More likely but less probable: The construction of relative versus absolute belief[1]

Craig R. Fox
University of California at Los Angeles

Jonathan Levav
Columbia University

Running Head: Absolute versus Relative Likelihood

Address correspondence to:
Craig R. Fox
UCLA Anderson School
110 Westwood Plaza #D511
Los Angeles, CA 90095-1481

voice: 310/206-3403
fax: 310/794-4257
e-mail: craig.fox@anderson.ucla.edu

[1] Early work on this project was completed when both authors were at Duke University, whose support we gratefully acknowledge. We also thank Manel Baucells, Florian Ederer, Yuval Rottenstreich, and Jack Soll for helpful comments on earlier drafts of this paper.

Abstract

In this paper we document two systematic anomalies in relative and absolute likelihood judgment. First, we provide examples in which event *A* is judged "more likely" than event *B* but event *B* is assigned a higher probability than event *A*, violating procedure invariance. Second, we provide examples in which event *A* is judged more likely to occur than event *B* while event not-*A* is also judged more likely than event not-*B*, violating ordinal complementarity. We attribute these "belief reversals" to the *complement neglect hypothesis:* evidence for complementary events is underweighted in relative but not absolute likelihood judgment. We advance the "Contingent Weighting of Support" (CWS) model in which support theory is embedded within the contingent weighting model, and derive conditions under which each kind of belief reversal is expected to occur. We then fit this model to data in two experiments, and provide direct evidence of complement neglect in relative but not absolute likelihood judgment.

The Construction of Absolute and Relative Belief

1. Introduction.

People are frequently called on to evaluate the relative likelihood of events. For instance, an investor might assess which of several mutual funds has the best chance of outperforming the market; a doctor might be asked which of three treatment options is most likely to cure a disease; a prospective buyer might wonder which of two car models is less likely to break down. Whether we infer a person's belief from his or her choices, explicit probability estimates, or verbal statements that qualify belief (e.g., "It is fairly likely that…"), we rely on several implicit assumptions to derive consensual meaning from these data. Among the most basic assumptions is *procedure invariance*: the belief ordering of two events, *A* and *B*, should coincide under normatively equivalent elicitation modes. Thus, absolute and relative expressions of belief should be consistent with one another. For instance, when a weather forecaster says she thinks there is a 70% chance of rain tomorrow in Seattle and a 40% chance of rain tomorrow in Los Angeles, we assume that she would also say that she thinks it is "more likely" to rain tomorrow in Seattle than Los Angeles. Second, the ordering of beliefs of any pair of events should be the reverse of the ordering of their complements, an assumption that we label *ordinal complementarity*. Thus, we expect the forecaster to say she thinks it is more likely to rain in Atlanta than Boston if and only if she would also say that it is more likely to remain dry in Boston than Atlanta.

The assumptions of procedure invariance and ordinal complementarity seem to be unassailable, if not tautological. In fact, axiomatic treatments of subjective probability do not typically distinguish between absolute and relative likelihood judgment, and they take ordinal complementarity for granted (Fishburn, 1986; Krantz, Luce, Suppes & Tversky, 1971).[2] However, as we will demonstrate, both assumptions can fail in systematic and predictable ways. These violations provide important clues about how people weight evidence when making absolute and relative likelihood judgments. In this paper we document several examples of violations of procedure invariance and ordinal complementarity, develop a model of belief construction that accommodates them, fit this model and measure its parameters. To do so we embed support theory (Tversky& Koehler, 1994; Rottenstreich & Tversky, 1997) within the contingent weighting model (Tversky, Sattath & Slovic & 1988) to develop a new model that characterizes the construction of absolute and relative likelihood judgment.

The psychology of likelihood judgment.

Research on heuristics and biases (Tversky & Kahneman, 1974; Kahneman, Slovic & Tversky, 1982; Gilovich, Griffin & Kahneman, 2002; Shah & Oppenheimer, 2008) asserts that people judge the *relative* likelihood of events by evaluating a substitute attribute such as the relative ease with which instances of each event come to mind or the relative similarity of each event to a relevant prototype (Frederick & Kahneman, 2002; Kahneman, 2003). For instance, a person may judge the likelihood of rain to be greater in Seattle than in Los Angeles because recent instances of rain are easier to recall in Seattle than Los Angeles.

Although studies of judgmental heuristics have been useful in predicting and explaining relative likelihood judgment, they do not explain how people quantify or qualify their *absolute*

---

[2]For instance, in Fishburn's (1986) review of the axioms of subjective probability he sets up a system of axioms defined using the $>$ operator so that $A > B$ is read as "event A is (regarded by the individual as) more probable than event B," but he does not explicitly distinguish absolute from relative probability judgment. Fishburn does explicitly advance ordinal complementarity as an axiom that he says "might quality [as] …so obvious and uncontroversial as to occasion no serious criticism" (p. 336). Likewise, Krantz, Luce, Suppes & Tversky (1971) define the "$\geq$" operator as "qualitatively at least as probable as" and defer "debates about the meaning of probability" as being "in reality, about acceptable empirical methods to determine $\geq$" (p. 200). They later articulate ordinal complementarity as Lemma 4 that follows from other axioms that are necessary for a representation of qualitative probability (see pp. 203, 212).

degree of belief, which is required for many choices. For instance, the decision whether or not to carry an umbrella requires one to assess how likely it is to rain today; the decision whether or not to purchase an insurance policy requires one to estimate how likely one is to make a claim; the decision whether or not to accept a legal settlement offer requires one to estimate how likely one is to prevail in court.

To model absolute likelihood judgment, support theory (Tversky & Koehler, 1994; Rottenstreich & Tversky, 1997) distinguishes between events and descriptions of events, called "hypotheses."[3] Support theory conceives of probability as a judgment of the proportion of evidence favoring a focal hypothesis (e.g., "rain in Seattle next Tuesday") relative to its complement ("no rain in Seattle next Tuesday"). People may recruit evidence for focal and complementary hypotheses using judgmental heuristics, explicit arguments, or objective data. For instance, the probability of rain in Seattle might be estimated by comparing the ease with which one can imagine a day with rain to the ease with which one can imagine a day without rain. If the former is equally easy as the latter then the probability might be estimated as ½ (i.e., odds of 1:1).

In support theory evidence for the focal hypothesis and evidence for the complementary hypothesis receive equal weight in probability judgment. In contrast, research on heuristics such as availability and representativeness suggests that judgment of relative likelihood (e.g., "where is it more likely to rain?") can be assessed by comparing evidence for each focal hypothesis (e.g., "rain" in Seattle versus "rain" in Los Angeles) without reference to corresponding complementary hypotheses (days without rain in Seattle/Los Angeles), though this research does not mention explicitly the neglect of complementary hypotheses.

---

[3] In our application of support theory we will not distinguish between different descriptions of the same event, so this distinction will not be especially relevant for most of this paper. However, we will return to the distinction between events and their descriptions in the discussion.

The notion that complementary hypotheses receive less weight in relative than absolute likelihood judgment is consistent with prior literature in judgment and decision making. Research on the *compatibility principle* holds that the weight that a particular stimulus feature receives is enhanced by its compatibility with the response mode (Tversky, Sattath & Slovic, 1988; see also Goldstein & Einhorn, 1987; Mellers, Ordóñez, & Birnbaum, 1992). Assessment of absolute likelihood, particularly when made on a percentage or fractional probability scale, is compatible with proportion judgments. Thus, it may be natural to consider the proportion of total evidence favoring the focal event when making such assessments, as in support theory. Assessment of relative likelihood, however, is more compatible with a comparison of evidence strength for the events being contrasted (e.g., whether it is easier to recall days with rain recently in Los Angeles or Seattle) than a comparison proportions of evidence favoring each target event over its corresponding complement.

Complement neglect is also consistent with the observation that more prominent attributes tend to be afforded greater weight for tasks whose goal is to differentiate among options (Fischer & Hawkins, 1993; Fischer, Carmon, Ariely & Zauberman, 1999). In this case the focal hypotheses (e.g., rain in Seattle and Los Angeles) are more salient than the corresponding complementary hypotheses (i.e., no rain in Seattle and Los Angeles) because the rainy scenarios are explicitly mentioned.

The notion that people afford greater weight or attention to focal information relative to contrary information, even when the alternative information is equally diagnostic, has precedence in literatures ranging from comparative social judgments (Chambers & Windchitl, 2004) to judgments of subjective well-being (Wilson, Wheatley, Meyers, Gilbert, & Axsom 2000) to consumer choice (Hoch, 1985) to judgmental confidence (Koehler, 1991) to the judged probability of grouped hypotheses (Brenner & Rottenstreich, 1999). We refer to the tendency to underweight complementary events in relative likelihood judgment as the *complement neglect hypothesis*.

When a judge has roughly equivalent knowledge concerning events under consideration, complement neglect will have little effect: beliefs in relative and absolute likelihood judgment will coincide. However, complement neglect can lead to violations of both procedure invariance and ordinal complementarity in situations where a person can summon more evidence for one pair of complementary hypotheses than for another pair of complementary hypotheses. Most commonly this can occur when one pair of complementary hypotheses is more familiar than another pair. For instance, suppose an American economist is asked to assess the future unemployment rates in both the United States and Bangladesh. This economist may be able to conjure several compelling reasons why the U.S. unemployment rate could *rise or hold steady* in the coming year and an even greater number of compelling reasons why it could *fall*. That same economist, if pressed, might summon a couple of weak reasons why the unemployment rate in Bangladesh could rise or hold steady in the coming year and a couple of weak reasons why it could fall. The complement neglect hypothesis suggests that this economist will say that unemployment is “more likely to rise or hold steady” in the U.S. than in Bangladesh (because it is easier to come up with compelling reasons for this scenario in the U.S. than Bangladesh), but also say that unemployment is “more likely to fall” in the U.S. than in Bangladesh (because it is easier to come up with compelling reasons for this scenario in the U.S. than Bangladesh). In contrast, assuming this economist affords equal consideration to the focal and complementary hypotheses in absolute likelihood judgment, consistent with support theory, she will assign a higher probability to unemployment rising or holding steady in Bangladesh than the U.S.

In sum, the complement neglect hypothesis (in conjunction with support theory) predicts a tendency to judge the more familiar event "more likely" to occur than the less familiar event while at the same time: (1) assigning the more familiar event a lower probability than the less familiar event (thus violating procedure invariance); and (2) judging the *non-occurrence* of the familiar event "more likely" than that of the less familiar event (thus violating ordinal complementarity). We refer to violations of procedure invariance and ordinal complementarity as "belief reversals" and to the specific tendency to deem more familiar events more likely as "familiarity bias" (see Fox & Levav, 2000).

The remainder of this paper is organized as follows. We begin with a review of prior evidence and present new evidence for these two forms of belief reversal. Next, we develop a theoretical model that generalizes support theory (Tversky & Koehler, 1994; Rottenstreich & Tversky, 1997) by embedding it within the contingent weighting model (Tversky, Sattath & Slovic, 1988), and show how and when this "Contingent Weighting of Support" model (CWS) predicts each kind of belief reversal. Third, we explicitly test the fit of the CWS model to data and measure its parameters, thus directly testing the complement neglect hypothesis. We conclude with a discussion of extensions and implications of the model.

2. Evidence for Belief Reversals.

A. Violation of ordinal complementarity: Between-subject demonstrations.

Belief reversals are more likely to occur when people are presented with one complementary pair of hypotheses that are high in support and another complementary pair of hypotheses that are low in support. To demonstrate, we constructed a pair of mutually exclusive and exhaustive college majors (engineering and literature, see below) and assembled profiles (courses that two different students had enrolled in) that were designed to be either highly representative of both majors or unrepresentative of both majors. Studies have shown that the relative likelihood of various descriptions of a person (e.g. that they are a bank teller or a feminist) are predicted by the perceived representativeness of the profile with the descriptions (Tversky & Kahneman, 1983). Thus, we expected that a student who had enrolled in classes that were more representative of both majors would also be perceived to be *more likely* to have subscribed to both majors.

Method. We asked 76 students passing through a major walkway on the UCLA campus to complete the following item in exchange for their choice of a candy bar or soda:

> Aaron and Ben live in a dormitory that houses only Engineering majors and Literature majors. During their first year of college each took the following courses (among others):

AARON:	**BEN:**
Calculus	Biology
Classical Mythology	Music Composition
Computer Science	Philosophy
English Composition	Psychology
Physics	Spanish
Seminar on Shakespeare	U.S. History

In the order presented above, Aaron's course load was designed to be representative of both engineering and literature whereas Ben's course load was designed to be representative of neither major. We counterbalanced the order in which these profiles were presented.

Results. Eighty-nine percent of the participants who were presented with the descriptions in the order above (n = 38) indicated that Aaron. Thus, in order to satisfy ordinal complementarity, roughly 11% should have reported that Ben was more likely to be a literature major. However, when a second group of participants (n = 38) was who was more likely to be a literature major, 58% indicated Aaron, a strong violation of ordinal complementarity (89% + 58% = 147% >> 100%, $p < .0001$ by one-tailed Fischer's exact test).

In a previous paper (Fox & Levav, 2000), we provided several between-participant demonstrations of violations of ordinal complementarity that are summarized in Table 1A and Figure 1A. Most of these demonstrations entailed assessments of natural events for which the arguments for and against each event were not made explicit. For instance, consider the following item that was administered to Duke University students (Study 1). Duke students devote a great deal of time following men's basketball, but most know little or nothing about the fencing team. Hence, we reasoned that our participants would find both the target question and complementary question about Duke basketball to be highly familiar (labeled H and $\overline{H}$, respectively) and that they would the find target and complementary question about Duke fencing to be much less familiar (labeled L and $\overline{L}$, respectively). Our results documented a strong familiarity bias: 75% of participants in a first group said they thought it was "more likely" that Duke would beat the University of North Carolina (UNC) in an upcoming men's basketball game than in an upcoming men's fencing tournament. Thus, if ordinal complementarity holds, one would expect 25% of participants in a second group to indicate that it was "more likely" that UNC would beat Duke in men's basketball than fencing. Instead, a significantly higher proportion, 44%, indicated this belief. Put another way, the proportion of participants indicating that H is more likely than L plus the proportion of participants indicating that $\overline{H}$ is more likely than $\overline{L}$ was 75% + 44% = 119% > 100%, indicating a strong familiarity bias that violates ordinal complementarity.

We replicated this pattern in all five studies reported in Table 1A. As can be seen in column 7, participants exhibited a significant bias to order high familiarity (focal and alternative) hypotheses above corresponding low familiarity hypotheses. This can also be seen in Figure 1A as the sum of proportions choosing the high familiarity hypothesis and its complement summing to more than one for every study.

The question arises what pattern one might expect for ratings of which event is "less likely." One could argue that such a response mode would direct attention to reasons why relevant target events would *not* occur (i.e. the complementary events), thereby showing a bias to rate more familiar events and their complements *less likely*. However, the complement neglect hypothesis suggests instead that because "less likely" judgments entail explicit relative likelihood judgment, will follow a pattern similar to "more likely" judgments in which evidence for complementary events is underweighted, thereby showing a bias to rate less familiar events and their complements *less likely*. Indeed Fox and Levav's (2000) data reveal such a pattern (Studies 4 and 5). It is noteworthy that the propensity to exhibit a familiarity bias when rating which of two hypotheses is more likely is roughly equal to the propensity to exhibit an "unfamiliarity" bias when rating which of two hypotheses is less likely.

Interestingly, the question which hypothesis is "more unlikely" appears to translate to an evaluation of which complementary hypothesis is more likely rather than which focal hypothesis is "more likely." In two demonstrations Yamagishi (2002; Study 3) asked participants to evaluate which of two profiles was "more likely" or "more unlikely" to belong to a target category (e.g., "feminist"). One profile was both highly representative of the category (e.g., an outspoken former philosophy major interested in social justice and participated in anti-nuclear demonstrations) and also high unrepresentative of the same category (e.g., pro-life, active in church, and supports prayer in school). The other profile was neither particularly representative nor unrepresentative. Participants tended to rate the representative and unrepresentative profile as both "more likely" and "more unlikely" to be a member of the target category.

B. Violation of ordinal complementarity: Within-subject demonstrations.

The violations of ordinal complementarity reviewed above all relied on between-subject designs. To test the robustness of these violations, we also explored whether familiarity bias persists in within-subject designs.

Participants. We recruited participants ($N = 35$) from bulletin boards of online groups dedicated to each of 29 National Basketball Association (NBA) teams, as well as two groups devoted to general professional basketball and fantasy league discussion. We offered participants a chance to win a $50 cash prize in exchange for completing the survey.

Procedure. We asked participants to rate eight events relating to the upcoming NBA collegiate player draft using a scale from 1 "most likely" to 8 "least likely," using each ranking only once. The questions concerned both familiar and unfamiliar players that had appeared on a list of NBA prospects culled and rated by a leading draft expert and national television commentator. The relative draft position of four players served as our target events: Elton Brand and Steve Francis (familiar; these players were the top two picks in the NBA draft that year) and Keith Carter and Roberto Bergerson (unfamiliar; Carter went undrafted and Bergerson was the 52nd pick). The order of the events was randomly determined and appeared as follows:

___ Elton Brand is chosen ahead of Steve Francis in the draft. (H)
___ Scott Padgett is a second round pick in the draft.
___ Keith Carter is chosen ahead of Roberto Bergerson in the draft. (L)
___ Ademola Okulaja is selected in the first round of the draft.
___ Tim Young is selected ahead of Heshimu Evans in the draft.
___ Steve Francis is chosen ahead of Elton Brand in the draft. ($\overline{H}$)
___ Roberto Bergerson is chosen ahead of Keith Carter in the draft. ($\overline{L}$)
___ Scott Padgett is a lottery pick (picks 1-13) in the draft.

We randomly assigned respondents either to the above questionnaire or to a version that presented the events in the opposite order.

Results. We coded participants' responses into three categories: familiarity bias (either listing both of the highly familiar over both of the less familiar, $H > H > L > L$, or alternating these rankings $H > L > H > L$), a reverse, "unfamiliarity" bias (either listing both the less familiar over both of the highly familiar, $L > L > H > H$, or alternating these rankings $L > H > L > H$), and a normatively defensible ranking (either listing $H > L > L > H$ or $L > H > H > L$).[4] Of the 35 participants, we found that 10 conformed to the familiarity bias pattern, only 1 conformed to the unfamiliarity bias pattern, and the remaining 24 indicated normatively defensible rankings. Although the familiarity bias here appears small, note that our study is quite conservative because of its transparency—it takes little effort for participants to see that the focal and alternative hypotheses appear in the same short list, creating a strong demand for consistency. If participants had given random responses, the proportion indicating each of these three categories would be roughly equal, a null hypothesis which clearly can be rejected ($\chi^2(2) = 13.4$, $p = .001$). Moreover, participants were ten times more likely to indicate a familiarity bias than an unfamiliarity bias ($p = .01$ by sign test).

C. Violation of procedure invariance: Between-subjects demonstrations.

Table 1B and Figure 1B present examples of violations of procedure invariance from Fox & Levav (2000). For these items we asked groups of respondents to judge the probability of the highly familiar hypothesis (and its complement) and less familiar hypothesis (and its

[4] Because we present complementary events and the designation of H versus $\overline{H}$ and L versus $\overline{L}$ is arbitrary, we do not distinguish between events and their complements in the analysis of these results (e.g. we treat $H > \overline{H} > L > \overline{L}$ the same as $\overline{H} > H > L > \overline{L}$).

complement).[5] For example, in Study 7, we asked a group of American participants whether they thought that it was "more likely" that "the winner of the next U.S. Presidential election [is/is not] a member of the Democratic Party" or "the winner of the next British Prime Ministerial election [is/is not] a member of the Labor party" [alternative wording in brackets]. Other groups were asked to judge the probabilities of the same events. Thus, roughly half the participants ordered their beliefs over these events directly and for the remaining participants we inferred these orderings from their judged probabilities. This design provided tests of both varieties of belief reversal (violations of procedure invariance and of ordinal complementarity). First, although 64% of participants in the relative likelihood condition indicated that a Democratic winner was "more likely" than a Labor winner, only 36% of participants in the absolute likelihood condition reported a higher probability for a Democratic winner than a Labor winner. Likewise, although 76% of participants in the relative likelihood condition indicated that a non-Democratic winner was "more likely" than a non-Labor winner, 73% of participants provided a higher probability for Democrat than Labor.[6] Taken together, there was a significant violation of ordinal complementarity evident in the relative likelihood conditions, but not in the absolute likelihood conditions: The proportion of participants rating Democrat more likely than Labor plus the proportion of participants rating non-Democrat more likely than non-Labor was a striking 140% ($p < .005$). Meanwhile, the proportion of participants providing a higher probability for Democrat than Labor plus the proportion providing a higher probability for non-Democrat than non-Labor summed to 109% (n.s.), a significant interaction

[5] In Study 6 we only elicited relative likelihood and probability judgments for a focal hypothesis and not the complementary hypothesis. This allowed a test of violations of procedure invariance, but not ordinal complementarity.

[6] We broke ties in the ordinal analysis of judged probabilities by assigning half to $H > L$ and half to $L > H$, and so forth. For a fuller description of the frequency of ties see Table 1 of Fox & Levav (2000).

($p < .05$). In each of the studies listed there is a significantly stronger tendency to rate the more familiar hypothesis "more likely" than assign it a higher probability.[7]

D. Violations of procedure invariance: Ranking versus rating likelihood.

We next attempt to replicate our within-participant demonstration of familiarity bias in relative likelihood judgment and determine whether this effect will be significantly attenuated when participants are asked instead to rate the absolute likelihood of each event.

Participants. One hundred and seventy-one MBA students at Duke University participated in this study, which was embedded in a larger survey packet. A $10 charitable contribution was made in the name of each respondent in exchange for his or her participation.

Procedure. Participants responded to one of two versions of a questionnaire item. In one condition ($n = 73$) participants were asked to rank a set of eight events from "most likely" to "least likely." In a second condition they ($n = 97$) were asked to provide their best estimate of the probability for each of the same eight events. Events concerned the future relative positions of pairs of graduate programs in the *U.S. News and World Report* graduate program ranking. Two target events and their complements (i.e., four events), one familiar and one unfamiliar, were combined with four filler events. The events are listed below, with the target events marked:

U. of Pennsylvania is ranked above Yale among medical schools.
Syracuse is ranked above Rutgers among library sciences programs. (L)
Wharton (U. Penn.) is ranked above Harvard among business schools. (H)
MIT is ranked above Michigan among physics departments.
Columbia is ranked above U. of Pennsylvania among medical schools.
Rutgers is ranked above Syracuse among library sciences programs. ($\overline{L}$)
UCLA is ranked above Cornell among physics departments.
Harvard is ranked above Wharton (U. Penn) among business schools. ($\overline{H}$)

[7] This tendency is obviously more pronounced in Studies 7 & 8 among pairs of hypotheses for which a minority of participants assigned a higher probability to the more familiar hypothesis. Naturally, when most participants rate the more familiar hypothesis more probable, ceiling effects will come into play and there will be "less room" for a stronger tendency among participants assessing relative likelihood.

We assumed that because our participants were incoming MBA students, the items regarding business schools would be quite familiar to them, while the items regarding library sciences would be wholly unfamiliar.[8] The order of event presentation was determined at random and counterbalanced.

Results. As in the NBA draft study, we coded our participants' responses into three categories in the ranking and probability conditions: familiarity bias, "unfamiliarity" bias, or a normatively defensible ranking. Once again, had participants been responding at random, the proportion indicating each of these three categories would be roughly equal. In contrast, of the 73 participants in the ranking condition, we found that 35 conformed to the familiarity bias pattern, only 1 conformed to the unfamiliarity bias pattern, and the remaining 37 indicated normatively defensible rankings ($\chi^2(2) = 35.1$, $p < .0001$). The corresponding numbers for the probability condition were 25, 2, and 70, respectively ($\chi^2(2) = 72.2$, $p < .0001$). We speculate that the residual familiarity bias in the probability condition was a result of some participants having anchored their probabilities on events that they had rated and adjusting according to whether subsequent items appeared more or less likely. More importantly, the results indicate that the familiarity bias was significantly attenuated in the probability condition relative to the ranking condition (25% versus 48%, respectively; $\chi^2(1) = 8.97$, $p < .005$), a highly significant interaction.

E. Restoring Symmetry.

---

[8] In a contemporaneous survey of students admitted to Duke's MBA program, 99% indicated that they had used *Business Week* and/or *U.S. News & World Report*'s published rankings of business schools in deciding which business school to attend.

According to the present account, belief reversals arise from a tendency to underweight the complementary hypothesis in relative likelihood judgment but not in absolute likelihood judgment. This leads us to the conjecture that we may induce greater symmetry in weighting of the focal and complementary hypotheses, and hence attenuate complement neglect, by drawing attention to the complementary hypothesis immediately prior to eliciting relative likelihood judgment (cf. Koehler, 1991). To test this notion we asked students to predict which set of films (domestic or foreign) was more likely to win its respective Academy Award. We drew some participants' attention to the complementary hypothesis by asking them to first indicate whether they preferred that one of the films in the focal set or complementary set would win its award category.

Participants. One hundred and seven MBA students at Columbia University participated in our study as unpaid volunteers.

Procedure. Participants were told that we were interested in their answers to questions about films that had been nominated for a Best Picture Academy Award ("The Aviator," "Finding Neverland," "Million Dollar Baby," "Ray" and "Sideways") and Best Foreign Film Academy Award ("As It Is in Heaven," "The Chorus," "Downfall," "The Sea Inside" and "Yesterday"). We anticipated that our U.S.-based participants would be more familiar with the films nominated for the Best Picture award than the Best Foreign Film award. We listed all films nominated in both categories so that participants would not be forced to rely on their memories to answer subsequent questions.

After presenting this background information we asked participants three questions. The relative likelihood question was worded as follows alternate wording in [brackets]:

> *Which of the following two options do you think is more likely to occur (check either Option A or Option B below):*

_____ Option A: [“The Aviator” OR “Ray” / “Finding Neverland” OR “Million Dollar Baby” OR “Sideways”] wins best picture.

_____ Options B: [“As It Is in Heaven” OR “The Sea Inside” / “The Chorus” OR “Downfall” OR “Yesterday”] wins best foreign film.

We also asked two questions about which cluster of films included the participant’s “personal preference” to win its category’s award, Best Picture and Best Foreign Film (e.g., Which do you personally prefer?: A) “The Aviator” or “Ray” wins Best Picture; or B) “Finding Neverland,” “Million Dollar Baby” or “Sideways” wins Best Picture).

Our manipulation consisted of varying the order of the likelihood and preference questions. In the control condition the likelihood question was asked before the preference questions; in the “complement salient” condition the preference questions were asked before the likelihood question. We expected that the preference questions would draw attention to the complementary hypotheses (i.e., non-focal movies), thereby attenuating complement neglect and the resulting familiarity bias.

Results. Table 2 and Figure 2 present the number of participants endorsing the familiar ($H/\overline{H}$) or unfamiliar ($L/\overline{L}$) option in each condition. As predicted, we found a strong familiarity bias in the control condition that was significantly attenuated in the complement salient condition ($z = 2.80$, $p < .005$). These results support the notion that complement neglect is a consequence of underweighting the complementary hypothesis and that it can be reversed by explicitly drawing attention to these hypotheses.

3. Theory

To formalize the complement neglect hypothesis we begin with the theoretical foundation of support theory (Tversky & Koehler, 1994; Rottenstreich & Tversky, 1997), in which subjective probability is not attached to events, as it is in other models, but rather to descriptions of events, called *hypotheses*. Hence, two descriptions of the same event may be assigned different probabilities (i.e., the model is nonextensional). For instance, the hypothesis "precipitation in Chicago on April 1" might be assigned a lower probability than the hypothesis "rain or snow or sleet or hail in Chicago on April 1" (For other examples, see Sloman et al., 2004; Fox & Rottenstreich, 2003; Fox & Levav, 2004). In the demonstrations presented in this paper we will assume a canonical description of each event, and will therefore it will not be necessary to distinguish between events and hypotheses, but for completeness we note that our model retains this feature of support theory and can therefore accommodate nonextensionality.

Support theory assumes that each hypothesis A has a nonnegative support value $s(A)$ corresponding to the strength of evidence for this hypothesis. The judged probability $P(A, \overline{A})$ that hypothesis A rather than $\overline{A}$ holds, assuming that one and only one of them obtains is given by:

(1) $$P(A,\overline{A}) = \frac{s(A)}{s(A)+s(\overline{A})}.$$

Thus, judged probability is interpreted as support for the focal hypothesis, A, relative to the complementary hypothesis, $\overline{A}$. For example, the probability of rain tomorrow (A) rather than no rain ($\overline{A}$) is assumed to be the support for rain divided by the sum of support both for and against rain. It is convenient to translate eq. (1) into an odds metric:

(2) $$R(A,\overline{A}) \equiv \frac{P(A,\overline{A})}{1-P(A,\overline{A})} = \frac{s(A)}{s(\overline{A})}.$$

Note that R is a notational device that is derived from judgments of probability, and that the ordering of hypotheses by odds is formally equivalent to the ordering of hypotheses by judged probability.

Contingent weighting in ordering beliefs.

Consider two (not necessarily exclusive) hypotheses H and L whose complements are $\overline{H}$ and $\overline{L}$, respectively. Let $\geq_i$ be the belief ordering of hypotheses under elicitation mode i ($i = P, M$), where P refers to the belief ordering inferred from separately evaluated judged probabilities and M refers to the direct assessment of which hypothesis is "more likely." It readily follows from eq. (2) that $R(H,\overline{H}) \geq R(L,\overline{L})$ iff $\frac{s(H)}{s(\overline{H})} \geq \frac{s(L)}{s(\overline{L})}$, so that

$$H \geq_P L \text{ iff } \log s(H) - \log s(\overline{H}) \geq \log s(L) - \log s(\overline{L}).$$

This is merely a special case of the contingent weighting model (Tversky, Sattath & Slovic, 1988, Equation 5):

$$(3) \qquad H \geq_i L \text{ iff } \alpha_i \log s(H) - \beta_i \log s(\overline{H}) \geq \alpha_i \log s(L) - \beta_i \log s(\overline{L}),$$

with $i = P$, $\alpha_P = 1$, and $\beta_P = 1$. Here α_i and β_i reflect the relative weight in response mode i of evidence favoring the focal and their complementary hypotheses, respectively. In support theory the focal and complementary hypothesis receive equal and opposite weight. However, the complement neglect hypothesis asserts that support for the complement receives less weight than the focal in relative likelihood judgment: $\beta_m / \alpha_m \leq 1$, whereas the null hypothesis is that this ratio will not differ significantly from unity. Our conjecture that the complementary hypothesis will loom larger in cardinal judgment (of probability) than in ordinal judgment (of which event is more likely) can be expressed as $\beta_P / \alpha_P \geq \beta_m / \alpha_m$, whereas the null hypothesis is that these ratios will not significantly differ. We refer to equation (3) as the Contingent Weighting of Support (CWS) model.

Implications.

Of course, if focal and complementary hypotheses receive equal weight in ordinal judgment (i.e., $\alpha_m = \beta_m$), then belief orderings are determined by the proportion of support for the hypothesis in question (as in support theory) and are not affected by the absolute amount of support for the focal and alternative hypotheses. On the other hand, if $\alpha_m >> \beta_m$, then belief ordering is essentially determined by support for the focal hypothesis. Second, note that whenever the total amount of support for the focal and complementary hypothesis is the same for both events (i.e., $s(H) + s(\overline{H}) = s(L) + s(\overline{L})$), then the belief ordering over *H* and *L* will be the same regardless of the relative weight to the focal versus complementary hypothesis.[9]

If the sum of support for and against one hypothesis is greater than support for and against a second hypothesis (i.e., $s(H) + s(\overline{H}) > s(L) + s(\overline{L})$) and if complement neglect is present in relative likelihood judgment and more pronounced than in absolute likelihood judgment (i.e., $\beta_m / \alpha_m \leq \beta_P / \alpha_P \leq 1$), then two patterns emerge. First, situations can arise in which the familiar event is deemed both more likely to occur than the unfamiliar event ($H >_m L$) and more likely *not* occur than the unfamiliar event ($\overline{H} >_m \overline{L}$), a violation of ordinal complementarity. Second, situations can arise in which the more familiar event is deemed more likely ($H >_m L$) but is assigned a lower probability ($L >_P H$), a violation of procedure invariance. We next derive more specific conditions under which these two forms of belief reversal are expected to emerge.

[9] To see why, suppose the total amount of support for both pairs of hypotheses equals some constant (i.e., $s(H) + s(\overline{H}) = s(L) + s(\overline{L}) = C$). In this case, $s(H) \geq s(L)$ iff $s(\overline{H}) \leq s(\overline{L})$ so that $H \geq_i L$ for all α>0 and $\beta \geq 0$. By way of analogy, if one assesses which city is rainier by counting the number of days per year of rain in each city or the proportion of days of rain per year, one reaches the same conclusions because the total number of days in a year is the same in both cities.

First, we assume the CSW model (Eq.3), and no complement neglect for judged probability, $\beta_m / \alpha_m \leq \beta_P / \alpha_P = 1$. Next, we need to define parameters measuring: (a) the extent of complement neglect in relative likelihood judgment and (b) the extent to which total support differs for the higher- versus lower-familiarity hypothesis pairs. The domain of belief reversals is expected to increase in both variables. Define the ratio of weight to the alternative to focal hypotheses:

$$\kappa_m \equiv \beta_m / \alpha_m \text{, where } 0 < \kappa_m \leq 1,$$

and total support for the high familiar and low familiarity domains as follows:

$$h \equiv s(H) + s(\overline{H}),\ l \equiv s(L) + s(\overline{L}) \text{, where } h \geq l.$$

Now, define the ratio of total support in the low familiarity to high familiarity domains,

$$\sigma \equiv l / h \text{ where by definition } 0 < \sigma \leq 1.$$

Finally, we define two variables that determine the conditions under which we expect belief reversals, given a particular combination of complement neglect (k_M) and support asymmetry (σ) parameter values: the proportion of support on the focal hypothesis for each domain. Define p and q as the proportion of support favoring the focal hypothesis for the high and low familiarity domains, respectively,

$$p \equiv \frac{s(H)}{s(H) + s(\overline{H})},\ q \equiv \frac{s(L)}{s(L) + s(\overline{L})},$$ (note that under support theory $H \geq_P L$ iff $p \geq q$).

It can be shown (see Appendix 1) that the following belief reversals are expected under the following conditions:

(4a) $H \geq_m L$ and $L \geq_P H$ whenever $\frac{p}{(1-p)^{\kappa_m}} \geq \frac{q\sigma}{[(1-q)\sigma]^{\kappa_m}}$ and $q \geq p$,

in which case we also get $\overline{H} \geq_m \overline{L}$, for a second belief reversal, and

(4b) $\overline{H} \geq_m \overline{L}$ and $\overline{L} \geq_P \overline{H}$ whenever $\dfrac{(1-p)}{p^{\kappa_m}} \geq \dfrac{(1-q)\sigma}{[q\sigma]^{\kappa_m}}$ and $p \geq q$ (i.e. 1-$q \geq$ 1-p),

in which case we also get $H \geq_m L$, for a second belief reversal.

Figure 3 depicts the conditions under which both forms of belief reversal are expected, according to Equations (4a) and (4b). In each panel, p (the proportion of total support favoring the focal in the case of the high familiarity event) is depicted on the vertical axis and q (the proportion of total support favoring the focal in the case of the low familiarity event) is depicted on the horizontal axis. Columns indicate different levels of the relative weight to alternative hypothesis, κ_m = .2, .5, .8. Rows indicate levels of total support for the low familiarity domain relative to the high familiarity domain, σ= .2, .5, .8. In each case the triangle above the identity line represents the region in which $H \geq_m L$, whereas the triangle below the identity line represent the region in which $\overline{H} \geq_P \overline{L}$. The upper triangle also represents the region in which $H \geq_m L$, which also extends to the lower "lip" that is bounded by the contour below the identity line. Similarly, the lower triangle represents the region in which $\overline{H} \geq_m \overline{L}$, which also extends to the upper "lip" that is bounded by the contour above the identity line.

Thus, the lips in each panel represent the regions of belief reversals: the lower lip is the region in which $L \geq_P H$ and $H \geq_m L$ and the upper lip is the region in which $\overline{L} \geq_P \overline{H}$ and $\overline{H} \geq_m \overline{L}$ (violating procedure invariance); both lips represent the region in which $H \geq_m L$ and $\overline{H} \geq_m \overline{L}$ (violating ordinal complementarity). Not surprisingly, these belief reversals are expected to be more prevalent as σ and κ_M decrease (i.e., we move to the left and up, respectively).

4. Fitting the CWS model.

In this section we fit the Contingent Weighting of Support model and directly test the complement neglect hypothesis. In the experiments that follow we asked participants to assess the relative likelihood that various teams in a sports tournament will win their respective match-ups, the absolute likelihood that each team will win its match-up, and the relative strength of these teams (as a proxy for support). We must first show how the CWS model can be tested from these data. Equation (3) can be restated as follows:

(6) $H \geq_i L$ iff $\alpha_i[\ln s(H) - \ln s(L)] - \beta_i[\ln s(\overline{H}) - \ln s(\overline{L})] \geq 0$.

To test the complement neglect hypothesis we must obtain measures of the weight afforded the focal hypothesis, α_i, and the weight afforded the alternative hypothesis, β_i, for absolute and relative likelihood judgment. To do so we must also obtain a measure of raw evidence strength. Recall that support for a hypothesis (e.g., UCLA beats USC in an upcoming basketball game) is a hypothetical construct that can only be inferred from corresponding probability judgments. However, Tversky & Koehler (1994) observe that by invoking minimal assumptions, support for hypothesis A derived from judged probability, $s(A)$, can be related to the raw rating of evidence strength (e.g., the perceived strength of the UCLA and USC teams), $\hat{s}(A)$, by a power function, $s(A) = [\hat{s}(A)]^k$. These assumptions were empirically verified by Fox (1999) in the context of a sports tournament (see also Koehler, 1996). Plugging this relation into Equation 6, we get:

$$H \geq_i L \text{ iff } \alpha_i^{'}[\ln \hat{s}(H) - \ln \hat{s}(L)] - \beta_i^{'}[\ln \hat{s}(\overline{H}) - \ln \hat{s}(\overline{L})] \geq 0,$$

where $\alpha_P^{'} = k\alpha_P$, $\beta_P^{'} = k\beta_P$. Thus, we can fit the CWS model using logistic regression:

(7) $\ln[(\Pr(H >_i L)/(1-\Pr(H >_i L))] = \alpha_i'[\ln\hat{s}(H) - \ln\hat{s}(L)] - \beta_i'[\ln\hat{s}(\overline{H}) - \ln\hat{s}(\overline{L})]$,

where the $\Pr(H >_i L)$ is the proportion of responses in which H is rated above L in response mode i. According to the complement neglect hypothesis, we should find that:

$$1 \approx \beta'_P/\alpha'_P \geq \beta'_M/\alpha'_M.$$

Recall the responses in equation (7) are coded by whether hypothesis H is rated above hypothesis L based on separate assessments of probability. The logistic analysis above assumes that the probability of rating H over L using judged probability (absolute likelihood) or direct assessment of relative likelihood is a function of weighted strength of focal and alternative teams.

We can also examine whether the extremity in participants' ratings can be predicted in the same manner. According to support theory, we can derive weights on the focal and complementary hypothesis from log-odds (derived from judged probability) of hypotheses H and L:

$$\ln R(H,\overline{H}) - \ln R(L,\overline{L}) = \alpha'_P[\ln\hat{s}(H) - \ln\hat{s}(L)] - \beta'_P[\ln\hat{s}(\overline{H}) - \ln\hat{s}(\overline{L})]$$

Likewise, we might derive these weights from raw ratings of relative likelihood on a likert scale:

$$m(H\overline{H}, L\overline{L}) = \alpha'_m[\ln\hat{s}(H) - \ln\hat{s}(L)] - \beta'_m[\ln\hat{s}(\overline{H}) - \ln\hat{s}(\overline{L})],$$

where $m(H\overline{H}, L\overline{L})$ is the rating of the extent to which hypothesis H is rated as "more likely" than hypothesis L.

Method

We conducted a pair of studies to parameterize the CWS model. In these studies we asked participants to make various judgments about two upcoming college basketball tournaments, the Atlantic Coast Conference (ACC) tournament and the 3rd round of the National Collegiate Athletic Association (NCAA) tournament. The studies were conceptual replicates with identical procedures, so we present them simultaneously. The sporting domain is attractive for running a test of the CWS model because previous studies (Fox, 1999; Koehler, 1996) have demonstrated that it allows for direct assessment of support as rated strength of teams. Moreover, this domain provides a conservative test of complement neglect because the evaluation of the relative likelihood of teams winning competitive games demands at least some consideration of whom each team is playing (i.e., the complementary hypothesis).

Participants. We recruited a total of 648 participants (N = 384 in the ACC tournament study and N = 264 in the NCAA tournament study) through advertisements on ACC and NCAA basketball fan websites two days prior to the first round of the ACC tournament and the third round of the NCAA tournament, respectively. Each advertisement included a link to a website that contained the study and its instructions. Participants were offered a chance to win a $100 cash prize in exchange for completing the experiment.

Procedure. The survey program comprised three sections: likelihood judgment, probability judgment, and strength ratings. Four pairs of squads that were set to meet in the upcoming round were selected as target teams. Teams were chosen so that a broad range of strengths would be represented across the pairings.

In the likelihood section participants were asked to judge which of two teams was more likely to win its game. On each trial two pairs of teams that were set to face each other were presented on opposite sides of the screen (e.g., team A vs. team B on the left; team C vs. team D on the right), with one team from each pair serving as the focal hypothesis (the team name was highlighted). Participants indicated their judgment by clicking a point on a twenty-one point scale that ranged from –10 (team A is “much more likely” to win its game) to 0 (the teams are "equally likely" to win their respective games) to +10 (team *C* is "much more likely" to win its game). Note that for each pair of games there were four possible pairs of results that could be presented: Team *A* beats *B* and *C* beats *D*; *B* beats *A* and *C* beats *D*; *A* beats *B* and *D* beats *C*; and, *B* beats *A* and *D* beats *C*. Note also that there were six possible ways to matchup two pairs of teams from the set of four: pairs 1&2, 1&3, 1&4, 2&3, 2&4, 3&4. Thus, in total there were 6 match-ups × 4 possible outcomes to be evaluated = 24 permutations. The sequence of presentation of outcomes, as well as the sequence of presentation of match-ups within each outcome, was randomly determined for each participant. The program chose one of the outcomes at random, followed by one of the six match-ups within that outcome, sampling without replacement.[10] The position of each game result on the scale (i.e., on the -10 or +10 side) was also randomly determined. Participants were allowed to proceed only after entering a response, and allowed to click back only to the previous screen. The instruction screen for the likelihood judgments was accessible throughout this phase of the experiment.

Next, participants were instructed to complete 16 probability judgments in which they were asked to estimate the probability that a particular team would win its game. The four target matchups were divided into two sets of four events (e.g., "team *A* beats team *B*") and four complements (e.g., "team *B* beats team *A*"), with these sets separated by eight filler questions about college basketball trivia and other predictions. Each question was presented individually on the screen. Responses were entered on a twenty-one point scale that ranged from 0 to 100 in increments of 5. The order of presentation of each set of four target matchups was randomized both between the two sets and within each respective set.

---

[10] For the ACC sample the order of match-ups was only partially randomized due to a minor programming error. Because we found no significant effects of orderings in either study we surmise that this error had no effect on the results.

Finally, participants were asked to rate the strength of each of the eight target teams (cf. Fox, 1999; Koehler, 1996) on a scale from 1 to 100. They were asked to label the strongest team "100," and assess the strength of the remaining 7 teams relative to the team deemed strongest. Because it was crucial that participants make careful assessments, they were instructed to verify each of their ratings after completion. Participants could access the instruction screen as often as they liked.

Results.

To fit the CSW model (Equation 3), for each participant we conducted an ordered logistic regression in which we classified judged probabilities and explicit relative likelihood judgments into three categories: (1) AB > CD, (2) AB < CD, (3) AB = CD, where AB represents the event "team A beats team B," and inequalities represent relative ordering of judged probabilities and relative likelihood judgments.

Table 3 lists the mean weight to the focal and complementary hypotheses for both studies in relative and absolute likelihood judgment (i.e., values of κ_M and κ_p). Consistent with the complement neglect hypothesis, when making explicit relative likelihood ("more likely") judgments, participants placed significantly less weight on the complementary than focal teams. In contrast, when making absolute likelihood (probability) judgments, there was no significant difference in weights afforded the strength of focal and complementary teams. Likewise, the median ratio of weight to the complementary hypothesis relative to the focal hypothesis in relative likelihood judgment was .77 and .89 for the ACC and NCAA studies, respectively, whereas the median ratio for absolute likelihood judgment was 1.00 in both studies.[11]

As another test of complement neglect, we examine the proportion of participants who placed less weight on the complementary hypothesis relative to the focal hypothesis in relative than absolute likelihood judgment. This condition held for 66% and 66% of participants in the ACC and NCAA studies, respectively[12], $p < .0001$ by sign test in each case).

An obvious disadvantage of the logit analysis is that by allocating relative and absolute likelihood judgments to categories that order belief strength, we discard information concerning the magnitude of these differences. Thus, we replicated the foregoing analyses using a linear regression using the ratings of relative likelihood and the differences in judged probabilities as dependent variables. In all cases the qualitative results are similar to those we have reported. Table 4 reports the major results of the linear regression analysis. It is noteworthy that the adjusted R^2 were quite high for both probability and likelihood judgments in both studies, indicating a good fit of the model to the data. We also note that the model tended to fit the probability judgments slightly better than the likelihood judgments. This suggests that the more even weighting of support in absolute than relative likelihood cannot merely be attributed to a regression effect.

Discussion

In this paper we articulate and test a model of absolute and relative likelihood judgment called the "Contingent Weighting of Support" (CWS) model that embeds support theory

---

[11] Note that the effects cannot readily be explained by an inferior fit of the model to judged probability than relative likelihood judgment: median log-likelihood was -12.13 for P and -17.41 for M in the ACC study and -9.86 for P and -10.67 for m in the NCAA study.

[12] There were 22 ties for the ACC study and 3 ties for the NCAA study; we distributed ties equally in compiling these proportions.

(Tversky & Koehler, 1994; Rottenstreich & Tversky, 1997) within the contingent weighting model (Tversky, Sattath & Slovic, 1988). Moreover, we advance the *complement neglect hypothesis,* according to which people afford more weight to a focal event than to its complement when making relative likelihood judgments but not when making absolute likelihood judgments. This account predicts two forms of belief reversals. First, people will sometimes rate a more familiar event *A* to be "more likely" than a less familiar event *B,* and rate the more familiar complement of *A* more likely than the less familiar complement of *B,* violating the normative axiom of ordinal complementarity. Second, people will sometimes rate a more familiar event *A* "more likely" than a less familiar event *B*, but judge the probability of *B* higher than the probability of *A.* We provide evidence of both of these kinds of belief reversals, and test the CWS model directly. Results of two studies show that expert sports fans afford less weight to the strength of the alternative team than the strength of the focal team when judging the relative likelihood of teams winning their respective match-ups than when assessing probabilities, consistent with the complement neglect hypothesis. In the remainder of this section we review other forms of belief reversal, discuss the normative implications of the present findings, characterize how these findings relate to axioms of subjective probability, and suggest directions for future work.

Other forms of belief reversal.

In this paper we have argued that absolute and relative likelihood judgment promote differential attention to the focal versus alternative hypothesis. It is also worth noting that these modes of judgment differ in the number of events under consideration—whereas relative likelihood naturally entails consideration of two distinct events (e.g., rain in Seattle versus Los Angeles), absolute likelihood entails consideration of a single event (e.g., rain in Seattle). In the empirical demonstrations of belief reversals reported in this paper participants either evaluated the relative likelihood of two events, the absolute likelihood of two events, or engaged in both tasks. This was a deliberate design choice so that we would not confound elicitation mode (relative versus absolute likelihood) with the number of events being evaluated.

It is also worth asking whether we might find belief reversals in joint versus separate evaluation of hypotheses. Prior research on preferences has documented a number of joint-separate preference reversals (Hsee, Loewenstein, Blount & Bazerman, 1999), in which important attributes are easier to evaluate when options are considered together so that these attributes receive more weight in joint than separate evaluation. For instance, in one study the number of entries in a music dictionary (an important attribute that is hard to evaluate by itself) exerted a greater influence on willingness-to-pay than the condition of the dictionary's cover (a less important attribute that is easy to evaluate), but only when two dictionaries were evaluated together rather than separately (Hsee, 1996).

We explored the possibility of evaluability effects in likelihood judgment by asking participants to provide judgments of probability based on a pair of cues. Participants ($N = 80$) were MBA students completing the survey as part of larger packet of surveys, in exchange for a donation to charity. In one condition participants were asked to read the profile of a single medical school applicant ("David" or "Jerry"), including his undergraduate grade point average (GPA) and his Medical College Admissions Test (MCAT; this is the standardized test used in all medical school admissions in the United States). A second version of the questionnaire provided each applicant's profile in succession (David and Jerry). We assumed that participants would find the GPA cue to be evaluable when presented in isolation because of their familiarity with the scale, whereas the evaluability of the unfamiliar MCAT cue would benefit from the joint presentation of the applicants. Note that the MCAT score is a more diagnostic cue for acceptance to medical school than GPA. The two conditions, as well their respective estimated probabilities, appear in Table 5.

The results conform to our prediction. In joint evaluation, the less evaluable—but more diagnostic—cue weighed more heavily in participants' judged probabilities. While Jerry and David were judged equally likely to be admitted in separate judgments, David was judged much more likely to be admitted than Jerry in joint evaluation ($t = 2.32$, $p < 0.05$).[13] It appears that participants were better able to evaluate the MCAT score when the profiles were presented together than when they were presented separately.

Violations of binary complementarity in probability judgment

In the present paper we have focused on violations of ordinal complementarity in relative likelihood judgment and found little evidence of such violations in absolute likelihood judgment. Although a number of researchers have found the judged probabilities of complementary events generally sum to unity (e.g., Wallsten, Budescu & Zwick, 1992; Tversky & Koehler, 1994; Tversky & Fox, 1995; Fox & Tversky 1998), a few studies have documented systematic violations of binary complementarity, which in principle imply possible violations of ordinal complementarity that could be due to distinct mechanisms from those described here.

Macchi Osherson & Krantz (1999) present a study in which judged probabilities for highly unfamiliar complementary events (e.g., "The freezing point of gasoline [alcohol] is not equal to that of alcohol [gasoline]. What is the probability that the freezing point of gasoline [alcohol] is greater than that of alcohol [gasoline]?)) sum to slightly less than one, a violation of binary complementarity. Furthermore, they find that rephrasing each statements to make the

[13] This t-statistic was calculated following Hsee (1996): $t = ((M_{JA} - M_{JB}) - (M_{SA} - M_{SB}))/[(S_J^2/N_J + S_{SA}^2/N_{SA} + S_{SB}^2/N_{SB})]^2$, where A and B refer to the different students, and J and S refer to joint and separate evaluation, respectively.

alternative hypothesis more salient (e.g., "Either the freezing point of gasoline is greater than that of alcohol or the freezing point of alcohol is greater than that of gasoline. What is the probability that the freezing point of gasoline is greater than that of alcohol?") restored binary complementarity. This finding suggests an asymmetry of attention to focal and complementary hypotheses in probability judgment. Indeed, Idson, Krantz, Osherson & Bonini (2001) report examples in which judged probabilities of highly unfamiliar hypotheses and their complements sum to less than one while the judged probabilities of highly familiar hypotheses and their complements sum to more than one. To accommodate these findings they advance a modification of support theory in which for a common domain of statements, the probability of attending to alternative hypothesis less than 1. Judged probability is modeled as a weighted average of two terms: (1) the proportion of support for the focal hypothesis relative to the complementary hypothesis, as in support theory, and (2) the proportion of support for the focal hypothesis relative to a constant, reflecting complete complement neglect.

Yamagishi (2002, Study 4) provided participants with two personality profiles and asked them to judge the probability that one of the two described individuals was a member of a particular category assuming that one belonged to that category. A separate group evaluated the probability that the target individual was *not* a member of that category. The profile that was both more similar and dissimilar to the category elicited probabilities that summed to slightly more than 100%. In a subsequent study Yamagishi (2005; Study 5) determined through regression that affirmation probabilities tended to weight more highly features of the profile that were similar than dissimilar to the category, whereas negation probabilities tended to weight more highly features of the profile that were dissimilar than similar to the category. This result is consistent with a form of complement neglect in which feature similarity (dissimilarity) as an indicator of support for the focal hypothesis when eliciting affirmation (negation) probabilities.

McKenzie (1998) distinguishes between different cognitive representation of confidence in two hypotheses, *A* and *B*: (1) *dependent confidence*, in which the perceived truth of hypotheses are represented as alternative poles of a single scale—thus, an increases in confidence that hypothesis *A* is true necessitates a complementary decrease in confidence that *B* is true; and (2) *independent confidence* in which the perceived truth or falsity of each hypothesis is represented on separate scale—thus an increase in confidence that hypothesis *A* is true may or may not coincide with a change in confidence in hypothesis *B*. He argues further that the representation of confidence can be influenced by the whether information concerning hypotheses is learned in a *contrastive* manner (e.g., participants learn which of two complementary diagnoses is correct for patients with various patterns of symptoms) or *noncontrastive* manner (e.g., participants learn whether a diagnosis is correct for patients with various patterns of symptoms, separately for each diagnosis). In a series of studies, McKenzie (1998) finds that contrastive learning leads to judged probabilities that more closely coincide with binary complementarity than noncontrastive learning, and a greater tendency to see evidence for (against) a focal hypothesis as evidence against (for) the complementary hypothesis. For nonconstrastive learners, a pattern of symptoms implicating both diagnoses led to higher judged probabilities than a pattern that implicated neither. Intererstingly, when probabilities were elicited in a symmetric way that explicit mentioned both the focal and complementary hypothesis ("How confident are you that the patient has puneria rather than zymosis?") noncontrastive learners provided much more complementary probabilities compared to when probabilities were elicited in an asymmetric way that only mentioned the focal hypothesis ("How confident are you that the patient has puneria?").

In this paper we have focused on violations of ordinal complementarity in relative likelihood judgment that follow from complement neglect in that response mode (i.e., $\beta_M / \alpha_M < 1$), and violations of procedure invariance that follow from the notion that complement neglect is more pronounced in relative than absolute likelihood judgment (i.e., $\beta_M / \alpha_M < \beta_P / \alpha_P$). However, the CWS model can accommodate violations of ordinal complementary in judged probability by relaxing the assumption that $\beta_P / \alpha_P \approx 1$, and instead allowing $\beta_P / \alpha_P < 1$. Such patterns would be interpreted to reflect underweighting of complementary hypotheses in absolute likelihood judgment, which captures the spirit of the aforementioned theoretical accounts.

Alternative measures of belief strength.

In the present paper we have emphasized the contrast between relative likelihood judgment and absolute likelihood judgment. Relative likelihood judgment has been operationalized by asking participants which hypothesis is "more likely" or "less likely" or by asking them to rank the relative likelihood of several hypothesis. Absolute likelihood has been operationalized by asking participants to judge the probability for a focal event. Windschitl (2000; Study 1) argues that binary complementarity of judged probabilities does not necessarily reflect binary complementarity of "internal perceptions of certainty." In particular, he found that participants who were asked to express their degree of certainty using a verbal scale (e.g., …"extremely likely ", "quite likely", "fairly likely"…) provided ratings for complementary events that were less additive than did participants asked to evaluate the same events using numerical probabilities. Moreover, verbal certainty for focal events increased as support for both focal and complementary hypotheses increased. Although the interpretation of a verbal probability scale is open to question, this finding is consistent with the notion that probabilistic thinking prompts more evenhanded attention to focal and complementary events, whereas more qualitative thinking prompts some degree of complement neglect.

Indeed, when Windschitl (2000; Study 2) presented participants with a personality profile of Pat who was either a teacher or librarian but representative of neither, and a profile of Sally who was either a politician or journalist but representative of both. Participants who exhibited binary complementarity in their probability judgments nevertheless tended to both prefer to bet that Sally was a politician than Pat a teacher and also bet that Sally was a journalist than Pat a librarian. We suggest that this result may partly reflect complement neglect because betting choices naturally prompt consideration of which hypothesis is "more likely", and partly a tendency to merely prefer betting on events when one feels more confident judging probability (cf. Heath & Tversky, 1991; Fox & Tversky, 1995; Fox & Weber, 2002). Interestingly, the bias toward betting on the more representative profile was significantly attenuated among participants who first judged probability then made betting choices (compared to those who completed the tasks in the opposite order), consistent with the notion that restoring symmetry can attenuate the familiarity bias as we saw in our Academy Awards example above.

In a similar vein, Denis-Raj and Epstein (1994) observed that most undergraduate students in their sample preferred betting (to win $1) that they would draw a red jelly bean from a bowl containing 9 red jelly bean out of 100 than from a bowl containing 1 red jelly bean out of 10. The authors note that "subjects commonly commented that in spite of he stated odds, they felt that they had a better chance of winning by picking from the bowl with the more winning (red) beans (p.823)," a sentiment that is certainly consistent with complement neglect in intuitive relative likelihood estimates.

A hierarchy of axioms.

Theories of subjective probability characterize the relationship between events in the world and expressed beliefs. If expressions of belief are to convey any meaning they must conform to semantic or at least syntactical rules that are shared by the judge and observer. The present demonstration of belief reversals challenges the foundation of any reasonable axiomatic system of subjective probabilities. To elaborate this strong assertion we can examine a hierarchy of axioms on which various models of subjective probability rely.

At the top of such a hierarchy are the strongest assumptions, the correspondence axioms: in an ideal world, probabilities relate precisely to propensity for events to obtain in the world. For instance, we might hold people to be calibrated in their assessments: when Jack says he thinks there is a "30 percent chance" that an event will occur (e.g., "the Lakers will win the Championship this year") the event should occur precisely thirty percent of the time. Decades of research on calibration of experts and novices (e.g., Lichtenstein, Fischhoff & Phillips, 1982; Griffin & Tversky, 1991; Liberman & Tversky, 1993; Moore & Healy, 2008) demonstrate that the perfect calibration assumption is routinely violated, and in most contexts people exhibit overconfidence in judging probabilities.

More often, theories of subjective probability jettison the assumption that subjective probabilities are calibrated (i.e., semantic correspondence) and instead rely on the assumption that these beliefs accord with standards of coherence (i.e., are syntactically consistent). The strictest of these models assume that subjective probability measures are additive so that if $A \cap B = \phi$ then $P(A \cup B) = P(A) + P(B)$. Thus, when Jane says "the Lakers have a 30 percent chance to win the championship this year" and she says "the Celtics have a 20 percent chance to win the championship this year" we would also expect her to say that "there is a 50 percent chance that either the Lakers or Celtics will win the championship." However, there is ample evidence that judged probabilities routinely exhibit *subadditivity* (Tversky & Koehler, 1994; Rottenstreich & Tversky, 1997; Fox, 1999), and choices under uncertainty also reflect this tendency (Tversky & Fox, 1995; Fox & Tversky, 1998).

Noting such violations forces a retreat to a nonadditive measure of subjective probability, and several have been developed (for a review see, e.g., Dubois & Prade, 1988). Such measures generally preserve the ordinal properties of a probability measure. For instance if Bob says "there is a 50 percent chance that the Lakers or Celtics will win the championship this year," then we expect him to assign a smaller probability to the less inclusive possibility that the Lakers will win. This basic assumption of monotonicity is sometimes violated, as when a conjunction of features (e.g., a socially conscious young political activist named Linda is "both a Bank Teller and a Feminist") is judged to be more likely than one of these features alone (e.g., Linda is "a Bank Teller"; Tversky & Kahneman, 1983). Moreover, recent studies have found systematic violations of the assumption of extensionality that is required to represent subjective probabilities as a function of Boolean mappings from a state space to a belief measure. For instance, unpacking the description of an event (e.g. "precipitation next April 1 in Chicago") into a disjunction of constituent events ("rain or snow or sleet or hail next April 1 in Chicago") tends to increase the judged probability of that event (Rottenstreich & Tversky, 1997; for a review and for reversals see Sloman et al., 2004).

Given such failures of extensionality, some models have attempted to discard Boolean algebra and instead interpret subjective probabilities as a function of some other primitive. As noted in the present paper, support theory (Tversky & Koehler, 1994; Rottenstreich & Tversky, 1997; Brenner 2003) interprets subjective probability as a function of descriptions of events called "hypotheses." Such a framework can allow for subjective probabilities that exhibit unpacking effects. In order to maintain some formal structure, support theory implicitly retains weak axioms including ordinal complementarity and procedure invariance. For instance, note that binary complementarity, $P(A, \bar{A}) + P(\bar{A}, A) = 1$, implies ordinal complementarity.

The CWS model retains coherence by modeling judged probability as a constructed process in which more attention is paid to the focal hypothesis relative to the alternative in absolute likelihood judgment compared to relative likelihood judgment. While this model introduces additional parameters that must be measured (relative weights in different response modes) it can accommodate the systematic patterns of belief reversals that we document in this paper. In this respect the constructive process that we outline is akin to prior literature on the construction of preferences (e.g., Payne, Bettman & Schkade, 1999; Slovic, 1995; Lichtenstein & Slovic, 2006).

Implications and Conclusions

The present research may have interesting application to research on ambiguity aversion, the notion that decision makers prefer betting in situations where they have more knowledge or information concerning outcome probabilities (Ellsberg, 1961; Camerer & Weber, 1992). This phenomenon has been attributed to a preference to bet in situations where one is more knowledgeable or competent (Heath & Tversky, 1991; Fox & Weber, 2002). Moreover, ambiguity aversion is more pronounced when in separate evaluations of more familiar and less familiar events than in joint evaluations (Fox & Tversky, 1995). It is possible that this joint-separate evaluation effect is driven partly by beliefs rather than preferences. That is, it may be that joint evaluation prompts an assessment of relative likelihood that favors more familiar events whereas separate evaluation prompts assessment of absolute likelihood which leads to more evenhanded consideration of evidence for familiar and unfamiliar events.

Failure to fully consider alternative events in relative likelihood judgment may also partially underlie the *alternative outcomes effect* (Windchitl and Wells, 1998). In this phenomenon, holding objective probability of a focal event constant (e.g., you hold 3 out of 10 tickets to a raffle), subjective feelings of confidence increase when the most likely alternative event is less likely (e.g., seven other people each hold one ticket) than when the most likely alternative event is more likely (e.g., one other person holds seven tickets). Meanwhile, judged probabilities are unaffected by this manipulation. It may be that many nonnumeric assessments of uncertainty and expressions of concern tend entail comparisons between focal outcome and strongest alternative, whereas numerical probabilities reflect more thorough consideration of the complementary hypothesis.

The present research also has practical implications. We have found that judged probabilities reflect a more evenhanded consideration of evidence for and against the hypotheses under evaluation than does a direct assessment of relative likelihood. Thus, when assessing relative belief, one ought to solicit or assess absolute rather than relative likelihood judgments. This disparity is especially pronounced when the judge has substantially different degrees of knowledge or information concerning events. For instance, when asking a doctor for an assessment of the relative effectiveness of an established treatment and an experimental treatment, it would be wiser to separately ask for an assessment of likelihood that each will be successful than to ask which one is “more likely” to cure the disease.

Appendix 1:

Deriving Equation 4 from the CWS Model and Complement Neglect

Assume the CSW model (Eq.3),

$$H \geq_i L \text{ iff } \alpha_i \ln s(H) - \beta_i \ln s(\overline{H}) \geq \alpha_i \ln s(L) - \beta_i \ln s(\overline{L}).$$

Also assume support theory for absolute likelihood, and complement neglect for relative likelihood:

$$\beta_m / \alpha_m \leq \beta_P / \alpha_P = 1$$

Now define the ratio of weight to the alternative to focal hypotheses, $\kappa_M \equiv \beta_m / \alpha_m$, where $0 < \kappa_m \leq 1$, so that we get:

$$H \geq_m L \text{ iff } \ln s(H) - \kappa_m \ln s(\overline{H}) \geq \ln s(L) - \kappa_m \ln s(\overline{L}).$$

Noting that $\ln A - c \ln B = \ln \frac{A}{B^c}$, and restricting attention for the moment to the right-hand inequality, we get: $\frac{s(H)}{s(\overline{H})^{\kappa_m}} \geq \frac{s(L)}{s(\overline{L})^{\kappa_m}}$ so that $\frac{s(H)}{s(L)} \geq \left[\frac{s(\overline{H})}{s(\overline{L})}\right]^{\kappa_m}$.

Now, define total support for the highly familiar and less familiar domains as follows:

$$h \equiv s(H) + s(\overline{H}),\ l \equiv s(L) + s(\overline{L}), \text{ where } h \geq l.$$

Also, define p and q as the proportion of support favoring the focal hypothesis for the high and low familiarity domains, respectively, $p \equiv \frac{s(H)}{s(H) + s(\overline{H})}$, $q \equiv \frac{s(L)}{s(L) + s(\overline{L})}$, (note that $H \geq_P L$ iff $p \geq q$). We get $\frac{ph}{ql} \geq \left[\frac{(1-p)h}{(1-q)l}\right]^{\kappa_m}$.

Finally, define the ratio of total support in the low-familiarity to high familiarity domains,

$\sigma \equiv l/h$, where by definition $0 < \sigma \leq 1$. We get: $\frac{p}{q\sigma} \geq \left[\frac{(1-p)}{(1-q)\sigma}\right]^{\kappa_m}$, so that adding the first

condition back we get:

(5a) $H \geq_m L$ iff $\frac{p}{(1-p)^{\kappa_m}} \geq \frac{q\sigma}{[(1-q)\sigma]^{\kappa_m}}$.

Note that these conditions will be met whenever $p \geq q$ (i.e., $H \geq_P L$), and sometimes when $q < p$ (i.e., $L \geq_P H$).

Conditions for complementary hypotheses can be generating by replacing *p* with (1-*p*) and *q* with (1-*q*), so that we get:

(5b) $\overline{H} \geq_m \overline{L}$ iff $\frac{(1-p)}{p^{\kappa_m}} \geq \frac{(1-q)\sigma}{[q\sigma]^{\kappa_m}}$.

Note that these conditions will be met whenever $q \geq p$ (i.e., $\overline{H} \geq_P \overline{L}$), and sometimes when $p < q$ (i.e., $\overline{L} \geq_P \overline{H}$).

Returning to Equation (5a), note that if $\kappa_M = 1$, then σ cancels out of the right-hand term, and we simply have an expression of odds on both sides, or:

$$H \geq_m L \text{ iff } H \geq_p L.$$

Note that if $\sigma = 1$, then the right-hand term becomes

$$\frac{p}{(1-p)^{\kappa_m}} \geq \frac{q}{(1-q)^{\kappa_m}},$$

which is trivial because $p \geq q$ iff $(1-p) \leq (1-q)$ iff $(1-p)^k \leq (1-q)^k$ for all k, so that again,

$$H \geq_m L \text{ iff } H \geq_p L$$

More generally, when $\sigma < 1$ and $\kappa_M < 1$ we obtain the following belief reversals:

(4a) $H \geq_m L$ and $L \geq_P H$ whenever $\frac{p}{(1-p)^{\kappa_m}} \geq \frac{q\sigma}{[(1-q)\sigma]^{\kappa_m}}$ and $q \geq p$,

in which case we also get $\overline{H} \geq_m \overline{L}$, for a second belief reversal, and

(4b) $\overline{H} \geq_m \overline{L}$ and $\overline{L} \geq_P \overline{H}$ whenever $\frac{(1-p)}{p^{\kappa_m}} \geq \frac{(1-q)\sigma}{[q\sigma]^{\kappa_m}}$ and $p \geq q$ (i.e. 1-$q \geq$ 1-p),

in which case we also get $H \geq_m L$, for a second belief reversal.

References

Brenner, L. (2003). A random support model of the calibration of subjective probabilities. *Organizational Behavior and Human Decision Processes, 90,* 87-110.

Brenner, L. & Rottenstreich, Y. (1999).Focus, repacking, and the judgment of grouped hypotheses. *Journal of Behavioral Decision Making, 12,* 141-148.

Camerer, C. F., & Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty, 5*, 325-70.

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: The role of nonmotivated factors in above-average and comparative optimism effects. *Psychological Bulletin*, *130*, 813-838.

Denes-Raj, V. & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology, 66,* 819-829.

Dubois, D. & Prade, H.(1988). Modelling uncertainty and inductive inference: A survey of recent non-additive probability systems. *Acta Psychologica 68,* 53-78.

Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms. *Quarterly Journal of Economics*, *75*, 643-69.

Fishburn, P.C. (1986). The Axioms of subjective probability. *Statistical Science, 1,* 335-358.

Fischer, G., Carmon, Z., Ariely, D., & Zauberman, G. (1999). Goal-based construction of preferences: Task goals and the prominence effect. *Management Science*, *45*, 1057–1075.

Fischer, G. W. & Hawkins, S. A. (1993). Scale compatibility, strategy compatibility and the prominence effect. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 580-597.

Fitts, P. M., & Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, *46*, 199–210.

Fox, C. R. (1999). Strength of evidence, judged probability, and choice under uncertainty. *Cognitive Psychology, 38*, 167–189.

Fox, C. R. & Levav, J. (2000). Familiarity bias and belief reversal in relative likelihood judgment. *Organizational Behavior and Human Decision Processes*, *86*, 268-292.

Fox, C.R., and Rottenstreich, Y. (2003). Partition priming in judgment under uncertainty. *Psychological Science, 14*, 195-200.

Fox, C. R., and Tversky, A. (1998). A Belief-based account of decision under uncertainty. *Management Science*, *44*, 879-895.

Fox, C. R., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *Quarterly Journal of Economics*, *110*, 585–603.

Fox, C.R., & Weber, M. (2002). Ambiguity aversion, comparative ignorance, and decision context. *Organizational Behavior and Human Decision Processes, 88,* 476-498.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.

Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomenon. *Psychological Review*, *94*, 236–254.

Griffin, D., & Tversky, A. (1992). The weighting of evidence and the determinants of confidence. *Cognitive Psychology*, *24*, 411–435.

Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty*, *4*, 4–28.

Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*, 719-731.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, *67*, 247-257.

Hsee, C. K, Loewenstein, G. F., Blount, S. & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluation of options: A review and theoretical analysis. *Psychological Bulletin*, *125*, 576-590.

Idson, L. C., Krantz, D. H., Osherson, D. & Bonini, N. (2001). The relation between probability and evidence judgment: An extension of support theory. *Journal of Risk and Uncertainty*, *22*, 227-249.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58* , 697-720.

Kahneman, D. & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). New York: Cambridge University Press.

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

Koehler, D. J. (1996). A strength model of probability judgments for tournaments. *Organizational Behavior and Human Decision Processes, 66,* 16-21.

Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin, 110*, 499-519.

Krantz, D.H., Luce, R.D., Suppes, P., & Tversky, A. (1971). *Foundations of Measurement, Vol. 1.* New York: Academic Press.

Liberman & Tversky (1993)

Lichtenstein, S., & Slovic, P. (Eds.) (2006). *The construction of preference.* New York: Cambridge University Press.

Lichtenstein, S., Fischhoff, B., & Phillips, L.D. (1982). Calibration of probabilities: The state of the art to 1980. Chapter 22 (pp. 306-334) of D. Kahneman, P. Slovic & A. Tversky (Eds.). *Judgment Under Uncertainty: Heuristics and Biases.* (Cambridge, UK: Cambridge University Press).

McKenzie, C.R.M. (1998). Taking into account the strength of an alternative hypothesis. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24,* 771-792.

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

Mellers, B.A., Ordóñez, L.D., & Birnbaum, M.H. (1992). A Change-of-Process Theory for Contextual Effects and Preference Reversals in Risky Decision Making. *Organizational Behavior and Human Decision Processes, 52,* 331-369.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502-517.

Payne, J. W., Bettman, J. R., & Schkade, D. A. (1999). Measuring constructed preferences: Towards a building code. *Journal of Risk and Uncertainty*, *19*, 243-270.

Rottenstreich, Y. & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review, 104*, 406–415.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory and Cognition*, *21*, 546–556.

Shah, A.K., & Oppenheimer, D.M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin, 134,* 207-222.

Sloman, S., Rottenstreich, Y., Wisniewski, E., Hadjichristidis, C., & Fox, C.R. (2004). Typical versus atypical unpacking and superadditive probability judgment. *Journal of Experimental Psychology: Learning, Memory & Cognition, 30*, 573-582.

Slovic, P. (1995). The construction of preference. *American Psychologist, 50*, 364-371.

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

Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, *102*, 269-283.

Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293-315.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, *101*, 547-567.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, *95*, 371-384.

Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *American Economic Review*, *80*, 204–217.

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

Wickens, C. D. (1984). *Engineering psychology and human performance*. Columbus, Ohio: Merrill.

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,* 195-225.

Windschitl, P.D. & Wells, G.L. (1998). The alternative-outcomes effect. *Journal of Personality and Social Psychology, 75,* 1411-1423.

Wilson, T. D., Wheatley, T. P., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 78,* 821-836.

Yamagishi, K. (2002). Proximity, compatability, and noncomplementarity in subjective probability. *Organizational Behavior and Human Decision Processes, 87,* 136-155.

Table 1A. Violations of Ordinal Complementarity
(adapted from Fox & Levav, 2000)

Study	*Elicitation mode*	$\Pr(H >_i L)$	*n*	$\Pr(\overline{H} >_i \overline{L})$	*n*	$\Pr(H >_i L) + \Pr(\overline{H} >_i \overline{L})$	*z*
1. Duke Sports	More likely	.75	69	.44	66	1.19	2.37
2. Mutual Funds	More likely	.38	115	.73	113	1.11	1.79
3. Corporate Theft	More likely	.41	71	.79	73	1.20	2.65
4. Temperatures	More likely	.50	84	.63	81	1.13	1.70
	Less likely	.50	82	.62	86	1.12	1.58
5. Academy Awards	More likely	.76	25	.48	31	1.24	1.94
	Less likely	.80	30	.57	30	1.37	3.18

Table 1B. Violations of Procedure Invariance
(adapted from Fox & Levav, 2000)

Study	*Elicitation mode*	$\Pr(H >_i L)$	*n*	$\Pr(\overline{H} >_i \overline{L})$	*n*	$\Pr(H >_i L) + \Pr(\overline{H} >_i \overline{L})$	*z*
6. Law Case	More likely	.66	62				1.88
	Probability	.49	55				
7. Politics	More likely	.64	22	.76	21	1.40	2.89
	Probability	.36	21	.73	22	1.09	0.64
8. College Basketball	More likely	.59	79	.74	76	1.33	4.41
	Probability	.24	73	.64	77	0.88	-1.62

Note. Tables 1A and 1B are adapted from Table 1 in Fox & Levav (2000) and present a summary of the authors' findings. Table 1Apresents examples of violations of ordinal complementarity. Table 1B presents violations of procedure invariance. For both tables, the first column lists the number and topic of the study. The second column lists the elicitation mode. The third column lists the proportion of respondents rating the hypothesis for which a high amount of evidence can be recruited (H) above the hypothesis for which a low amount of evidence can be recruited (L). The fourth column lists the sample size on which that proportion is based. The fifth column lists the proportion of respondents rating the complement of the H hypothesis ($\overline{H}$) above the complement of the L hypothesis ($\overline{L}$). The sixth column lists the sample size on which that proportion is based. The seventh column reports the summation of respondents ordering Hand its complement above L and its complement. A sum that is significantly greater than unity indicates a bias toward H. The final column lists the z score of the difference between the proportions reported in the third column and one minus the proportion reported in the fifth column (for details see the original source).

Table 2. Restoring Symmetry.

Condition	*Elicitation mode*	$\Pr(H >_i L)$	*n*	$\Pr(\overline{H} >_i \overline{L})$	*n*	$\Pr(H >_i L) + \Pr(\overline{H} >_i \overline{L})$	*z*
Control	More likely	.72	25	.70	27	1.42	2.83
Complement Salient	More likely	.39	28	.52	27	.91	.015

Note. The first column lists the condition and the second column lists the elicitation mode. The third column lists the proportion of respondents rating the more familiar hypothesis (H) above the less familiar hypothesis (L). The fourth column lists the sample size on which that proportion is based. The fifth column lists the proportion of respondents rating the complement of the H hypothesis, $\overline{H}$*, above the complement of the L hypothesis (*$\overline{L}$*). The sixth column lists the sample size on which that proportion is based. The seventh column reports the summation of respondents ordering H and its complement above L and its complement. A sum that is significantly greater than unity indicates a bias toward H. The final column lists the z score of the difference between the proportions reported in the third column and one minus the proportion reported in the fifth column.*

Table 3. Mean weights and t-statistics of differences, computed for each participant using ordered logistic regression.[14]

Study	Relative Likelihood			Absolute Likelihood		
	Focal weight	*Complementary weight*	*t-value*	*Focal weight*	*Complementary weight*	*t-value*
ACC	14.64	11.19	4.79	17.84	17.68	0.25
NCAA	25.93	22.08	2.60	22.22	20.97	0.83

Note. Focal weight refers to mean logistic regression weight, averaging over participants, on the focal hypotheses; alternative mean weight to weight on the complementary hypothesis. T-values refer to t-statistics for the paired sample t-test of a difference in means.

Table 4. Mean weights computed from linear regression.

Study	Likelihood				Probability			
	Focal weight	*Alternative weight*	*t-value*	*Adj R^2*	*Focal weight*	*Alternative weight*	*t-value*	*Adj R^2*
ACC	6.52	5.12	>1000	.81	38.08	36.57	2.01	.90
NCAA	10.62	8.80	>1000	.65	60.27	58.37	1.50	.71

Note. Focal weight refers to mean linear regression weight, averaging over participants, on the focal hypotheses; alternative mean weight to weight on the complementary hypothesis. T-values refer to t-statistics for the paired sample t-test of a difference in means. Adj-R^2 refers to the adjusted R-square statistic for the designated regression.

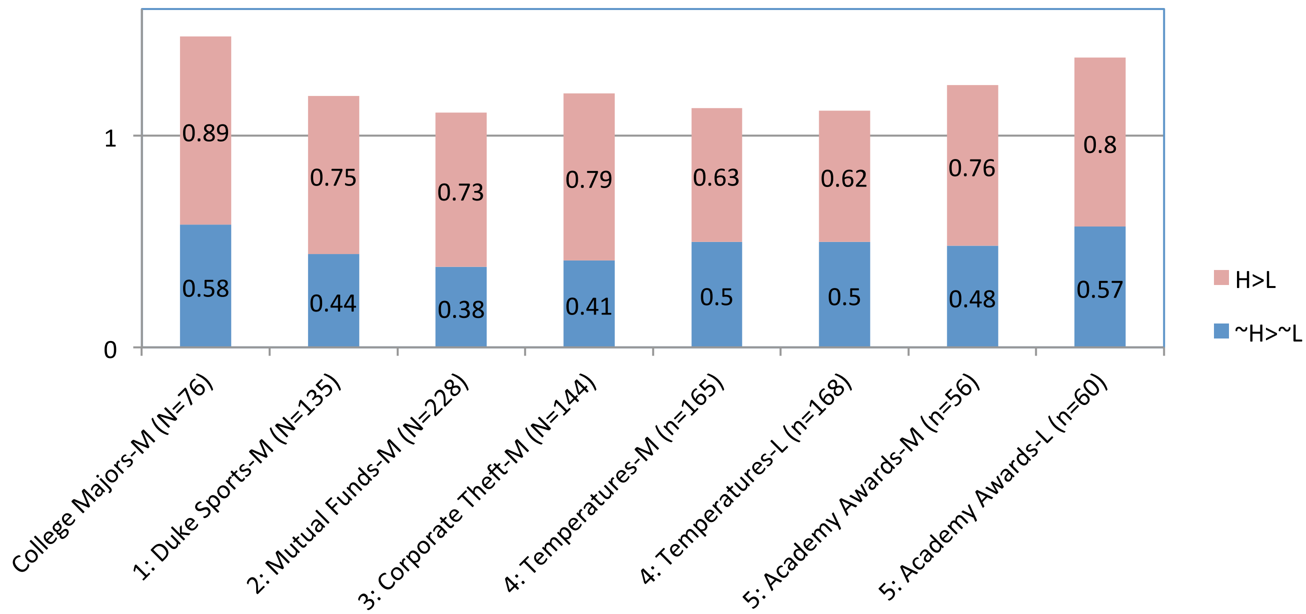
[14] We used ordered logistic regression to analyze, for the NCAA study, 670 ties out of 6336 cases for relative likelihood, 347 ties out 6336 cases for probability; for ACC study, 1756 out of 18432 cases for relative likelihood, 676 ties out of 18432 cases for probability.

Table 5. Joint Versus Separate Evaluation of Probability

<table>
<tr><th>Evaluation Mode</th><th>Jerry
(MCAT = 29, GPA = 3.81)</th><th>David
(MCAT = 38, GPA = 3.55)</th></tr>
<tr><td>Separate
(n = 29, 24)</td><td>52.7
(25.2)</td><td>51.8
(22.4)</td></tr>
<tr><td>Joint
(n = 27)</td><td>45.7</td><td>63.8</td></tr>
<tr><td></td><td colspan="2">(25.5)</td></tr>
</table>

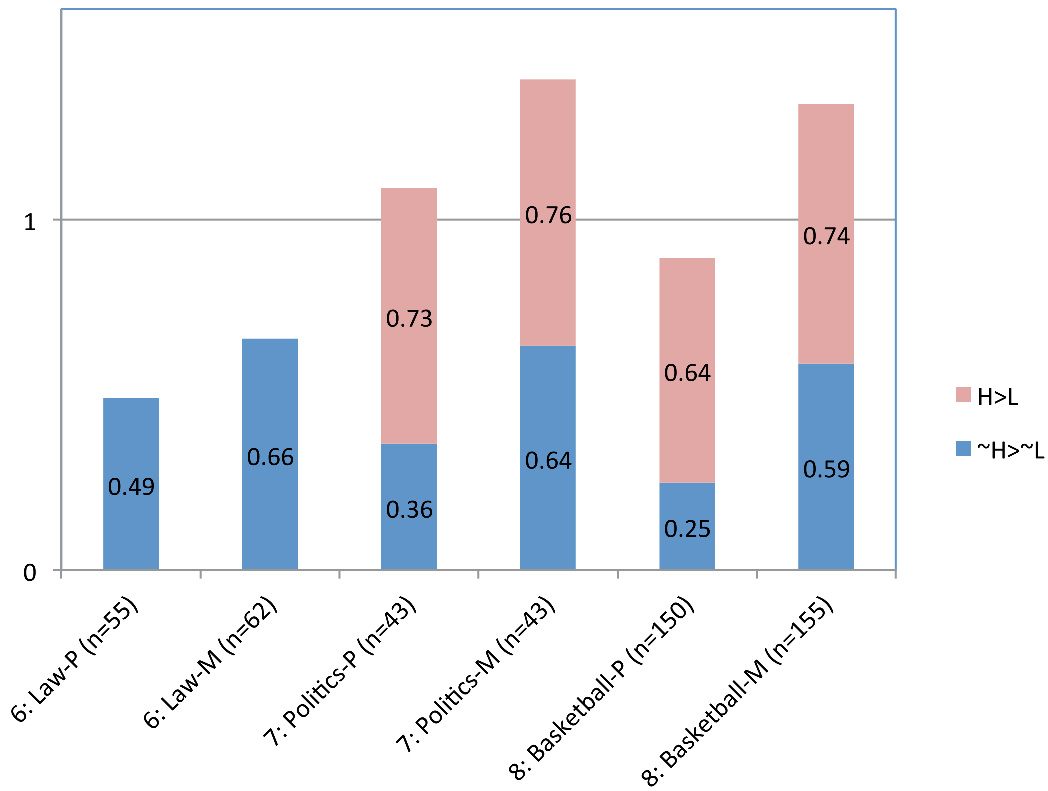
Note. The first column lists the evaluation mode and sample size. The second and third columns list mean judged probabilities (standard deviations in parentheses).

Figure 1A
Violations of Ordinal Complementarity



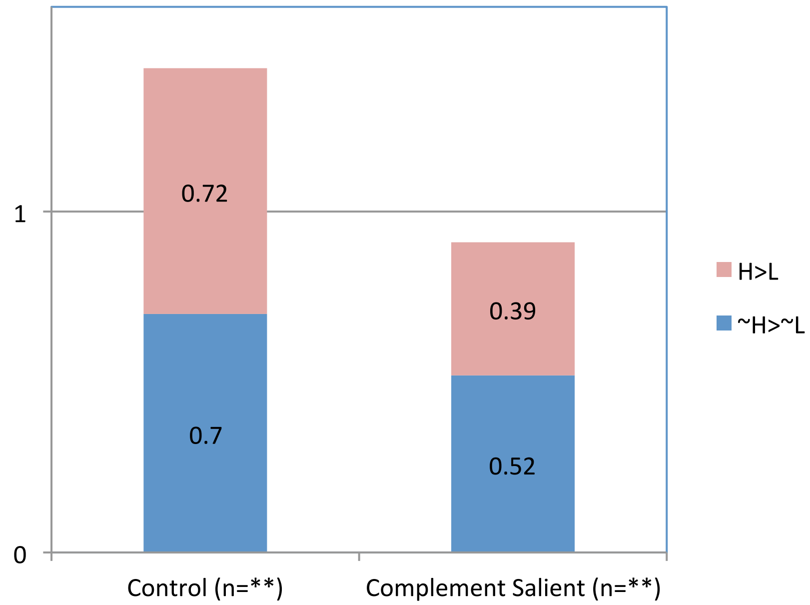
Data presented are proportions of participants in each condition indicating that the highly familiar hypothesis (H or ~H) is more likely (Mu) or the less familiar less likely (Lambda) to occur than the less familiar hypothesis (L or ~L, respectively). Numbers 1-5 indicate study numbers from Fox & Levav (2000). Note that in each case the sum of the bars exceeds 1, representing a violation.

Figure 1B
Violations of Procedure Invariance



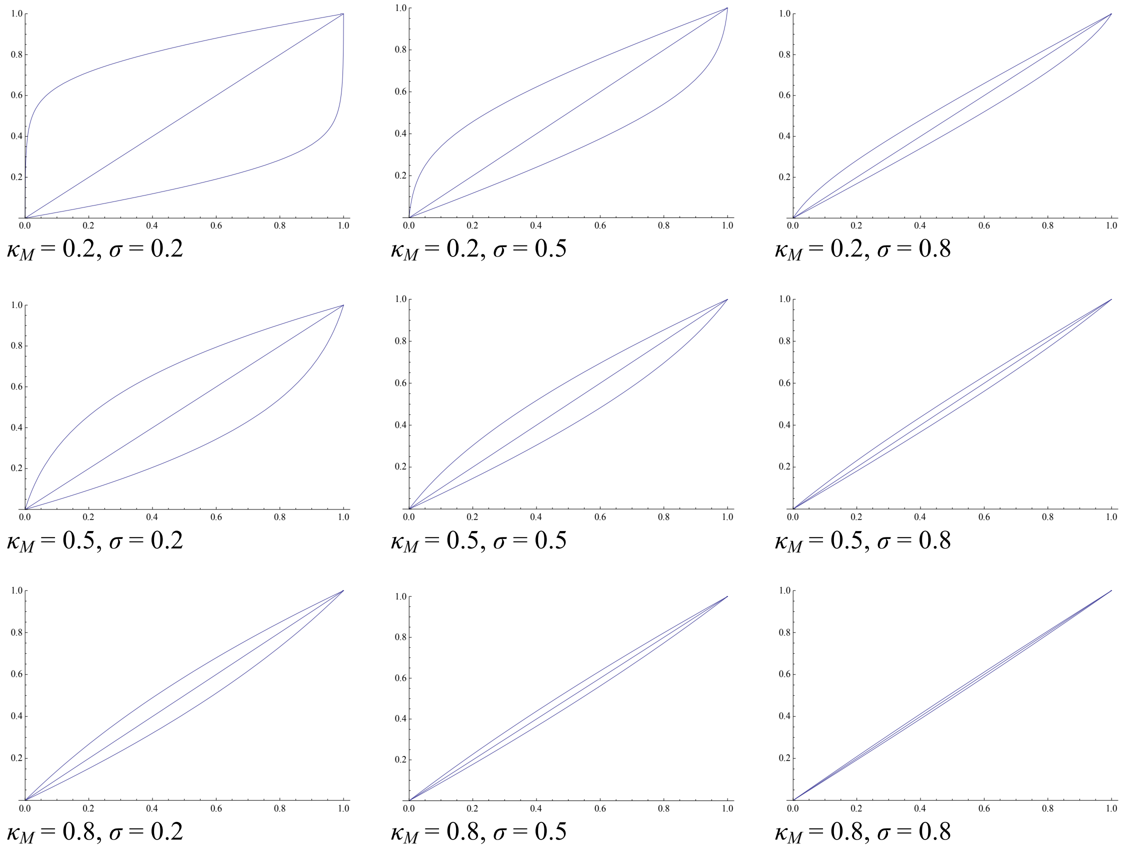
Data presented are proportions of participants in each condition indicating that the highly familiar hypothesis (H or ~H) is more likely (Mu) or has a higher probability (P) than the less familiar hypothesis (L or ~L, respectively). Numbers 6-7 indicate Study numbers from Fox & Levav (2000).

Figure 2
Restoring Symmetry



Proportion of participants saying one of a group of Best Picture nominees (high familiarity hypotheses) or Best Foreign Film nominees (low familiarity hypothesis) is more likely to win. These data are presented for the condition in which this question is the first item on the survey (Control) or follows a question about one's preferences for which film wins in each award category, thereby making the complementary hypothesis more salient (Complement Salient).

Figure 3
Belief Reversal Contours Implied by the CWS Model



The panels above illustrate the instances in which the CWS model predicts belief reversals. In each panel the *y*-axis represents *p,* the proportion of support of the high-familiarity domain that favors the focal hypothesis; the *x*-axis represents *q,* the proportion of support of the low-familiarity domain that favors the focal hypothesis. Thus, points above the identity are regions where $H \geq_P L$ (and $\overline{L} \geq_P \overline{H}$), and points below the identity are regions where $L \geq_P H$ (and $\overline{H} \geq_P \overline{L}$). Columns indicate different levels of σ, the ratio of total support for the less familiar to more familiar hypotheses. Rows indicate different levels of κ_M, the ratio of weight afforded the alternative to focal hypotheses. Contours are defined by inequalities (5a) and (5b). In each panel the lower lip is the region in which $L \geq_P H$ and $H \geq_m L$ and the upper lip is the region in which $\overline{L} \geq_P \overline{H}$ and $\overline{H} \geq_m \overline{L}$ (violating procedure invariance); both lips represent the region in which $H \geq_m L$ and $\overline{H} \geq_m \overline{L}$ (violating ordinal complementarity).