

People Express More Bias in Their Predictions than in Their Likelihood Judgments

Inkyung Park^a, Paul D. Windschitl^a, Jane E. Miller^a, Andrew R. Smith^b,

Jillian O'Rourke Stuart^c, Mark Biangmano^a

^aUniversity of Iowa

^bAppalachian State University

^cVirginia Military Institute

Currently in press at *Journal of Experimental Psychology: General*

OSF: <http://osf.io/ecnjr>

© 2022, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/xge0001258

Author Note

Inkyung Park (<https://orcid.org/0000-0002-1681-6288>), Paul D. Windschitl (<https://orcid.org/0000-0002-4058-3779>), Jane E. Miller (<https://orcid.org/0000-0002-3487-9602>), Andrew R. Smith (<https://orcid.org/0000-0001-5302-3343>), Jillian O'Rourke Stuart (<https://orcid.org/0000-0002-8005-4986>), Mark Biangmano

Data and materials for all studies can be found at:
https://osf.io/ecnjr/?view_only=cda6a7ad3b3c4285adb21925378cc51b.

The preregistration for Study 1 can be viewed at:
https://osf.io/2a9cu/?view_only=f94ae5ebf5e94e388193315e6577aeed.

The preregistration for Study 3 can be viewed at:
https://osf.io/ygm5/?view_only=5b4f776d465d4f0c95b200b32b040149.

The preregistration for Study 4 can be viewed at:
https://osf.io/hk4dm/?view_only=d02e8263151c4fa1bbcc24e8ecca4404.

This work was supported by Grant SES-1851738 to Paul Windschitl and Andrew Smith from the National Science Foundation and the National Institutes of Health T32 pre-doctoral training grant (T32GM108540 supporting Inkyung Park).

Correspondence concerning this article should be addressed to Inkyung Park, Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA, 52242. E-mail: inkyung-park@uiowa.edu

Declarations of interest: none.

Abstract

The *desirability bias* refers to when people's expectations about an uncertain event are biased by outcome preferences. Prior work has provided limited evidence that the magnitude of this motivated bias depends on (is moderated by) how expectations are solicited—as discrete outcome predictions or as likelihood judgments expressed on more continuous scales. The present studies extended the generalizability and understanding of the moderating process. The authors proposed that solicitations of predictions and likelihood judgments have different connotations that ultimately affect how much bias is expressed; this varies from a prior account that attributed the moderation effect to response scale differences (dichotomous vs. continuous). Study 1 confirmed the connotation difference, with predictions being viewed as more affording of hunches. Studies 2-4 directly tested the moderation effect, and unlike prior work focusing on expectations for purely stochastic events, the present studies involved more naturalistic events for which likelihood information was not supplied or directly knowable. Before viewing scenes from a basketball game (Study 2) or an endurance race (Studies 3-4), participants were led to prefer one contestant over another. After viewing most of the closely-fought contest, they made either a prediction or likelihood judgment about the outcome. Participants' tendency to forecast their preferred contestant to win was significantly stronger among those making predictions rather than likelihood judgments. In support of the proposed account, this effect persisted even when both types of solicitations offered only dichotomous response options. Broader implications for measuring and understanding people's expectations/forecasts are discussed.

Key Words: desirability bias, wishful thinking, uncertainty, likelihood judgment, motivated reasoning

People Express More Bias in Their Predictions than in Their Likelihood Judgments

People's expectations or forecasts about how uncertain events will turn out can be solicited in at least two types of ways—as predictions or as likelihood judgments (Howell & Burnett, 1978). Queries about outcome predictions ask for a belief about what will happen. For example, in predicting the outcome of a game between Teams A and B, people might have to respond with either “A wins” or “B wins.” Queries about likelihood judgments ask people to estimate the likelihood of a specific possible outcome or to compare the likelihoods of possible outcomes. The responses might involve numbers (0-100%) or words (e.g., Team A is very likely to win, or Team A is much more likely than B).

Perhaps it goes without saying, but probabilities should inform predictions.¹ If a person thinks Team A has a 60% chance of beating Team B, that person should logically predict that Team A will win, not Team B. Indeed, researchers have often assumed that people's predictions are derived from subjective probabilities (e.g., see Kahneman & Tversky, 1982). The assumption is reflected in a standard paradigm for measuring overconfidence, when participants are given only the upper half of the probability scale (50-100%) to express their confidence about whether their prediction/answer for a binary item was correct (Brenner et al., 1996; Dunning et al., 1990; Lichtenstein et al., 1977; but see Sieck & Yates, 2001; Juslin et al., 2000). The lower half of the scale is not always included, because it is assumed that participants would have given the alternative prediction if their confidence was less than 50%.

However, there are findings that question the assumption that predictions are straightforwardly derived from subjective probabilities. For example, when people make

¹ We will use the term “subjective probability” in a slightly more general way than “likelihood judgment”—with the former referring to an internal belief and the latter a judgment to be communicated or indicated as response.

predictions across a set of repetitions of a binary event, their responses often reflect a probability matching strategy (i.e., predicting events proportional to the probability rates of two events) rather than a probability maximizing strategy (i.e., always predicting the event that is known to have the higher probability rate) (Gaissmaier & Schooler, 2008; Koehler & James, 2009, 2010). Also, different methods of assessing subjective probabilities are differentially predictive of behavior and reveal distinct facets of how people think about uncertainty-relevant phenomena (Howell & Burnett, 1978; Kahneman & Tversky, 1982; Windschitl & Chambers, 2004; Windschitl & Wells, 1998). Given this, it seems overly simplistic to assume that people's predictions are always solely derived from subjective probabilities.²

The present work fits in the vein of findings that suggest that predictions are not merely derived from subjective probabilities. Our interest is specifically in how predictions deviate from likelihood judgments when motivated concerns are in play because people prefer one outcome over the other. The key finding, as suggested by the paper's title, is that under specific circumstances, people express more bias in their predictions than in their likelihood judgments.

The Impact of Motivated Factors

The notion that people's expectations about future events might be biased by their outcome preferences is known as the *desirability bias* or *wishful thinking*. The relationship between outcome preferences and likelihood beliefs have been studied in both correlational and

² Another empirical example suggesting that predictions do not always follow subjective probabilities is rather specific but also quite interesting. It is from work on industry experts' forecasts about the success of petroleum well exploration (Milkov, 2020). Dichotomous predictions about whether a given well exploration would be successful did not always reflect the same experts' probability estimates of the success. Many experts predicted the well explorations to be successful (implying a probability estimate of 50% or above), even when their probability estimates were less than 30%.

experimental designs (e.g., Babad, 1987; Harris & Hahn, 2011; Hayes Jr, 1936; Krizan et al., 2010; Krizan & Windschitl, 2007; Markman & Hirt, 2002; Massey et al., 2011). Here, we focus on experimental studies to avoid causal ambiguities that can affect interpretations of correlational studies (for discussion, see Windschitl & Stuart, 2015). In studies that use experimental manipulations of outcome desirability to examine their influence on expectations about impending events, the results are decidedly mixed (Bar-Hillel & Budescu, 1995; Bar-Hillel et al., 2008b; Vosgerau, 2010). Most important for the present work are studies that have compared different ways of measuring expectations and what they suggest about the magnitude of desirability biases. The common finding is that when expectations are solicited as discrete outcome predictions, there are robust desirability biases detected. However, when expectations are solicited as likelihood judgments, the tests for the desirability biases are often not significant (Bar-Hillel & Budescu, 1995; Pruitt & Hoge, 1965; but see Stuart et al., 2017). A meta-analytic comparison of the magnitude of desirability biases detected with the two types of measures revealed a significant difference (Krizan & Windschitl, 2007). The potential difference in results between discrete outcome predictions and likelihood judgments is of substantial theoretical importance. It suggests quite different conclusions about whether people's beliefs or expectations about the future are readily vulnerable to motivated distortions.

Unfortunately, however, the experiments that have compared how different measures impact the desirability bias have relied almost exclusively on one type of paradigm—the marked-card paradigm—or close variants of it (Irwin, 1953; Marks, 1951; Price & Marquez, 2005; Windschitl et al., 2010). In the marked-card paradigm, participants are told how many cards in a deck have a particular marking (which is either desired or not depending on condition) before they are asked about their expectations regarding a card draw. That is, relevant probabilities are directly supplied to respondents; this is also true of variants of the paradigm.

The fact that these prior studies have always provided specific probability information is quite limiting, especially because across many contexts in everyday life, people need to form expectations when probabilities are not simply provided.

Our studies address this crucial gap. We examined whether outcome predictions and likelihood judgments are differentially affected by motivated factors, but we do this for events quite different from card draws in a marked-card paradigm. The events were more naturalistic and did not involve known probabilities. In one study, for example, people were asked about a college basketball game. Despite using different paradigms, we predicted our results would be consistent with results from the marked-card paradigm—namely that motivated factors would have a stronger impact on predictions than likelihood judgments. As described in the next section, our work advances new ideas about why this might happen.

Why Predictions and Likelihood Judgments Might be Differentially Biased

In a prior publication, the *biased-guessing account* was proposed to explain why, in the marked-card paradigm, predictions were more impacted than likelihood judgments by outcome desirability (Windschitl et al., 2010). The account suggested that when people construe their forecast as a guess or as subjectively arbitrary, they would tend to be biased in an optimistic direction. In a stochastic circumstance like that encountered in the marked-card paradigm, unless the evidence is strongly favoring one outcome over another, dichotomous predictions about whether a drawn card would be marked or not can have an arbitrary feel. According to the account, this led to, or allowed for, optimistic predictions. However, the account also suggested that, when a likelihood-judgment query is accompanied by a continuous response scale (as they typically are), the continuous nature of the scale removes the arbitrary feel of the response and hence removes the tendency for people to respond in an optimistic way (Stuart et al., 2017; Windschitl et al., 2010).

Our updated account embraces the general outline of the biased-guessing account, but it differs in an important respect. Whereas the biased-guessing account assumes that the continuous scaling of the response options for a likelihood-judgment query is the crucial reason for observed differences in the impact of bias on predictions vs. likelihood judgments, our updated account does not consider this scaling issue to be crucial. It assumes that, irrespective of this type of scaling or response-option distinction, queries about predictions and likelihood judgments carry slightly different connotations that shape their vulnerability to motivated bias. Specifically, we propose that whereas people broadly understand that answers to queries about both likelihood judgments and predictions require careful assessments of the available evidence, they also perceive queries about predictions as allowing for relatively more impact from hunches or other arbitrary considerations.

By definition, predictions require people to anticipate the outcome as a discrete state (e.g., Team A wins, Team B wins). In most situations in which a prediction is requested, there is at least some stochasticity involved in the determination of the target outcome. This means the outcome is partly indeterminate. There is no prediction that can be guaranteed to be correct, and even the same prediction would vary from being correct vs. incorrect across repeated plays of the same situation. Given all this, we suggest that people are accustomed to thinking of predictions as affording some arbitrariness, such as a hunch or guess. In contrast, likelihood judgments are not typically viewed as affording this arbitrariness, and we propose this is true even when a likelihood judgment is solicited on a dichotomous “scale” (e.g., Team A is more likely to win, Team B is more likely to win). Despite the fact that, from a formal perspective, asking for this type of likelihood judgment on a dichotomous scale is equivalent to asking for a prediction, people will perceive this likelihood question as more constrained to focus only on an assessment of currently available evidence.

This idea that predictions and likelihood judgments evoke different emphases is not entirely new. Howell and Burnett (1978) speculated that predictions and probability estimates about a single, upcoming outcome—what they called a short-span situation—could differ because the probability estimate would, to a greater degree than a prediction, reflect beliefs about what the evidence implies about outcomes over longer spans (e.g., outcomes over 100 repetitions). They viewed predictions as more open to unstable factors, including momentary feelings or hunches.³ We do not necessarily agree that the reason probability estimates would be different from predictions is because of what they imply about longer spans, but we do agree that soliciting probability estimates, as opposed to predictions, could focus people more on the available evidence.

We further hypothesize that because predictions (vs. likelihood judgments) are viewed as slightly more accommodating of hunches or other arbitrary considerations, they are also more vulnerable to influence from extra-evidentiary motivations. By “extra-evidentiary motivations,” we are referring to motivated factors tied to outcome desirability. The motivation to feel good about the future, to be an optimistic person, or to lend symbolic support to a preferred side in a competition are all examples of motivated factors that could bias people’s predictions toward the desired outcome (Anderson et al., 2012; Armor et al., 2008; Helweg-Larsen et al., 2002; Miller et al., 2021; Morewedge et al., 2018; Taylor & Brown, 1988; Tyler & Rosier, 2009).⁴ Past work

³ Yet, Howell and Burnett (1978) also included a comment that seems in tension with this view: “Since prediction is basically an all-or-none response in the case of short spans, we must assume that the predicted outcome is perceived by the observer as more likely than any of the alternatives at that moment.” (p. 63).

⁴ Although we focus on the instances in which desire inflates optimism, there is also work suggesting that extra-evidentiary factors could exert pessimistic influences over people’s forecasts (Bilgin, 2012; Bonnefon & Villejoubert, 2006; Harris et al., 2009; Harris & Corner, 2011; Risen & Gilovich, 2008; Vosgerau, 2010). In these studies, an asymmetric loss function is

supports this step in our account, which is also shared by the biased-guessing account. Studies have shown that instructional manipulations that explicitly encourage a reliance on hunches vs. statistical assessments produce more optimistic forecasts (Rose & Aspiras, 2020; Windschitl et al., 2010). However, no studies have tested whether, with scaling differences neutralized and without explicit encouragement of hunches or assessments, the impact of outcome desirability would be greater when expectations are solicited as predictions rather than as likelihood judgments.

It is important to emphasize the key distinction between our updated account and the original biased-guessing account. The biased-guessing account explicitly attributed the observed differences in motivated bias (in the marked-card paradigm) to the fact that predictions have dichotomous response options whereas likelihood judgments involve response options that are more continuous (Windschitl et al., 2010). This is not assumed in the updated account. Instead, the account suggests that queries about predictions and likelihood judgments are distinct enough to yield different vulnerabilities to motivated factors, even when the response-scale differences are removed.

Overview of Current Studies

As a set, the studies presented here tested whether—in paradigms that are more naturalistic than the marked-card paradigm—predictions are more vulnerable to motivated bias than are likelihood judgments. The studies also tested hypotheses from an updated account as to why this might be true. We present seven studies in all: Studies 1-4 in the main text and Studies A, B and C in the Supplemental Materials. The first study in the main text was a scenario study

often stipulated, whereby the harm of being too optimistic is considered greater than the harm of being too pessimistic (Haselton & Nettle, 2005; Hertwig & Fox, 2006; Sweeny et al., 2006).

showing that people are more inclined to associate hunches with predictions than with likelihood judgments. Studies 2-4 involved actual events in which preferences for target outcomes were manipulated and expectations were solicited. In Study 2, participants were shown footage from near the end of a college basketball game and were asked for either a prediction or likelihood judgment about the outcome. Critically, both the prediction question and likelihood-judgment question offered the same number of response options (i.e., two). Despite both formats offering dichotomous options, the desirability bias was larger for predictions than for likelihood judgments. Study 3 replicated those findings using a new competition context. Consistent with a preregistered expectation, it also showed that the enhanced vulnerability of predictions (vs. likelihood judgments) to the desirability bias was attributable to participants whose desired outcome was the less probable one. Finally, Study 4 addressed an alternative explanation for our findings.

Transparency and Openness

For all studies, we justify our sample sizes and report all data exclusions, manipulations, and measures. The reporting is consistent with Journal Article Reporting Standards (JARS) (Kazak, 2018). All data, analysis code, and research materials are available at https://osf.io/ecnjr/?view_only=cda6a7ad3b3c4285adb21925378cc51b. For all studies, data were analyzed using IBM SPSS Statistics for Windows, version 27.0 (IBM Corp., Armonk, N.Y., USA). All studies were conducted under the approval of the University of Iowa Institutional Review Board.

Study 1

Our account for why predictions and likelihood judgments might be differentially influenced by motivated concerns starts with the notion that they hold slightly different connotations. We propose that whereas people broadly understand that answers to queries about

both likelihood judgments and predictions require careful assessments of the available evidence, they also perceive queries about predictions as allowing for relatively more impact from arbitrary considerations like hunches. Study 1 tested this claim using a short hypothetical scenario.

Participants imagined being asked by a coworker for a forecast—which was either a prediction or likelihood judgment. We simply asked them to indicate the extent to which the request seemed to be asking for an assessment or a hunch. We expected that participants would perceive the coworker’s question as more about their hunch when it solicited a prediction than when it solicited a likelihood judgment. Critically, we phrased both forecast types as dichotomous.

Methods

Participants and Design

A total of 118 participants were recruited on Amazon Mechanical Turk through CloudResearch (Litman et al., 2017). They participated in a 2-cell within-subject design, which also included a counterbalancing described below. Consistent with our pre-registration, we excluded data from 17 participants who gave nonsensical responses to an open-ended question in a task that followed this study. This left a final sample size of 101, which provided over 80% power to detect a small difference ($d = .25$) between two within-subject cells in a one-tailed test. The design, hypotheses, and analysis of the study were preregistered on OSF (https://osf.io/2a9cu/?view_only=f94ae5ebf5e94e388193315e6577aeed).

Procedure and Material

After consenting to participate, participants read:

Imagine a coworker asks you a question about an upcoming football game between Team A and Team B. It’s a close matchup. The teams have similar records, and either team could win this upcoming game. For this close matchup, the coworker asks:

At this point, participants read a coworker's question; it was either for a dichotomous prediction ("Which team will win? Will team A win or will Team B win?") or for a dichotomous likelihood judgment ("Which team is more likely to win? Is Team A more likely to win or is Team B more likely to win?"). The participant was then asked whether the coworker seemed "to be asking for your assessment or for your hunch?" on a 5-point scale anchored by "Definitely asking for an assessment" and "Definitely asking for a hunch."

On the subsequent screen, participants read "Now, please imagine that the co-worker had asked a slightly different question." The other type of forecast question was then presented and rated on the same assessment-hunch scale. The order of the prediction and likelihood questions was counterbalanced, as was the positioning of "assessment" and "hunch" in the rating question.

Result

We recoded all ratings on the assessment-vs.-hunch scale such that higher values reflected the hunch side. As predicted, a preregistered, paired samples t-test revealed that ratings on the assessment-vs.-hunch scale differed as a function of whether participants were rating the prediction question or the likelihood question, $t(100) = 2.18, p = .016$ one-sided, $d = .22$, 95% CI [.04, .87]. Specifically, the extent to which participants interpreted a query to be more about a hunch (vs. assessment) was greater when the query asked for a prediction ($M = 3.34, SD = 1.40$) than when it asked for a likelihood judgment ($M = 2.88, SD = 1.51$).⁵ This effect, although modest in size, supports the notion that questions about predictions and likelihood judgments have slightly different connotations in the eye of respondents, even when both are described as having dichotomous response options.

⁵ A significant Forecast x Order interaction also revealed that the influence of forecast was greater when people responded about a prediction before responding about a likelihood judgment, $F(1, 97) = 6.36, p = .013, \text{adj } \eta_p^2 = .06$.

A slightly different way of describing the findings from Study 1 is that participants interpreted likelihood judgments as relatively more tied to assessments of evidence than are predictions. A short follow-up study, which is detailed in the Supplemental Materials, provided additional evidence for this. In a hypothetical scenario, a protagonist was said to claim that the evidence for an upcoming game favored one team over the other, but the protagonist nonetheless made a forecast in the opposite direction. Participants rated this sequence to be more sensible (less illogical) when the forecast was in the form of a dichotomous prediction than in the form of a dichotomous likelihood judgment.

Study 2

Study 1 established that predictions and likelihood judgments can be perceived as having connotations that differ in whether they accommodate a hunch. Past work that manipulated whether people were explicitly encouraged to base their forecasts on hunches vs. statistical assessments has shown that the desirability bias is larger when the former is encouraged (Rose & Aspiras, 2020; Windschitl et al., 2010). However, no studies have tested whether, with scaling differences neutralized and without explicit encouragement of hunches or assessments, the impact of outcome desirability would be greater when expectations are solicited as predictions rather than as likelihood judgments. This was addressed in Study 2.

Critically, we used a more naturalistic paradigm than the marked-card paradigm in which participants are provided with numeric summaries for evidence about the possible outcomes (Price & Marquez, 2005; Windschitl et al., 2010). The new paradigm centered around a college basketball game, for which participants viewed video footage. Another crucial feature of Study 2 was that we removed major scale differences in how participants were asked to provide either predictions or likelihood judgments. Regardless of whether people were asked for a prediction or likelihood judgment, their response options were dichotomous (the likelihood question asked

which team was more likely to win, with only two response options). For reasons discussed earlier, the original biased-guessing account would not anticipate differences in vulnerability to bias between the predictions and likelihood judgments in this study, given that both forecast formats involved a dichotomous response format. Evidence for differential vulnerability to bias would support our updated account.

The basketball-game paradigm used here was adapted from Stuart et al. (2017). Participants watched scenes from a closely-fought game. A manipulation was used to make one team winning more desirable. The scenes that were viewed led up to the final moments of the game when there was uncertainty about which team would win. Stuart et al. (2017) used the paradigm to show that people often chose to prepare as if a preferred (vs. not preferred) outcome will happen, but they did not address what we tested here—whether predictions vs. likelihood judgments are differentially influenced by bias.

Methods

Participants and Design

Two hundred and sixty-nine undergraduate students at the University of Iowa participated in a 2 (Assigned Team: Red, White) \times 2 (Forecast Format: Prediction, Likelihood Judgment) between-subjects design. We note that Study 2 combines the data from two studies (2a and 2b). Study 2b was a direct replication of Study 2a, and the procedure, materials and dependent measures used in the study were identical to Study 2a. Here we describe and report both studies combined as Study 2, but in the Supplemental Materials we also include per-study breakdowns for main findings. The sample size of 269 allowed 80% power on a one-tailed test to detect a 50% vs. 65% difference between the proportion of people forecasting their preferred team in the prediction versus likelihood. Neither 2a nor 2b were pre-registered.

Procedure and Material

Participants were instructed that they would be watching video footage of a basketball game in which one team was wearing red and the other white. They were told that they will be assigned to one of the two teams, and if “their” team won the match, they would get a candy bar. Then the experimenter flipped a coin to determine participants’ team, one side assigning them to the red team and the other to the white. They were shown a basket of candy bars and reminded that if their team won, they would be able to pick one. Participants then watched about 6 minutes of video footage from the game.

The game was from the 2010 tournament of the Women’s Northeast College Basketball Conference. This particular game was used because it was unlikely that participants knew the outcome and because the score was very close at the end. When there were just 10 seconds left in the game and the white team had the ball but was down by 1 point, the video stopped, and participants completed the main dependent measure followed by others. Then they saw the ending and finished the rest of the study, including the receipt of the candy bar when warranted.

Main Dependent Measure (the Forecast)

Participants in the prediction condition were asked “What is your prediction about the outcome of the game? That is, which team do you think will win?” and indicated their choice between the two options “Red Team” or “White Team.” Those in the likelihood judgment condition were asked “Is it more likely that the red or white team will win? Please indicate your best assessment by clicking on one of the options below.” They made a choice between two options that read “More likely that the RED team will win” and “More likely that the WHITE team will win.”

Other Measures

After the main measure, participants completed various other measures that were not of primary interest and for which reporting can be found in the Supplemental Materials. These

measures included a request for a rationale for their forecast, the same behavioral measure used by Stuart et al. (2017), 7-point likelihood judgments, demographics, reactions to the task, and three trait scales.

Results

Preliminary Analyses

Responses to a manipulation check that participants encountered near the end of the study confirmed that desire was successfully manipulated. When asked to indicate which team they hoped to win (1=definitely hope red wins to 5=definitely hope white wins), participants who were assigned to the white team as “theirs” responded significantly higher ($M = 4.21$, $SD = .89$) than those assigned to the red team ($M = 1.70$, $SD = .88$), $t(267) = 23.13$, $p < .001$, $d = 2.82$, 95% CI [2.30, 2.72].

Main Analyses of Forecasts

An overall analysis collapsing across the format conditions revealed a significant desirability bias, with participants tending to forecast a win for their own team (62.1%) rather than the other team (binomial test, $p < .001$). Recall, however, that our main interests were in whether the desirability bias was significant within the prediction condition and whether it would be significantly smaller in the likelihood judgment condition compared to the prediction condition.

As expected, within the prediction condition, there was a desirability bias, with significantly more than half of the participants (75.4%) predicting that their team would win (binomial test, $p < .001$).⁶ The left section of the Figure 1 breaks this down in a more specific

⁶ This study also included an exploratory behavioral-choice measure adapted from Stuart et al. (2017), which was solicited after the main forecast measures. Details about that measure and its

way. The proportion of people predicting a red win was significantly higher among people assigned to the red team (76.3%) than among people assigned to the white team (25.9%), $X^2(1) = 33.13, p < .001$.⁷

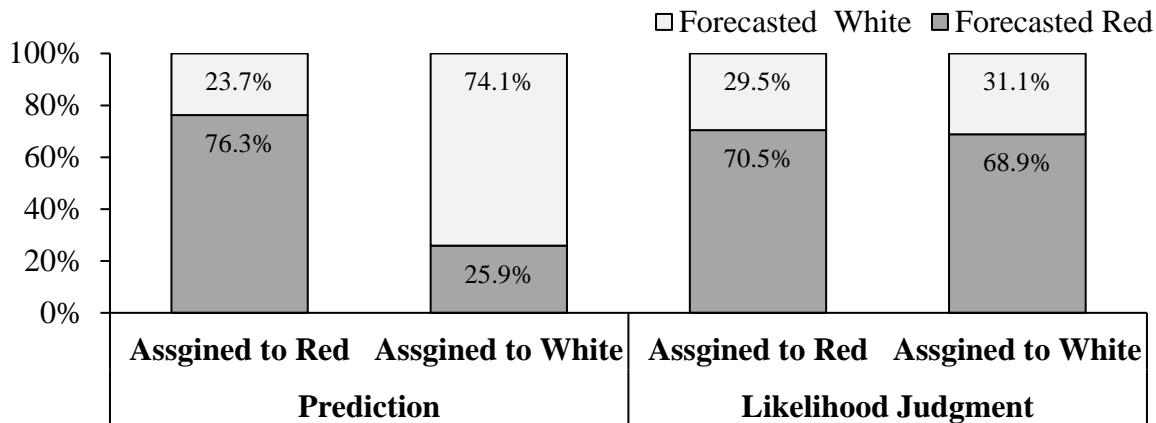


Figure 1. From Study 2, the proportions of responses in which participants forecasted either the white (in light gray) or red team (in dark gray) winning the game, as a function of the assigned team (i.e., whether the participant's team was the white or red team) and forecast format (i.e., whether participant made a prediction or likelihood judgment). Note that each vertical bar represents all responses from participants assigned to a given cell; therefore, the two numbers per bar add to 100%.

Crucially, and as expected, the desirability bias was significantly larger in the prediction condition than in the likelihood-judgment condition. This conclusion was supported by a significant Forecast Format x Assigned Team interaction in logistic regression, with the forecasted team as the dependent variable (0 = forecasted white, 1 = forecasted red team), $\beta = -2.14, SE = .55, Wald \chi^2(1) = 14.93, odds ratio (OR) = 8.51, p < .001$. Said differently, the rate of forecasting the red team to win was more strongly affected by team assignment among people in

results can be found in the Supplemental Materials, but here we simply note that the behavioral-choice results followed the prediction results very closely.

⁷ The binomial test and chi-square tests in this paragraph are conceptually overlapping. However, due to noise in random assignment, more people in the prediction condition were assigned to the red team ($n=141$) than to the white team ($n=128$). This makes the latter of the two inferential tests more relevant—although both suggest the same conclusion.

the prediction condition (see left half of in Figure 1) than among people in the likelihood condition (see right half).

The simple effect of desirability was not significant in the likelihood judgment condition, with only 48.9% of the participants forecasting their team to win (binomial test, $p = .863$). More specifically, the rate at which participants gave a greater likelihood for the red team winning was about the same whether their desired team was the red team (70.5%) or the white team (68.9%), $X^2(1) = .04, p = .843$.⁸

Overall, participants were more likely to give a forecast that favored the red team (62.8%) rather than the white team (37.2%) (binomial test, $p < .001$). This is not terribly interesting because it is a natural consequence of one team—the red team—appearing to be in a slightly stronger position than another at the point at which the video stopped. However, this interfaces with another finding in an interesting way. The other finding comes from the alternative decomposition of the aforementioned Forecast Format x Assigned Team interaction. It reveals that forecast format had a larger effect on people assigned to the white team rather than the red team. For people assigned to the red team, they gave similar forecasts regardless of the forecast-format condition (see first and third bars of Figure 1), $X^2(1) = .59, p = .441$. For people assigned to the white team (see second and fourth bar of Figure 1), they were more likely to forecast a white win when giving predictions than when giving likelihood judgments, $X^2(1) = 23.09, p < .001$. Although we did not anticipate this particular finding prior to running the study, it is potentially interpretable as follows. Again, the red team was generally viewed as the team in the better position to win. Among people assigned to that better-positioned team, most

⁸ This study included an exploratory pair of numeric likelihood questions after the set of main dependent variables and other exploratory measures. Details about that measure and its results can be found in the Supplemental Materials, but here we simply note that the main effect of desirability on a composite from these measures was not significant.

participants would forecast that team to win regardless of whether they are asked for a likelihood judgment or asked for a prediction. Among people who were assigned to the worse-positioned team (the white team), they might predict their team to win even though a solicitation of a likelihood judgment would elicit the objective assessment that the other team has the better chance.

Discussion and Follow-Up Study

In two ways, the findings from Study 2 extend past work that used variants of the marked-card paradigm (Price & Marquez, 2005; Windschitl et al., 2010). First, it shows that the effect of forecast format (i.e., the greater vulnerability of predictions vs. likelihood judgments to motivated bias) holds even when people are asked about naturalistic events for which relevant evidence is not explicitly provided in summarized numeric values. Second, we used a dichotomous scale for soliciting both predictions and likelihood judgments. This supports our updated account and extends the implications of the current findings beyond a distinction between discrete vs. continuous response scales.

However, there were two features of the wording used in Study 2 that require more scrutiny. It could be argued that there were subtle wording confounds. In the likelihood judgment condition, the use of the pronoun “it” at the beginning of the likelihood question suggests an impersonal, third-person perspective (“Is it more likely...”). Near the beginning of the prediction question, the word “your” was used rather than “it” (“What is your prediction...”). On an intuitive basis, we did not think this was a critical difference, but we recognized the general plausibility of this concern, given work by Juanchich et al. (2017) and Løhre and Teigen (2016). They showed that subtle differences in interpretation and usage of probability statements made from a first- or third-person perspective. For example, a statement made from a first-person perspective (“I am certain that...”) versus an impersonal third-person perspective (“It is certain

that...”) is interpreted as reflecting more internal than external uncertainty. Another potential confound was that in the likelihood-judgment condition, but not the prediction condition, the wording instructed people to give their “best assessment.”

Because it was logically possible that these confounds were the key drivers of the effect in Study 2, we ran a small follow-up study ($N = 53$) in which everyone was asked for a prediction about the basketball game’s outcome. However, the wording we used for the prediction question was modified. It started from an impersonal perspective and included the words “best assessment.” Specifically, people were asked “*Will the red or white team win? Please indicate your best assessment by clicking one of the options below: ‘The red team will win’ / ‘The white team will win.’*” If either of the two potential confounds explained the results of Study 2, we’d expect the predictions in this follow-up to show no desirability bias—as in the likelihood-judgment condition of Study 2. However, the desirability bias was approximately as strong as in the prediction condition in Study 2. Overall, 77.3% of participants predicted that their assigned team would win the game, significantly greater than 50% (binomial test, $p < .001$). Caution must be emphasized about making across-study comparisons, but these findings appear to rule out the possibility that the potential confounds explain the effects in Study 2.

Study 3

Again, the fact that Study 2 used a naturalistic event and involved only dichotomous measures makes it an important extension of previous work that used a marked-card paradigm (Price & Marquez, 2005; Windschitl et al., 2010). However, testing the generalizability beyond just one naturalistic event would add further credence to the observed phenomena. Therefore, we conducted additional studies that centered on participants’ forecasts about a completely different type of competition. Specifically, we asked people for forecasts about the outcome of a Spartan Race. Spartan Races are obstacle-laden endurance races that often have hundreds of recreational

participants but also elite athletes; the races have sometimes been nationally televised in the United States. The particular race we used in our studies featured a close competition between two athletes.

Three initial studies that we conducted with the new paradigm produced results that were, overall, quite compatible with the effects detected in Study 2. A full description of that work can be found in the Supplemental Materials (Studies B1-B3). Here, as Study 3, we present a well-powered, pre-registered follow-up to those studies, yielding the same conclusions as those studies.

A critical feature of Study 3 and its preregistration relates to the unanticipated but explainable finding in Study 2, where the impact of the forecast format was driven by people assigned to the team that was—on a sample-wide basis—viewed as less likely to win. For the race featured in our Spartan Race paradigm, one runner (named Hobie) held a slight and steady lead over the other featured runner (Cody), causing Cody to be generally perceived as less likely to win (an *underdog*; Goldschmied & Vandello, 2012). Therefore, it is logical to expect that the impact of the format manipulation would be driven by participants who were assigned to Cody as the desired athlete. Among these people, they might predict their athlete to win even though a solicitation of a likelihood judgment would elicit the objective assessment that the other athlete has the better chance. Among people assigned to the “overdog” Hobie, who led the race, most participants would forecast him to win regardless of whether they are asked for a likelihood judgment or asked for a prediction. Consequently, a key preregistered prediction for Study 3 was that the effect of the forecast format would be exclusively found among participants assigned to the underdog, Cody.

Another new feature of this study was that we added continuous likelihood judgments in a third format condition. We did this because we thought some portion of participants might

interpret the dichotomous-judgment questions as essentially a prediction, but that this type of interpretation is made less plausible by the presence of a continuous response scale.

In sum, we preregistered that there would be a desirability bias in the prediction condition, that the bias would be smaller in the dichotomous likelihood judgment condition, and that the bias would be even smaller in the continuous likelihood judgment condition. However, as noted, we also predicted that the effect of the forecast format would be exclusively found among participants assigned to the underdog, Cody.

Methods

Participants and Design

We preregistered a sample size of 324 participants, the design, hypotheses, and analyses on the OSF (https://osf.io/ygmx5/?view_only=5b4f776d465d4f0c95b200b32b040149). After completion of all posted sessions, our final sample was comprised of 331 University of Iowa undergraduates. The design was a 2 (Assigned Athlete: Cody vs. Hobie) x 3 (Forecast Format; Dichotomous Prediction vs. Dichotomous Likelihood Judgment vs. Continuous Likelihood Judgment) between-subjects design. Although the preregistration mistakenly claims otherwise, we did not include the right/left counterbalancing of athlete names.⁹ After dropping some participants (described below), the remaining sample size of 314 provided 84% power on a one-tailed test to detect a 50% vs. 75% proportion difference between two cells.

Procedure and Material

Participants were given introductory information about Spartan Races before learning they would be viewing scenes from one race, focusing on the top two competitors. They were also informed that they will be assigned to one of the athletes. First, however, the desirability

⁹ Counterbalancing of athlete names was included in other studies using this paradigm (see Study 4 and the Supplemental Materials), where it did not have a significant impact on results.

manipulation was also introduced. Participants were told that they would be doing either an appealing or unappealing secondary task depending on who won the race. If their athlete won, they would rate pictures of sports game winners or animal memes, but if their athlete lost, they were to complete a narrative writing task on their experience of losing a competition or opportunity.

For the athlete-assignment process, participants were offered a blind choice between two boxes labeled “Athlete A” and “Athlete B.” The name of the athlete they apparently choose was then revealed along with the other athlete’s name. Although this process hinted to participants that their blind selection determined the athlete assignment, they were assigned to either Hobie or Cody at random.

Once assigned an athlete, participants viewed information about their athlete and his competitor (e.g., age, height, weight, jobs, and sponsors). Then, they were presented with a series of still images captured from a video of the race between Cody and Hobie.¹⁰ Hobie was always in the lead throughout the race while Cody was in a close second. When the presentation of images reached the last obstacle in the race (with Hobie still in the lead), participants moved on to complete the main dependent measures and other measures (see descriptions in next two sections). After the completion of the measures, participants learned the winner of the race (Hobie). Depending on the assigned athlete, participants moved on to do either the appealing or unappealing secondary task.

Main Dependent Measure (Forecast)

¹⁰ A key reason why we chose a Spartan Race was because we could show people still images that meaningfully showed the progress of the race. We selected this race in particular because it quickly devolved into a close two-person competition.

Depending on condition, participants indicated their expectation about the outcome of the race either as either a dichotomous prediction, dichotomous likelihood judgment, or continuous likelihood judgment. Table 1 shows the exact wordings and options used in each condition. Critically, we avoided the possible wording confounds that were mentioned in the Discussion section for Study 2.

Table 1

Prompt and the Response Options Used in Study 3 for Each Forecast Format Condition

Forecast Format	Prompt	Response Options
Dichotomous Prediction	Which competitor will win the race?	Hobie Call will win, Cody Moat will win
Dichotomous LJ	Which competitor is more likely to win the race?	Hobie Call is more likely to win, Cody Moat is more likely to win
Continuous LJ	Which competitor is more likely to win the race?	Hobie Call is much more likely to win (1) - Cody Moat is much more likely to win (6)

Note. LJ = Likelihood Judgments.

Other Measures. After the main measure, participants completed other measures that were not of primary interest and for which reporting can be found in the Supplemental Materials. These measures included a short rationale for their forecast, items about prior knowledge of Spartan Races and competitors, likelihood estimates on a 0%-100% scale, and items about who participants hoped would win the race and how much they cared.

Participants also answered memory checks to ensure that they had clear knowledge about which athlete they were assigned to and whether they paid adequate attention during the study.

Results

Preliminary Analysis

As preregistered, we excluded participants (5.1%, $n = 17$) who failed the memory check of which athlete they were assigned to, leaving 314 participants for analysis.

The manipulation check confirmed that participants' desire was successfully manipulated. When asked to indicate which athlete they hoped to win (1= 'strongly hope Hobie wins' to 7= 'strongly hope Cody wins'), participants who were assigned Cody as "theirs" responded significantly higher ($M = 5.75$, $SD = 1.42$) than those assigned to Hobie ($M = 2.13$, $SD = 1.28$), $t(310) = 23.71$, $p < .001$, $d = 2.69$, 95% CI [3.32, 3.92].

Main Analyses of Forecasts

To allow for direct comparisons between data from the continuous likelihood judgment condition and data from the other conditions, we first dichotomized the continuous likelihood judgments, using the midpoint of the 6-item scale as the cut-point. This dichotomized version was used in the main analyses that follow, but see Supplemental Materials for the original, non-dichotomized values. In an analysis collapsing across all conditions, we found a significant desirability bias with participants tending to forecast a win for their own athlete (71.8%) rather than the other athlete (28.2%) (binomial test, $p < .001$). Within just the prediction condition, the desirability bias was quite robust, with 82.4% of participants predicting that their athlete would win, $X^2(2) = 44.83$, $p < .001$. The two leftmost bars in Figure 2 break this down in a more specific way—showing the prediction rates as a function of those assigned to Hobie or Cody.

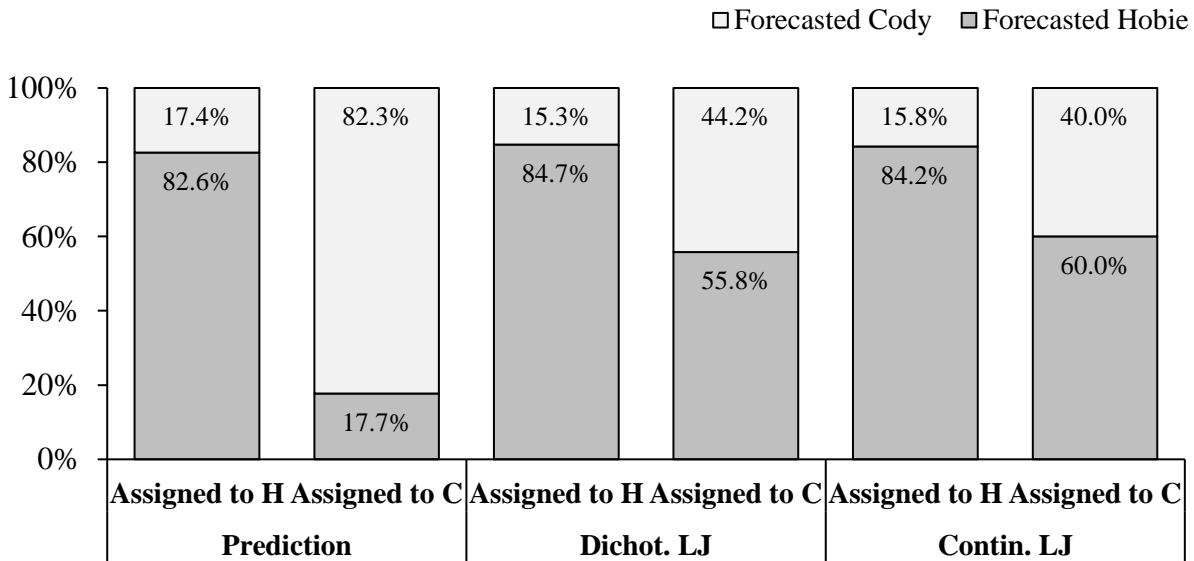


Figure 2. From Study 3, the proportions of responses in which participants forecasted either Cody (in light gray) or Hobie (in dark gray) winning the race, as a function of the assigned athlete (i.e., whether the participant's athlete was Cody or Hobie) and forecast format (i.e., whether participant made a prediction, dichotomous likelihood judgment, or continuous likelihood judgment). Note that each vertical bar represents all responses from participants assigned to a given cell; therefore, the two numbers per bar add to 100%. H = Hobie; C = Cody; Dichot. = Dichotomous; LJ = Likelihood Judgments; Contin. = Continuous.

The effect of desirability was also significant in both the dichotomous and continuous likelihood conditions, where 67.6% (binomial test, $p < .001$) and 64.7% (binomial test, $p = .004$) of participants forecasted their own athlete to win. However, as expected, the magnitude of the desirability effect was significantly larger in the prediction condition than in either of the likelihood judgment conditions—i.e., the 82.4% rate from the prediction condition was greater than the 67.6% rate, $X^2(1) = 6.13$, $p = .013$, and the 64.7% rate, $X^2(1) = 8.50$, $p = .004$.

This same conclusion—that the desirability effect was significantly larger in the prediction condition than in either of the likelihood conditions—was also verified with more detailed tests of Forecast Format x Assigned Athlete interactions in logistic regressions.¹¹ The

¹¹ The omnibus test of 2 x 3 interaction was also significant, $Wald \chi^2(2) = 7.98$, $p = .018$.

interaction test involving just the prediction and dichotomous likelihood conditions was significant, $\beta = -1.61$, $SE = .70$, $Wald \chi^2(1) = 5.33$, $OR = .20$, $p = .021$, as was the interaction test involving just the prediction and continuous likelihood conditions, $\beta = -1.82$, $SE = .70$, $Wald \chi^2(1) = 6.84$, $OR = .16$, $p = .009$. The interaction test involving just the two likelihood conditions was not significant, meaning that the desirability bias was not significantly different between those two conditions, $\beta = .21$, $SE = .67$, $Wald \chi^2(1) = .10$, $OR = 1.24$, $p = .751$.¹²

Most important, we tested our key prediction that the effect of the forecast format would be exclusively found among participants assigned to Cody the underdog, who was forecasted to win by only 36.5% of participants overall. Crucially, among people assigned to Cody, the percent of people who forecasted a Cody win varied substantially by the format condition, $X^2(2) = 24.42$, $p < .001$. The rates were 82.3% in the prediction condition, but only 44.2% and 40.0% in the dichotomous and continuous likelihood judgment conditions (see lighter gray area of the second, fourth, and sixth bars in Figure 3). When we restricted the analysis to look at only the participants assigned to Hobie, they gave similar forecasts—most indicating a Hobie win—regardless of the forecast format, $X^2(2) = .092$, $p = .955$ (see darker gray area of the first, third, and fifth bars in Figure 2).

Discussion

As expected, there was a significant desirability bias among people whose forecasts were solicited as predictions, and this bias was significantly reduced among people whose forecasts were solicited as either dichotomous or continuous likelihood judgments. This is an important generalization of the key results from Study 2, given that a different type of event was used in

¹² This study included an exploratory pair of numeric likelihood questions after the set of main dependent variables. Details about that measure and its results can be found in the Supplemental Materials, but here we simply note that the main effect of desirability on a composite from this pair of measures was significant ($p < .001$).

the study. It is also noteworthy that we avoided the possible wording confounds that were mentioned in the Discussion section for Study 2. All three forecasts were solicited using wording from an impersonal, third-person perspective. Equalizing the perspective wording across the formats was important given work by Juanchich et al. (2017) and Løhre & Teigen (2016).

Also as expected, the impact of the prediction vs. likelihood judgment formats was exclusively observed among participants assigned to the underdog (i.e., Cody). Underdog fans who were asked for predictions tended to express an optimistic response, even though underdog fans who were asked for likelihood judgments did not show an overall tendency to give responses that favored their athlete. Among people assigned an overdog, they tended to forecast that outcome regardless of whether they were asked for a likelihood judgment or asked for a prediction, because the evidence was already compatible with the preferred outcome.

Interestingly, although the results differed between the prediction condition and the likelihood judgment conditions, there were no significant differences between the dichotomous and continuous likelihood conditions. We had thought that likelihood judgments of either type would encourage objectivity and an attention to the weight of evidence, but we also thought the continuous format would minimize the possibility of some participants interpreting a likelihood judgment as essentially a prediction. Apparently, with regard to vulnerability to a motivated bias, soliciting a likelihood judgment on a dichotomous scale might not be appreciably different from soliciting the judgment on a continuous scale.

Study 4

In Study 4, we tested an alternative explanation for why likelihood responses in the previous studies were less vulnerable to motivated biases than were the predictions. The alternative explanation concerns the use of the word “more” when soliciting likelihood judgments. Using the term “more” in soliciting a likelihood judgment about a binary event seems

natural; in Study 3 the query for likelihood judgments asked, “Which competitor is more likely to win...?”. Meanwhile, the prediction question did not include the word more (“Which competitor will win...?”). A similar difference in queries existed in Study 2. Could the presence vs. absence of “more” be a crucial factor? Our reason for addressing this issue is because the word “more” in likelihood-judgment queries may implicitly emphasize the complementarity between the likelihoods of the possible outcomes. Granted, regardless of how people’s forecasts were solicited in our prior studies, we suspect it was very clear to people that the two possible outcomes in each sports event were mutually exclusive and exhaustive. Nonetheless, the complementarity of people’s responses regarding mutually exclusive and exhaustive outcomes can be affected by contextual features (Macchi et al., 1999; Smith et al., 2020; Windschitl, 2000). Therefore, we thought it was important to determine whether the presence/absence of the word “more” could have been the feature that played a role in the results of the previous studies.

To this end, for Study 4 we used the same paradigm but added likelihood-question conditions that excluded the word “more” (see Table 2). Specifically, two conditions used the same prediction and dichotomous-likelihood questions that were used in Study 3, but two conditions involved likelihood questions that did not contain the word “more” in the stem or response options.¹³ The response options for one of these new conditions were dichotomous, and for the other, they were continuous.¹⁴

Table 2

Prompt and the Response Options Used in Study 4 for Each Forecast Format Condition

¹³ We note that the labeling of the four forecast types was slightly modified from what was used in our preregistration for brevity and coherence.

¹⁴ The reason we included a new continuous likelihood question along with the new dichotomous one—despite the previous null effects for this continuous-vs.-dichotomous distinction—was because the previous null effects might have been driven by the fact that the word “more” was present in both versions.

Forecast Formats	Prompt	Response Options
Dichot. Prediction	Which competitor will win the race?	Hobie Call will win, Cody Moat will win
Dichot. LJ with “more”	Which competitor is more likely to win the race?	Hobie Call is more likely to win, Cody Moat is more likely to win
Dichot. LJ without “more”	Which competitor is likely to win the race?	Hobie Call is likely to win, Cody Moat is likely to win
Contin. LJ without “more”	Which competitor is likely to win the race?	Hobie Call is very likely to win (1) – Cody Moat is very likely to win (6)

Note. Dichot. = Dichotomous; LJ = Likelihood Judgments; Contin. = Continuous.

Our main prediction was that the effect of desire would be more evident in the prediction condition than in any of the three likelihood conditions. This would support the view that the prediction vs. likelihood judgment distinction (and not the presence of the wording “more”) is a crucial determinant of people’s vulnerability to motivated biases. However, if the presence of the word “more” had been exerting a crucial influence on how people implicitly interpret the solicitations of forecasts, then we would find that the new likelihood questions that excluded the word “more” would show about the same impact of desire as found in the prediction condition.

Additionally, for efficiency, we assigned all participants to Cody instead of randomly assigning them to either Cody or Hobie. Given our previous studies showed that forecast format played a crucial role exclusively among those who were assigned to the team/athlete with an underdog status, it stands to reason that this was where we should test again for the predicted effects.

Methods

Participants and Design

A total of 480 participants (334 Amazon Mechanical Turk workers and 146 undergraduates at the University of Iowa) participated in a 4-cell between-subject design, where

each cell was defined by the type of forecast wording that was used (see Table 2 for the four formats). There was also a counterbalancing of where names for the two athletes appeared on response scales (on left or right). This counterbalancing had no significant effect on the main findings and will not be considered further. After dropping some participants (described below), the remaining sample size of 354 provided 86% power to detect a proportion difference of 20% between two cells with a one-tailed test. The design, hypotheses, and analysis of the study were preregistered on OSF (https://osf.io/hk4dm/?view_only=d02e8263151c4fa1bbcc24e8ecca4404).

Material, Procedure and Dependent Variables

The materials and procedures were identical to Study 3, with the following two exceptions. First, all participants were assigned to Cody (the underdog). Participants underwent the same protocol where they made what appeared as a blind choice between “Athlete A” and “Athlete B,” but the results were that they were always assigned to Cody. Second, as mentioned, there were four forecast formats with wording as shown in Table 2.

Results and Discussion

Preliminary Analysis

The preregistered exclusion criteria were slightly more stringent than for Study 3. We excluded those who failed to recall which athlete they were assigned to ($n = 81$) or answered fewer than 2 of 5 memory items about race details ($n = 64$). In total, data from 26.3% of the participants (126 of 480) were dropped, with the rate of exclusion being similar across the 4 forecast types. The results reported below are for the final sample of 354 participants.

Main Analyses of Forecasts

As in Study 3, we dichotomized the continuous likelihood judgments using the midpoint of the scale to allow direct comparisons across conditions. Because all participants were assigned to Cody, we could not use the same analysis for testing the desirability bias as we did in Study 3,

which used 50% as a threshold of indicating presence of the bias. Instead, we assumed that any differences across conditions in the proportions of participants whose forecast favored Cody would reflect the differential impact of the desirability bias.

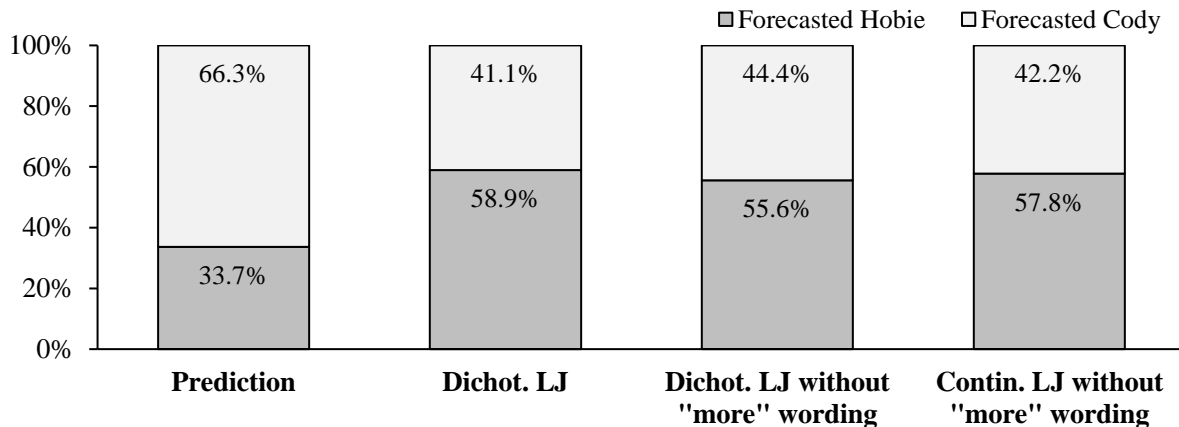


Figure 3. From Study 4, the proportions of participants who forecasted either Cody (in light gray) or Hobie (in dark gray) winning the race, as a function of the forecast format condition. Unlike in Study 3, all participants were assigned to Cody. Dichot. = Dichotomous; LJ = Likelihood Judgments; Contin. = Continuous.

As expected, the proportion of participants forecasting a Cody win was significantly impacted by forecast format, $X^2(3) = 14.91$, $p = .002$ (see Figure 3; the lighter gray areas indicating the proportion of people forecasting a Cody win). More targeted analyses further revealed that the proportion of participants predicting Cody was significantly larger in the prediction condition than in each of the three other conditions—dichotomous ($p < .001$), dichotomous without “more” ($p = .004$), continuous without “more” ($p = .002$) conditions. The proportions from the latter three likelihood-judgment conditions did not statistically differ, $p = .894$.

These results show that the removal of the word “more” in dichotomous and continuous likelihood judgment prompts did not lead to different results from what we observed using the original way of soliciting the likelihood judgment. This rules out the possibility that the pattern of findings from the previous studies were driven by a specific characteristic of the way likelihood judgments were measured (i.e., using the word “more” to solicit answers about which

outcome was more likely). In line with previous studies, Study 4 again demonstrated that people are more impacted by the desirability bias when they make predictions about an outcome as opposed to when they make likelihood judgments.

General Discussion

Whether and when expectations are biased by preferences or utilities are of substantial concern (de Molière & Harris, 2016; Harris & Hahn, 2011; Hastie, 2001; Logg et al., 2018). Past work had provided limited evidence that the impact of people's preferences on their expectations might be moderated by how those expectations are solicited (Windschitl et al., 2010). The present work tested this possibility more substantially, and the results support an updated account of when and why this moderation holds. This updated account attributes the moderation effect to slightly different connotations attached to questions about predictions vs. likelihood judgments—not to differences between dichotomous vs. continuous response formats.

Study 1 provided evidence that people view predictions as more accommodating of hunches than are likelihood judgments. This is an important finding given that past work suggests that encouraging people to rely on hunches can exacerbate a desirability bias (Rose & Aspiras, 2020; Windschitl et al., 2010). In Studies 2-4, participants' outcome preferences were manipulated, and forecasts were solicited as predictions or likelihood judgments. Predictions showed larger degrees of desirability bias than did likelihood judgments. This was true when expectations were about the outcome of a close basketball game (Study 2) or the outcome of an endurance race (Studies 3-4). Critically, these effects emerged even when both predictions and likelihood judgments were solicited using a dichotomous response format.

The difference in how people made predictions and likelihood judgments was primarily observed among people who had been assigned to the “underdog”—the team/athlete who appeared relatively disadvantaged at the time the forecast was solicited. This fits with our

account. In this situation, the assessment of evidence prompted by a likelihood-judgment query leads to a forecast that essentially acknowledges the disadvantaged position of the preferred team/athlete, whereas the extra flexibility afforded by a prediction allows for a forecast that fits with an extra-evidentiary consideration, even though it contradicts the disadvantaged position of the preferred team/athlete. Among participants assigned to the “overdog,” both the evidence and extra-evidentiary considerations (e.g., outcome desirability) point in the same direction and lead to similar likelihood judgments and predictions.

Our results substantially generalize prior findings of differences between likelihood judgments and predictions. The current research involved forecasts about single, naturalistic events. Past work investigating the impact of format (likelihood judgments vs. predictions) on desirability bias have relied on the marked-card paradigm or its close variants, in which the probabilities about the event outcomes were supplied to the respondents across multiple trials (Price & Marquez, 2005; Windschitl et al., 2010). This kind of set-up may be useful in executing more controlled experiments but does not fully reflect real-life contexts where probabilities regarding the event outcomes are not readily accessible. By demonstrating a robust impact of forecast-solicitation types on desirability bias in contexts where explicit numeric information is absent, our results provide additional evidence of validity to past findings and generalize them closer to real-world environments.

The current findings have implications for the biased-guessing account, an account previously proposed to explain why, in the marked-card paradigm, predictions were more impacted than likelihood judgments by outcome desirability (Windschitl et al., 2010). This account suggested that the desirability bias is most likely to arise on occasions where the situation allows individuals to make arbitrary guesses for predictions (such as for a stochastic event where the chances of two outcomes are equal or almost equal), whereas the continuous

nature of a likelihood response scale does not allow for guessing. Because the present findings demonstrated that predictions and likelihood judgments are differentially impacted by motivated concerns even when both forecasts were solicited on dichotomous scales, the biased-guessing account cannot explain the findings without modification. Interestingly, in our studies that used both continuous and dichotomous scales (Studies 3-4), this scaling feature did not make a significant difference in results.

At What Stage Does Bias Enter?

As noted by Krizan and Windschitl (2007), motivated factors may lead to a biased expectation by altering evidence search, evidence evaluation, and/or response formation. For the present experiments, the bias that differentially impacted predictions vs. likelihood judgment was presumably most influential in the last stage (response formation). Recall that in Studies 2-4, participants witnessed the event-relevant information before being exposed to the specific forecast question. That is, participants did not know what the format of the queried forecast was until they already had all the available information about the event. Given this setup, we can assume that participants' initial evidence searches and evaluations were conducted in the same manner whether they were in a condition asking for predictions or likelihood judgments. This implicates the last stage of the processing as crucial for the differential influence of bias.

With that said, it is possible that upon seeing the forecast query, participants tried to recall and evaluate specific pieces of evidence that they had already witnessed. There are empirical findings suggesting that presence of motivations to arrive at certain conclusions can alter the way in which the evidence is perceived, searched, recalled, and evaluated (Balcetis & Dunning, 2006; Kunda, 1990; Lench et al., 2014; Scherer et al., 2013; Voss et al., 2008; Windschitl et al., 2013). Therefore, to the extent that people re-engaged these processes after being asked to report a forecast, there is the potential that those processes were differentially

biased, depending on the format of the forecast. For example, perhaps the flexibility afforded by a prediction format (vs. the assessment-like emphasis of a likelihood format) made people more interested in justifying a preferred response; they may then have engaged in a confirmatory evidence search that supported (rightly or wrongly) a prediction that matched their preference (Dawson et al., 2002; Gilovich, 1991). An interesting question is whether the impact of format would be even stronger than it was in the present studies, if participants were shown the format for the forecast before they viewed most of the information about the event.

Interestingly, an additional study suggests that participants accept that motivated bias can be differentially expressed in a response stage—as a function of response type. In that study (see Study C in the Supplemental Materials), participants responded to a hypothetical scenario in which they were asked to imagine that their predictions conflicted with their likelihood judgments. When there was a potential motivated reason for this conflict, they found the conflict between the two responses to be more acceptable. Specifically, they found the conflict less odd when the prediction favored a team they liked rather than one they disliked or felt neutral about.

A Curious Inconsistency

A curious inconsistency in one element of results from Study 2 and 3 was that there was no sign of a desirability bias as measured from likelihood judgments in the basketball paradigm (Study 2), but there was a reliable desirability bias on likelihood judgments in the Spartan Race paradigm (Studies 3). This fact does not threaten the validity of any of our main conclusions, but it does raise the question of why this happened. There are too many differences between the paradigms to pinpoint an answer now, but we will note one difference between the two paradigms that could be investigated further. In the Spartan Race paradigm, participants were generally forewarned that they would be making a prediction about the race they would witness. This prompt was kept consistent across those who were later asked to make predictions or

likelihood judgments and therefore cannot account for the main results in those studies. But nonetheless, the forewarning might have set in motion a biased evaluation of the evidence as it was initially encountered, resulting in a bias in likelihood judgments. There was no such forewarning in the basketball paradigm.

What is a Desirability Bias/Wishful Thinking?

More generally, the present findings have important implications for the debates about whether, how, and how strongly motivations bias expectations (Bar–Hillel et al., 2008a; Bilgin, 2012; Biner et al., 2009; de Molière & Harris, 2016; Harris, 2017; Harris et al., 2009; Harris & Corner, 2011; Lench, 2009; Logg et al., 2018; Massey et al., 2011; Simmons & Massey, 2012; Slovic, 1966). Terms like expectations and forecasts are broad ones. They can be measured by soliciting either predictions or likelihood judgments, and this choice by a researcher could substantively change the results of their study(ies) and the conclusions that are drawn. In defining the terms *desirability bias* or *wishful thinking* in their papers, researchers often use the terms in a general or loose way. This can be justified if the definition is meant to be broad and inclusive, but given the present results, this inclusivity probably needs to be qualified.

Some researchers might argue that convincing evidence for wishful thinking must show that desires impact likelihood judgments (or similar), because predictions are fickle and may not directly reflect internal subjective probabilities. This certainly has a compelling ring to it, but we also note that if behavior follows more closely from people's predictions than from their likelihood judgments (see Footnote 5, see also Stuart et al., 2017), it seems overly limiting to downplay the importance of people's predictions.

Ties to Related Dichotomies

Our findings fit within a broader category of research examining how variations in the way that uncertainty is measured or expressed can have non-obvious implications (Juanchich et

al., 2017). Here, we discuss three subsets of research relevant to our distinction between the predictions versus likelihood judgments.

First, our findings bear some resemblance to work on the Wells effect, which examines how verdicts relate to subjective probability judgments (Arkes et al., 2012; Niedermeier et al., 1999; Wells, 1992). Logic and sometimes legal directives indicate that verdicts about guilt should follow from subjective probability, but studies reveal divergences. A key divergence involves whether statistical evidence in a case is presented in the form of a base-rate or a reliability rate. This distinction in form has been shown to produce empirical differences on people's verdicts even when it does not affect relevant subjective probabilities (Arkes et al., 2012). Like with our findings, this version of the Wells effect is a phenomenon in which people's likelihood judgments seem to resist influence of a normatively-irrelevant factor even while that same factor did influence a related response.

A second line of work that deserves mention is on reversals of preferences measured via choice vs. judgment ratings (Lichtenstein & Slovic, 1971; Slovic & Lichtenstein, 1983). These reversals have often been attributed to differences in the cognitive processes triggered by the two measures (Billings & Scherer, 1988; Schkade & Johnson, 1989; Wedell & Senter, 1997). Even more germane for the present work might be a proposal by Sood and Forehand (2005) about the role of motivation, suggesting that the discrepancy between two measures occurs because choice elicits more self-referent processing than do judgments. They suggest that choice implies a personal consumption, prompting responders to consider connections between themselves and the option, whereas judgments would remain more as an abstract evaluation. Our proposal about predictions could be put in similar terms, by saying when people make predictions rather than likelihood judgments, they pay slightly more attention to the self-relevancy of the prediction they

are about to make (e.g., a sense of allegiance to their desired athlete) and to the self-relevancy of the consequences of the possible outcomes.

The third line of relevant work concerns how different natural-language choices for communicating uncertainty relate to the types of uncertainty being described (Fox & Ülkümen, 2011; Juanchich et al., 2017; Løhre & Teigen, 2016). In natural language, people can choose from at least two ways of expressing their level of uncertainty—confidence statements (e.g., “I am fairly confident,” “I am 90% sure”) and likelihood statements (e.g., “I believe it is fairly likely,” “I’d say there is a 90% chance”). Ülkümen, Fox, and Malle (2016) found that confidence statements tend to be used by speakers who are primarily communicating about epistemic uncertainty, whereas likelihood statements tend to be used by speakers who are primarily communicating about aleatory uncertainty (see also Fox & Ülkümen, 2011; Løhre & Teigen, 2016). Notice that when we solicited people’s uncertainty in the present studies, we used likelihood phrasing. Yet, we could have asked questions like “Are you more confident that X or Y will win?” See Study B in the Supplemental Materials for a description of a study that found similar levels of bias impacting forecasts solicited as likelihood judgments or confidence judgments. More studies would be needed to know if forecast solicitations phrased in confidence language reliably differ from predictions in their vulnerability to motivated bias, but this initial study hints that they might.

Application and Conclusion

The current findings have practical implications for soliciting forecasts, whether it is in a research context, polling context, or just a conversation in which one person wants to know what a friend or advisor thinks is going to happen. The lesson from the findings is that the impact of motivated biases can be reduced by asking for a likelihood judgment rather than a prediction. This lesson is not the only reason why it might be wise to ask for likelihood judgments, because

likelihood judgments—as usually measured on a continuous scale—provide a finer degree of information. Recent research suggests that people indeed prefer advisors who add information about their degree of uncertainty rather than just provide dichotomous predictions (Gaertig & Simmons, 2018). With that said, consideration needs to be given as to what is potentially lost in asking for likelihood judgments. If people’s behavioral inclinations about an uncertain event (e.g., whether they prepare or not for an undesired outcome) are biased like their predictions are, then likelihood judgments might be less biased but not as helpful for predicting behavior (Stuart et al., 2017). In short, this and related work suggest that, when maximally informative forecasts are sought, the choice between soliciting likelihood judgments versus predictions should be carefully considered.

References

- Anderson, C., Brion, S., Moore, D. A., & Kennedy, J. A. (2012). A status-enhancement account of overconfidence. *Journal of Personality and Social Psychology*, 103(4), 718–735.
- Arkes, H. R., Shoots-Reinhard, B., & Mayes, R. S. (2012). Disjunction between probability and verdict in juror decision making. *Journal of Behavioral Decision Making*, 25(3), 276–294.
- Armor, D. A., Massey, C., & Sackett, A. M. (2008). Prescribed optimism: Is it right to be wrong about the future? *Psychological Science*, 19(4), 329–331.
- Babad, E. (1987). Wishful thinking and objectivity among sports fans. *Social Behaviour*.
- Balcetis, E., & Dunning, D. (2006). See what you want to see: motivational influences on visual perception. *Journal of Personality and Social Psychology*, 91(4), 612–625.
- Bar-Hillel, M., & Budescu, D. (1995). The Elusive Wishful Thinking Effect. *Thinking & Reasoning*, 1(1), 71–103. <https://doi.org/10.1080/13546789508256906>
- Bar-Hillel, M., Budescu, D. V., & Amar, M. (2008a). Predicting World Cup results: Do goals seem more likely when they pay off? *Psychonomic Bulletin and Review*, 15(2), 278–283. <https://doi.org/10.3758/PBR.15.2.278>
- Bar-Hillel, M., Budescu, D. V., & Amar, M. (2008b). Wishful thinking in predicting world cup results: Still elusive. *Rationality and Social Responsibility: Essays in Honor of Robyn Mason Dawes*, 175–186. <https://doi.org/10.4324/9780203889695>
- Bilgin, B. (2012). Losses loom more likely than gains: Propensity to imagine losses increases their subjective probability. *Organizational Behavior and Human Decision Processes*, 118(2), 203–215.
- Billings, R. S., & Scherer, L. L. (1988). The effects of response mode and importance on decision-making strategies: Judgment versus choice. *Organizational Behavior and Human Decision Processes*, 41(1), 1–19.

- Biner, P. M., Johnston, B. C., Summers, A. D., & Chudzynski, E. N. (2009). Illusory control as a function of the motivation to avoid randomly determined aversive outcomes. *Motivation and Emotion*, 33(1), 32–41.
- Bonnefon, J.-F., & Villejoubert, G. (2006). Tactful or doubtful? Expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science*, 17(9), 747–751.
- Brenner, L. A., Koehler, D. J., Liberman, V., & Tversky, A. (1996). Overconfidence in probability and frequency judgments: A critical examination. *Organizational Behavior and Human Decision Processes*, 65(3), 212–219.
- Dawson, E., Gilovich, T., & Regan, D. T. (2002). Motivated Reasoning and Performance on the was on Selection Task. *Personality and Social Psychology Bulletin*, 28(10), 1379–1387.
- de Molière, L., & Harris, A. J. L. (2016). Conceptual and direct replications fail to support the stake-likelihood hypothesis as an explanation for the interdependence of utility and likelihood judgments. *Journal of Experimental Psychology: General*, 145(4), e13–e26.
- Dunning, D., Griffin, D. W., Milojkovic, J. D., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology*, 58(4), 568–581.
- Fox, C. R., & Ülkümen, G. (2011). Distinguishing two dimensions of uncertainty. Brun W, Keren G, Kirkebøen G, Montgomery H, eds. In *Perspectives on Thinking, Judging, and Decision Making: A Tribute to Karl Halvor Teigen*. Universitetsforlaget.
- Gaertig, C., & Simmons, J. P. (2018). Do people inherently dislike uncertain advice? *Psychological Science*, 29(4), 504–520. <https://doi.org/10.1177/0956797617739369>
- Gaissmaier, W., & Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition*, 109(3), 416–422. <https://doi.org/10.1016/j.cognition.2008.09.007>
- Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*.

Free Press.

Goldschmied, N. P., & Vandello, J. A. (2012). The Future is Bright: The Underdog Label, Availability, and Optimism. *Basic and Applied Social Psychology*, 34(1), 34–43.

<https://doi.org/10.1080/01973533.2011.637726>

Harris, A. J. L. (2017). Understanding the coherence of the severity effect and optimism phenomena: Lessons from attention. *Consciousness and Cognition*, 50, 30–44.

<https://doi.org/10.1016/j.concog.2016.10.014>

Harris, A. J. L., & Corner, A. (2011). Communicating environmental risks: Clarifying the severity effect in interpretations of verbal probability expressions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(6), 1571–1578.

Harris, A. J. L., Corner, A., & Hahn, U. (2009). Estimating the probability of negative events. *Cognition*, 110(1), 51–64.

Harris, A. J. L., & Hahn, U. (2011). Unrealistic optimism about future life events: A cautionary note. *Psychological Review*, 118(1), 135–154.

Haselton, M. G., & Nettle, D. (2005). The paranoid optimist: an integrative evolutionary model of cognitive biases. *Manuscript under Review*.

Hastie, R. (2001). Problems for judgment and decision making. *Annual Review of Psychology*, 52(1), 653–683.

Hayes Jr, S. P. (1936). The predictive ability of voters. *The Journal of Social Psychology*, 7(2), 183–191.

Helweg-Larsen, M., Sadeghian, P., & Webb, M. S. (2002). The stigma of being pessimistically biased. *Journal of Social and Clinical Psychology*, 21(1), 92–107.

Hertwig, R., & Fox, C. R. (2006). “Decisions from experience” = sampling error + prospect theory. *Judgment and Decision Making*, 1(2), 159–161.

Howell, W. C., & Burnett, S. A. (1978). Uncertainty measurement: A cognitive taxonomy.

Organizational Behavior and Human Performance, 22(1), 45–68.

[https://doi.org/10.1016/0030-5073\(78\)90004-1](https://doi.org/10.1016/0030-5073(78)90004-1)

Irwin, F. W. (1953). Stated expectations as functions of probability and desirability of outcomes.

Journal of Personality, 21, 329–335.

Juanchich, M., Gourdon-Kanhukamwe, A., & Sirota, M. (2017). I am uncertain” vs “it is

uncertain”. How linguistic markers of the uncertainty source affect uncertainty

communication. *Judgment and Decision Making*, 12(5), 445–465.

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence

research: A critical examination of the hard–easy effect. *Psychological Review*, 107(2),

384–396.

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, 11(2), 143–157.

[https://doi.org/10.1016/0010-0277\(82\)90023-3](https://doi.org/10.1016/0010-0277(82)90023-3)

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, 73(1),

1–2. <https://doi.org/http://dx.doi.org/10.1037/amp0000263>

Koehler, D. J., & James, G. (2009). Probability matching in choice under uncertainty: Intuition

versus deliberation. *Cognition*, 113(1), 123–127.

<https://doi.org/10.1016/j.cognition.2009.07.003>

Koehler, D. J., & James, G. (2010). Probability matching and strategy availability. *Memory and*

Cognition, 38(6), 667–676. <https://doi.org/10.3758/MC.38.6.667>

Krizan, Z., Miller, J. C., & Johar, O. (2010). Wishful thinking in the 2008 US presidential

election. *Psychological Science*, 21(1), 140–146.

Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism.

Psychological Bulletin, 133(1), 95–121. <https://doi.org/10.1037/0033-2909.133.1.95>

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

Lench, H. C. (2009). Automatic optimism: The affective basis of judgments about the likelihood of future events. *Journal of Experimental Psychology: General*, 138(2), 187–200.

<https://doi.org/10.1037/a0015380>

Lench, H. C., Smallman, R., Darbor, K. E., & Bench, S. W. (2014). Motivated perception of probabilistic information. *Cognition*, 133(2), 429–442.

<https://doi.org/10.1016/j.cognition.2014.08.001>

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1977). Calibration of probabilities: The state of the art. *Decision Making and Change in Human Affairs*, 275–324.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 89(1), 46–55.

<https://doi.org/10.1037/h0031207>

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 49(2), 433–442.

Logg, J. M., Haran, U., & Moore, D. A. (2018). Is overconfidence a motivated bias? Experimental evidence. *Journal of Experimental Psychology: General*, 147(10), 1445–

1465. <https://doi.org/10.1037/xge0000500>

Løhre, E., & Teigen, K. H. (2016). There is a 60% probability, but I am 70% certain:

Communicative consequences of external and internal expressions of uncertainty. *Thinking & Reasoning*, 22(4), 369–396.

Macchi, L., Osherson, D., & Krantz, D. H. (1999). A note on superadditive probability judgment. *Psychological Review*, 106(1), 210–214.

Markman, K. D., & Hirt, E. R. (2002). Social prediction and the “allegiance bias.” *Social*

Cognition, 20(1), 58–86.

Marks, R. W. (1951). The effect of probability, desirability, and “privilege” on the stated expectations of children. *Journal of Personality*, 19(3), 332–351.

Massey, C., Simmons, J. P., & Armor, D. A. (2011). Hope over experience: Desirability and the persistence of optimism. *Psychological Science*, 22(2), 274–281.

<https://doi.org/10.1177/0956797610396223>

Milkov, A. V. (2020). Forecasting Abilities of Individual Petroleum Explorers: Preliminary Findings From Crowdsourced Prospect Assessments. *Journal of Petroleum Geology*, 43(4), 383–400. <https://doi.org/10.1111/jpg.12771>

Miller, J. E., Park, I., Smith, A. R., & Windschitl, P. D. (2021). Do People Prescribe Optimism, Overoptimism, or Neither? *Psychological Science*, 32(10), 1605–1616.

Morewedge, C. K., Tang, S., & Larrick, R. P. (2018). Betting your favorite to win: Costly reluctance to hedge desired outcomes. *Management Science*, 64(3), 997–1014.

<https://doi.org/10.1287/mnsc.2016.2656>

Niedermeier, K. E., Kerr, N. L., & Messé, L. A. (1999). Jurors’ use of naked statistical evidence: Exploring bases and implications of the Wells effect. *Journal of Personality and Social Psychology*, 76(4), 533–542.

Price, P. C., & Marquez, C. (2005). Wishful thinking in the predictions of a simple repeatable event: Effects on deterministic versus probabilistic predictions. *Unpublished Manuscript*. California State University, Fresno.

Pruitt, D. G., & Hoge, R. D. (1965). Strength of the relationship between the value of an event and its subjective probability as a function of method of measurement. *Journal of Experimental Psychology*, 69(5), 483–489.

Risen, J. L., & Gilovich, T. (2008). Why people are reluctant to tempt fate. *Journal of*

Personality and Social Psychology, 95(2), 293–307.

Rose, J. P., & Aspiras, O. (2020). “To hope was to expect”: The impact of perspective taking and forecast type on wishful thinking. *Journal of Behavioral Decision Making*, 33(4), 411–426.

<https://doi.org/10.1002/bdm.2170>

Scherer, A. M., Windschitl, P. D., & Smith, A. R. (2013). Hope to be right: Biased information seeking following arbitrary and informed predictions. *Journal of Experimental Social Psychology*, 49(1), 106–112. <https://doi.org/10.1016/j.jesp.2012.07.012>

Schkade, D. A., & Johnson, E. J. (1989). Cognitive processes in preference reversals.

Organizational Behavior and Human Decision Processes, 44(2), 203–231.

Sieck, W. R., & Yates, J. F. (2001). Overconfidence effects in category learning: A comparison of connectionist and exemplar memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(4), 1003–1021.

Learning, Memory, and Cognition, 27(4), 1003–1021.

Simmons, J. P., & Massey, C. (2012). Is optimism real? *Journal of Experimental Psychology: General*, 141(4), 630–634. <https://doi.org/10.1037/a0027405>

Slovic, P. (1966). Value as a determiner of subjective probability. *IEEE Transactions on Human Factors in Electronics*, 1, 22–28.

Slovic, P., & Lichtenstein, S. (1983). Preference reversals: A broader perspective. *The American Economic Review*, 73(4), 596–605.

Smith, A. R., Windschitl, P. D., & Rose, J. P. (2020). An integrated approach to biases in referent-specific judgments. *Thinking & Reasoning*, 26(4), 581–614.

<https://doi.org/10.1080/13546783.2019.1691053>

Stuart, J. O. R., Windschitl, P. D., Smith, A. R., & Scherer, A. M. (2017). Behaving Optimistically: How the (Un)Desirability of an Outcome Can Bias People’s Preparations for It. *Journal of Behavioral Decision Making*, 30(1), 54–69. <https://doi.org/10.1002/bdm.1918>

- Sweeny, K., Carroll, P. J., & Shepperd, J. A. (2006). Is optimism always best? Future outlooks and preparedness. *Current Directions in Psychological Science*, 15(6), 302–306.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin*, 103(2), 193–210.
- Tyler, J. M., & Rosier, J. G. (2009). Examining self-presentation as a motivational explanation for comparative optimism. *Journal of Personality and Social Psychology*, 97(4), 716–727.
- Ülkümen, G., Fox, C. R., & Malle, B. F. (2016). Two dimensions of subjective uncertainty: Clues from natural language. *Journal of Experimental Psychology: General*, 145(10), 1280–1297.
- Vosgerau, J. (2010). How prevalent is wishful thinking? Misattribution of arousal causes optimism and pessimism in subjective probabilities. *Journal of Experimental Psychology: General*, 139(1), 32–48. <https://doi.org/10.1037/a0018144>
- Voss, A., Rothermund, K., & Brandtstädter, J. (2008). Interpreting ambiguous stimuli: Separating perceptual and judgmental biases. *Journal of Experimental Social Psychology*, 44(4), 1048–1056. <https://doi.org/10.1016/j.jesp.2007.10.009>
- Wedell, D. H., & Senter, S. M. (1997). Looking and weighting in judgment and choice. *Organizational Behavior and Human Decision Processes*, 70(1), 41–64.
- Wells, G. L. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology*, 62(5), 739–752.
- Windschitl, P. D. (2000). The binary additivity of subjective probability does not indicate the binary complementarity of perceived certainty. *Organizational Behavior and Human Decision Processes*, 81(2), 195–225.
- Windschitl, P. D., & Chambers, J. R. (2004). The dud-alternative effect in likelihood judgment. *Journal of Experimental Psychology: Learning Memory and Cognition*, 30(1), 198–215.

<https://doi.org/10.1037/0278-7393.30.1.198>

Windschitl, P. D., Scherer, A. M., Smith, A. R., & Rose, J. P. (2013). Why so confident? The influence of outcome desirability on selective exposure and likelihood judgment.

Organizational Behavior and Human Decision Processes, 120(1), 73–86.

<https://doi.org/10.1016/j.obhdp.2012.10.002>

Windschitl, P. D., Smith, A. R., Rose, J. P., & Krizan, Z. (2010). The desirability bias in predictions: Going optimistic without leaving realism. *Organizational Behavior and Human Decision Processes*, 111(1), 33–47. <https://doi.org/10.1016/j.obhdp.2009.08.003>

Decision Processes, 111(1), 33–47. <https://doi.org/10.1016/j.obhdp.2009.08.003>

Windschitl, P. D., & Stuart, J. O. (2015). Optimism biases: Types and causes. *The Wiley Blackwell Handbook of Judgment and Decision Making*, 2, 431–455.

Blackwell Handbook of Judgment and Decision Making, 2, 431–455.

Windschitl, P. D., & Wells, G. L. (1998). The Alternative-Outcomes Effect. *Journal of Personality and Social Psychology*, 75(6), 1411–1423.