ = 3.06, SD = 1.19; $M_{uncommitted}$ = 3.38, SD = 1.25; $t(121) = 1.44$, $p = .151$). The interaction between Commitment and agreement was not significant in the expert sample ($b_{interaction} = 0.46$, $t(160)$ = 1.03, $p = .306$).

[10] 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.

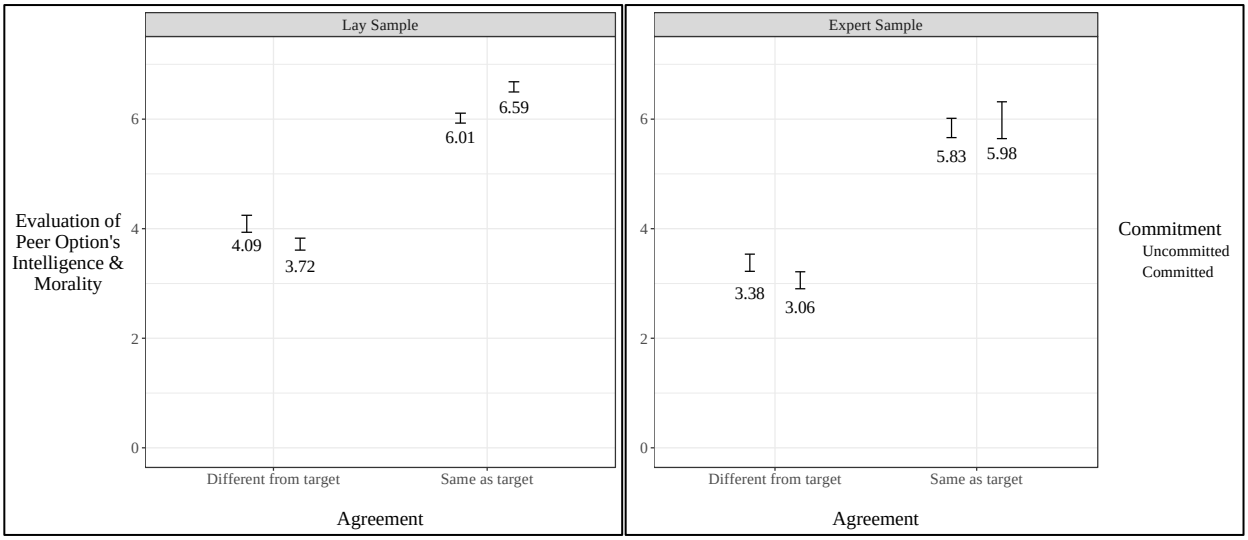

Figure 6. Group means (in text) of participant's evaluations of the target's morality and intelligence, using a composite measure of participant's rating of the target of the peer options intelligence, thoughtfulness, ethicality, and morality. The two panels show the results for the two types of samples 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 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 1-3), 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 2 & 3) 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.

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 judgment and decision-making processes beyond the well-documented assimilation of judgments. 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 their 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. Typically, studies using the JAS paradigm to 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. However, different studies have reached divergent conclusions regarding how disagreement impacts this process, 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. 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 both 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.

Conclusion

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 factor might have a harmful effect on another.

Context

The question of how commitment to one's own ideas affects evaluations of peer input arose from the professional and research experiences of both authors. The first author, an active-duty military officer, noticed that evaluations of proposals in varied domains from budgeting to military strategy (e.g., Study 5) seemed to be more negative to the extent that evaluators had ideas of their own. He was interested in exploring whether this was in fact the case and, if so, what psychological mechanisms accounted for it. The second author, who studies the psychology of disagreement, had long suspected a hidden social cost to the policy that individuals should render independent judgments before working in groups. Both authors were interested in exploring task order as a basic, structural factor influencing group outcomes. Our theoretical starting point was naïve realism, which holds that individuals treat their own understanding of the world as reflecting objective reality. We thought that naïve realists would evaluate others negatively to the extent that those others disagreed. Given anchoring, we suspected that disagreement would vary as a function of the order in which individuals rendered judgments and evaluations. While we felt the logic was clear, the prediction was tentative at the start of the project – we were dealing with novel judgments, while naïve realism was developed in contexts where individuals possessed long-held, often deeply important beliefs. Our studies extended the theory to recently-formed judgments, and to domains ranging from basic consumer estimates to moral dilemmas and military strategy.

References

Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, *193*(5), 31-35.

Austen-Smith, D., Feddersen, T., Galinsky, A., Liljenquist, K. (2010). The kidney case. Evanston: Kellogg School of Management, Dispute Resolution Research Center.

Barnes, C. M., Reb, J., & Ang, D. (2012). More than just the mean: Moving to a dynamic view of performance-based compensation. *Journal of Applied Psychology*, *97*(3), 711.

Blackman, S. F. (2014). *Seeing the subjective as objective: Naïve realism in aesthetic judgments* (Working Paper, Princeton University).

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, *101*(2), 127–151.

Budescu, D. V., & Rantilla, A. K. (2000). Confidence in aggregation of expert opinions. *Acta psychologica*, *104*(3), 371-398.

Budescu, D. V., Rantilla, A. K., Yu, H. T., & Karelitz, T. M. (2003). The effects of asymmetry among advisors on the aggregation of their opinions. *Organizational Behavior and Human Decision Processes*, *90*(1), 178-194.

Burgoon, E. M., Henderson, M. D., & Wakslak, C. J. (2013). How do we want others to decide? Geographical distance influences evaluations of decision makers. *Personality and Social Psychology Bulletin*, *39*(6), 826-838.

De Dreu, C. K. W., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. *Journal of Applied Psychology*, *88*(4), 741–749.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629.

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, *12*(5), 391–396.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological science*, *17*(4), 311-318.

Feng, B., & MacGeorge, E. L. (2006). Predicting receptiveness to advice: Characteristics of the problem, the advice-giver, and the recipient. *Southern Communication Journal*, *71*(1), 67-85.

Frederick, S., Kahneman, D., & Mochon, D. (2010). Elaborating a simpler theory of anchoring. *Journal of Consumer Psychology*, *20*(1), 17–19.

Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology: General*, *141*(1), 124–133.

Gigone, D., & Hastie, R. (1997). Proper analysis of the accuracy of group judgments. *Psychological Bulletin*, *121*(1), 149–167.

Goldsmith, D. J., & Fitch, K. (1997). The normative context of advice as social support. *Human communication research*, *23*(4), 454-476.

Griffin, D. W., & Ross, L. (1991). Subjective construal, social inference, and human misunderstanding. In *Advances in experimental social psychology* (Vol. 24, pp. 319-359). Academic Press.

Heath, C., & Gonzalez, R. (1995). Interaction with others increases decision confidence but not decision quality: Evidence against information collection views of interactive decision making. *Organizational Behavior and Human Decision Processes*, *61*(3), 305-326.

Janiszewski, C., & Uy, D. (2008). Precision of the anchor influences the amount of adjustment. *Psychological Science*, *19*(2), 121–127.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, *1*(1), 54-86.

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

Koehler, D. J., & Beauregard, T. A. (2006). Illusion of confirmation from exposure to another's hypothesis. *Journal of Behavioral Decision Making*, *19*(1), 61-78.

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

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management science*, *52*(1), 111-127.

Liberman, V., Minson, J. A., Bryan, C. J., & Ross, L. (2012). Naïve realism and capturing the "wisdom of dyads". *Journal of Experimental Social Psychology*, *48*(2), 507-512.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, *37*(11), 2098.

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, *108*(22), 9020-9025.

Loschelder, D. D., Friese, M., Schaerer, M., & Galinsky, A. D. (2016). The too-much-precision effect: when and why precise anchors backfire with experts. *Psychological Science*, *27*(12), 1573–1587.

Loy, J. (2017). The Economy of Dogs. *StateCollege*. http://www.statecollege.com/news/columns/the-economy-of-dogs,1472142/.

Luan, S., & Reb, J. (2017). Fast-and-frugal trees as noncompensatory models of performance-based personnel decisions. *Organizational Behavior and Human Decision Processes*, *141*, 29-42.

Mahler, J., (2015). What Do We Really Know About Osama bin Laden's Death? *New York Times*. https://www.nytimes.com/2015/10/18/magazine/what-do-we-really-know-about-osama-bin-ladens-death.html.

Minson, J. A., Liberman, V., & Ross, L. (2011). Two to tango: Effects of collaboration and disagreement on dyadic judgment. *Personality and Social Psychology Bulletin*, *37*(10), 1325-1338.

Minson, J. A., & Mueller, J. S. (2013). Groups Weight Outside Information Less Than Individuals Do, Although They Shouldn't: Response to Schultze, Mojzisch, and Schulz-Hardt (2013). *Psychological science*, *24*(7), 1373-1374.

Minson, J. A., Mueller, J. S., & Larrick, R. P. (2017). The Contingent Wisdom of Dyads: When Discussion Enhances vs. Undermines the Accuracy of Collaborative Judgments. *Management Science*.

Mochon, D., & Frederick, S. (2013). Anchoring in sequential judgments. *Organizational Behavior and Human Decision Processes*, *122*(1), 69–79.

Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, *17*(1), 39-57.

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others. *Psychological Review*, *111*(3), 781–799.

Rader, C. A., Soll, J. B., & Larrick, R. P. (2015). Pushing away from representative advice: Advice taking, anchoring, and adjustment. *Organizational Behavior and Human Decision Processes*, *130*, 26-43.

Robinson, R. J., Keltner, D., Ward, A., & Ross, L. (1995). Actual versus assumed differences in construal: "Naive realism" in intergroup perception and conflict. *Journal of Personality and Social Psychology*, *68*(3), 404–417.

Ross, L., Lepper, M., & Ward, A. (2010). History of social psychology: Insights, challenges, and contributions to theory and application. *Handbook of social psychology*.

Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and misunderstanding. In E. S. Reed, E. Turiel, & T. Brown (Eds.), *Values and knowledge* (pp. 103–135). Hillsdale, NJ: Erlbaum.

Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., ... & Maechler, M. (2015). robustbase: Basic Robust Statistics. R package version 0.92-3.

Simmons, J. P., LeBoeuf, R. A., & Nelson, L. D. (2010). The effect of accuracy motivation on anchoring and adjustment: Do people adjust from provided anchors? *Journal of Personality and Social Psychology*, *99*(6), 917–932.

Sniezek, J. A., & Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational behavior and human decision processes*, *84*(2), 288-307.

Snyder, R. C., Bruck, H. W., Sapin, B., Hudson, V. M., Chollet, D. H., & Goldgeier, J. M. (2002). *Foreign policy decision making*. New York: Palgrave Macmillan.

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 780.

Sunstein, C. R., & Hastie, R. (2015). *Wiser: Getting Beyond Groupthink to Make Groups Smarter*. Harvard Business Press.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis.

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

Vasel, K., (2018). It costs $233,610 to raise a child. *CNN*. http://money.cnn.com/2017/01/09/pf/cost-of-raising-a-child-2015/index.html.

Appendix A – Answer Process in Studies 1 & 2

Cost of dog ownership seven-step estimation process
Step 1: The person estimated how much dog food costs every year.
Step 2: The person estimated how much veterinarian bills cost every year.
Step 3: The person then added estimates from Steps 1 & 2 to get an estimate of annual costs (they were allowed to use a calculator).
Step 4: The person estimated how many years the dog would live.
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.
Step 6: The person estimated the one-time costs of owning the dog, such as collar, bedding, leash, etc.
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.

Table 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”).

Appendix B – Kidney Allocation

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 poly-cystic 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 poly-cystic 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."

Appendix C – 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."