Brief article

When good evidence goes bad: The weak evidence effect in judgment and decision-making

Philip M. Fernbach *, Adam Darlow, Steven A. Sloman

Brown University, United States

A R T I C L E   I N F O

A B S T R A C T

An indispensable principle of rational thought is that positive evidence should increase belief. In this paper, we demonstrate that people routinely violate this principle when predicting an outcome from a weak cause. In Experiment 1 participants given weak positive evidence judged outcomes of public policy initiatives to be less likely than participants given no evidence, even though the evidence was separately judged to be supportive. Experiment 2 ruled out a pragmatic explanation of the result, that the weak evidence implies the absence of stronger evidence. In Experiment 3, weak positive evidence made people less likely to gamble on the outcome of the 2010 United States mid-term Congressional election. Experiments 4 and 5 replicated these findings with everyday causal scenarios. We argue that this "weak evidence effect" arises because people focus disproportionately on the mentioned weak cause and fail to think about alternative causes.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Deciding on a course of action often requires a prediction about the future. For instance, deciding whether or not to support a public policy initiative depends on the state of affairs likely to obtain if the policy is adopted. Will the economy improve if a stimulus bill is passed? Will a war-torn country eventually achieve political stability if troops are committed? Across a wide variety of cognitive tasks people tend to be myopic, basing judgments exclusively on whatever is in the immediate context (Dawes, 2001). Fernbach, Darlow, and Sloman (2011) explored the extent to which people display this tendency in predictive and diagnostic causal reasoning. The key finding that emerged across a variety of materials and manipulations

was an asymmetry in the extent to which people attend to alternative causes when making predictions versus diagnoses. While diagnostic judgments displayed exquisite sensitivity to the strength of alternative causes, predictive judgments did not vary with the strength or even with the presence versus absence of alternative causes (Fernbach, Darlow, & Sloman, 2010). In these studies, the stimuli tended to embody fairly strong causes, attenuating the potential error due to the neglect of alternatives. The open question that we address in this paper is whether the neglect of alternative causes in prediction can lead to substantial errors in judgment and whether those errors also impact decision-making.

To understand why neglect of alternative causes might prove insidious, consider a political debate about a troop commitment aimed at stabilizing the government in Afghanistan. Afghanistan's political environment is impacted by many factors some of which seem insufficient to achieve the goal of stability when considered individually. For instance The European Union's commitment of

* Corresponding author. Address: Department of Cognitive, Linguistic & Psychological Sciences, Brown University, Box 1821, Providence, RI 02912, United States. Tel.: +1 401 863 1101.

E-mail address: philip_fernbach@brown.edu (P.M. Fernbach).

9000 troops – while a net positive – might seem woefully inadequate when considered in isolation. When assessing the utility of the troop commitment, focusing on this one cause of stability and neglecting other relevant factors might cause voters' belief in the likelihood of stability to be lower than if they did not know about the troop commitment. In that case, the troop commitment has probably increased the actual likelihood of stability in Afghanistan, but paradoxically made voters less likely to support it.

The magnitude of potential error increases as the likelihood of the outcome increases and the causal strength of the focal cause decreases. The example illustrates the extreme case where a weak cause – one that raises the likelihood of the effect only slightly – actually reduces a person's confidence that the effect will occur. Demonstrating such cases requires three types of judgments: (i) The *conditional* likelihood of the outcome given a cause (e.g. How likely is a stable government in Afghanistan given that the EU pledged 9000 troops?), (ii) The *marginal* likelihood of the outcome – the likelihood of the outcome when no causes are mentioned (e.g. How likely is a stable government in Afghanistan?), and (iii) A *probability-raising* judgment to verify that the cause in question is indeed seen as raising the likelihood of the outcome (e.g. Does the EU pledging 9000 troops raise or lower the likelihood of a stable government in Afghanistan?). It is inconsistent to judge the conditional lower than the marginal but judge the cause as probability-raising. Nonetheless, we predicted that this pattern would emerge due to a disproportionate focus on the weak cause. We call it the *weak evidence effect*.

2. Experiment 1: public policy

We created stimuli based on four public policy themes in the public consciousness at the time of study: the economy, the climate, healthcare, and the war in Afghanistan. For each theme we collected judgments of the conditional probability of an effect given a weak cause, the marginal probability of the effect, and whether the cause is probability-raising. We predicted that participants would display the weak evidence effect and judge conditional likelihoods lower than marginal likelihoods while judging the causes to be probability-raising.

2.1. Methods

Conditional, marginal, and probability-raising questions were created for each theme. Each conditional probability question consisted of three sentences; the first stated some background information, the second the presence of a weak cause, and the third requested the likelihood of the outcome. The conditional questions for each theme are shown in Table 1. Marginal questions were identical except they excluded the second sentence. Two questionnaires were created, each containing two conditional items, two marginal items and two filler items, all from different themes. Instructions at the top of each questionnaire asked participants to judge each question on a 0 ('impossible') to 100 ('definite') scale.

Probability-raising questions were identical to the conditional questions except that the question read 'does that raise or lower the likelihood that. . .?' Participants judged the questions on a 7-point scale ranging (left to right) from 'it lowers it a lot,' to 'it raises it a lot'. The four probability-raising questions were all included on a third questionnaire with two filler items, for a total of six questions.

Fifty-one members of the Brown University community were approached on campus and participated voluntarily. They were assigned at random to one of the three questionnaires and completed it in 5–10 min.

2.2. Results and discussion

The probability-raising questions were analyzed by converting the responses to numeric values from 1 to 7, with four corresponding to the scale mid-point, 'it neither raises nor lowers the likelihood.' As intended, the causes were seen as slightly probability-raising. The judgments were significantly higher than the mid-point, Mean = 4.9, $t(14) = 6.4$, $p < 0.001$, Cohen's $d = 1.6$, and the means of all themes were above the mid-point. Of the 60 judgments across all the themes only two were below the mid-point.

Probability judgment means and standard errors by theme are shown in Fig. 1. As predicted, conditional judgments (Mean = 33.7) were lower than marginal judgments (Mean = 48.7) when we collapsed over themes and compared participant means, $t(35) = 5.1$, $p < 0.001$, Cohen's

Table 1
Stimuli from Experiment 1. The weak cause is italicized.

| Theme	Conditional
Economy	Approximately 10% of the US population is currently using food stamps. *The Congress has recently approved a 15-cent increase in the federal minimum wage (from $7.25 to $7.40).* How likely is it that the percentage of people using food stamps will be less than 9% by the beginning of 2011?
Climate	Widespread use of hybrid and electric cars could reduce worldwide carbon emissions. *One bill that has passed the Senate provides a $250 tax credit for purchasing a hybrid or electric car.* How likely is it that at least one fifth of the US car fleet will be hybrid or electric in 2025?
Healthcare	The infant mortality rate in the United States is currently 6.3 deaths per 1000 live births. *The health care reform bill that is likely to pass into legislation includes funding for an education program to teach prospective mothers about prenatal nutrition.* How likely is it that the infant mortality rate in the United States will be below 5.5 deaths per 1000 live births by 2020?
Afghanistan	The democratic government of Afghanistan is embroiled in a protracted conflict with Taliban insurgents. *The European Union recently pledged 9000 troops to provide added security in population centers.* How likely is it that Afghanistan will have a stable government in 5 years?
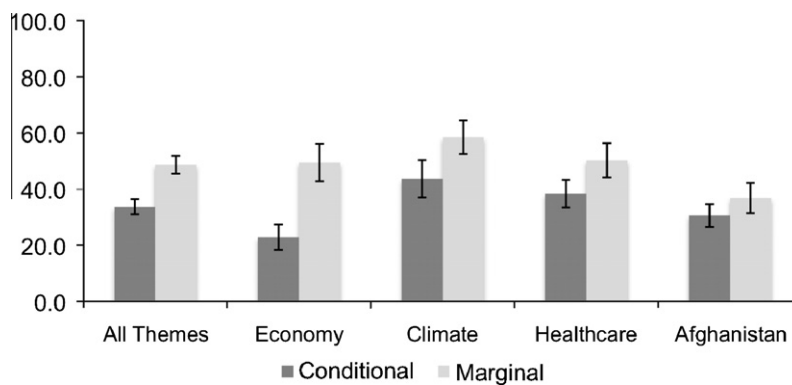
Fig. 1. Means and standard errors by theme for Experiment 1.

$d = 1.7$. The difference was also significant when collapsing over participants, $t(3) = 3.5$, $p < 0.05$, Cohen's $d = 4.0$.

3. Experiment 2: a pragmatic alternative

Grice's (1975) maxim of relevance suggests two alternative accounts of Experiment 1 based on possible pragmatic implicatures in the questions. The first possibility is that people interpreted the mention of a target cause as a request to ignore alternative causes when judging the conditional. In other words, they may have interpreted the conditional probability question as requesting a judgment of causal power (Cheng, 1997). We consider this possibility in Experiment 3.

A second alternative explanation is that the mention of the weak cause suggests the absence of stronger causes that participants might otherwise think of. For instance, mentioning the commitment of 9000 EU troops might suggest that a larger troop commitment will not be forthcoming, whereas, in the marginal condition, participants might assume a larger commitment. On this account, participants are not neglecting alternative causes, but rather are interpreting the evidence in the conditional as negative relative to their prior expectation. This account also requires that participants consider their prior expectation irrelevant to the probability-raising question. To test this possibility we asked people to first judge the marginal question for each item from Experiment 1 and then introduced the weak evidence and asked for a second (conditional) judgment of the likelihood of the outcome. If the pragmatic account is correct and participants interpret the evidence in the conditional as negative relative to their expectation then we would expect them to revise their judgment down and show the same pattern of judgments as Experiment 1. If the weak evidence effect is due to the neglect of alternative causes, we would expect them to judge the conditional equal to or higher than the marginal because judging the marginal would make them consider alternatives; the weak cause mentioned in the conditional should provide additional positive evidence.

3.1. Methods

Nineteen participants were recruited via Amazon Mechanical Turk (Paolacci, Chandler, & Ipeirotis, 2010)

and participated online for a small monetary reward. Each of the six themes from Experiment 1 (4 test items, two filler items) was presented on one screen. For each item participants first judged the marginal and then the conditional on a 0–100 scale. Here is an example:

> "The democratic government of Afghanistan is embroiled in a protracted conflict with Taliban insurgents. How likely is it that Afghanistan will have a stable government in 5 years?
> Imagine you found out that the European Union recently pledged 9000 troops to provide added security in population centers. How likely is it that Afghanistan will have a stable government in 5 years?"

The only change in the stimuli from Experiment 1 was in the "Economy" theme. Due to time elapsed since Experiment 1 was conducted, we asked people to judge the likelihood that the percentage of people using food stamps will be less than 9% by the beginning of 2012 rather than 2011.

3.2. Results and discussion

Results by theme are shown in Fig. 2. As predicted by the neglect hypothesis, conditional judgments were significantly higher than marginal judgments, $t(18) = 2.8$, $p = 0.01$. Means of marginal questions were almost identical to Experiment 1 except that the mean for the economy theme was lower in Experiment 2. We suspect this was due to relatively poor expectations about the economy as compared to when Experiment 1 was conducted. One participant judged each of the four conditional questions lower than its analogous marginal question. Otherwise, only one of the remaining 72 conditional judgments was lower than its analogous marginal judgment.

In summary, after responding to the marginal question, participants either did not change or raised their judgment of the conditional. This is consistent with an interpretation of the evidence as weakly positive and supports the idea that the weak evidence effect is due to the neglect of alternative causes. When evidence brought to mind when answering the marginal question was available for conditional judgment, judgments were higher. The results are inconsistent with the pragmatic explanation that weak evidence is interpreted as negative with respect to a default expectation.
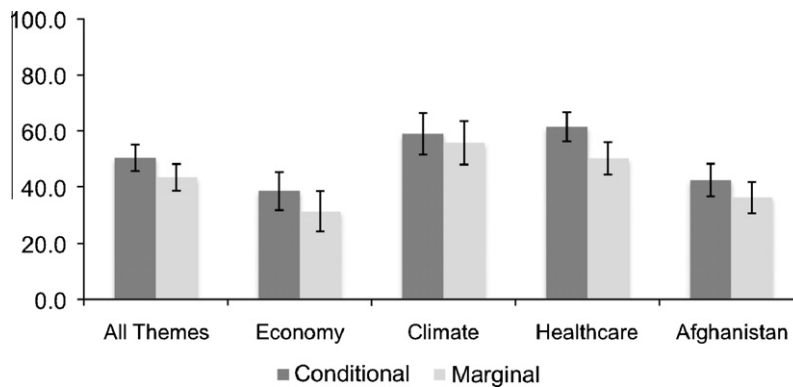
Fig. 2. Means and standard errors by theme for Experiment 2.

4. Experiment 3: decision-making

The goal of this experiment was twofold: first, we wanted to establish the weak evidence effect in a decision-making paradigm where participants choose a gamble with a real payoff whose value depends on whether an outcome occurs. Approximately one month before the 2010 United States mid-term Congressional election we asked people to gamble on whether the Republicans would retake a majority in the House of Representatives. Participants were given either weak evidence (conditional condition) or no evidence (marginal condition) about the election and chose whether they wanted a high-stakes gamble on Republicans retaking the House or a low-stakes gamble on a fixed amount of money. In the conditional condition we told people about a real event that had recently transpired, a newspaper endorsement of a Republican candidate in a hotly contested district. We predicted that a smaller proportion would choose to gamble on the Republicans taking the House when told about this weak evidence.

The second purpose of Experiment 3 was to test the remaining pragmatic account of the weak evidence effect, that people are interpreting the conditional probability question as a request for causal power. Establishing the effect in a decision-making paradigm would rule out such an explanation because participants can only reasonably gamble on the likelihood of the outcome, regardless of what cause brought it about.

4.1. Methods

One-hundred and eighteen United States residents were recruited via Amazon Mechanical Turk and participated online for a small monetary reward. Twenty participants were assigned to judge whether the evidence was probability-raising using the same scale as Experiment 1. The remaining 98 participants were assigned to the decision-making portion of the experiment. Upon logging onto the experiment, they read the following instructions:

> In this experiment you will be given some information and then be asked to choose a gamble. A subset of participants will be chosen to win a prize based on the gamble they choose. This is not a hypothetical study.

We will actually test the gamble and pay based on what happens.

Participants were then randomly assigned to one of the conditions. In the conditional condition they read the following:

> Mid-term elections for the House of Representatives are coming up in November. Recently, Ryan Frazier, a Republican candidate in Colorado's hotly contested 7th district House race won the endorsement of the Denver Post, Colorado's largest newspaper.

They were then asked to judge the likelihood that the "Republicans will win control of the House of Representatives" on a 0–100 scale. After making the likelihood judgment, they were asked to choose whether they wanted to gamble on the Republicans winning the House as below:
Please Choose One Option:

1. You get 10 dollars no matter what happens.
2. You get 30 dollars if the Republicans win control of the House of Representatives in mid-term elections.

The order of response options was randomized. The marginal condition was identical except that the sentence about the newspaper endorsement was omitted.

After choosing a gamble, participants proceeded to another page where they identified their political affiliation from the following options: "Democrat," "Republican," "Independent," "Tea Party" and "Other." After the mid-term elections (won by the Republicans) we paid out prizes to a random subset of participants based on the option they chose.

4.2. Results and discussion

We tested the main prediction of the Experiment by comparing the proportion of people in each condition choosing to gamble on the Republicans taking the House (Fig. 3). As predicted, a smaller proportion of participants in the conditional condition chose to gamble (25.5%) than in the marginal condition, 51.2%, $z = 2.6$, $p < 0.01$. Likelihood judgments also displayed the weak evidence effect, with participants in the conditional condition ($M = 56.3$) judging the likelihood of the Republicans winning the
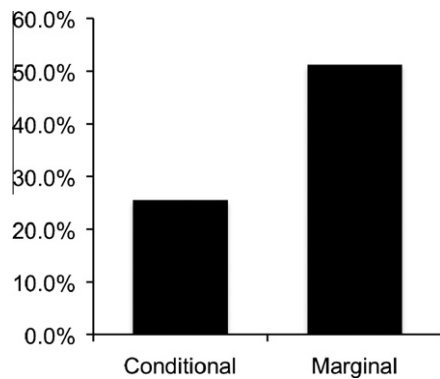
Fig. 3. Percentage of participants in each condition of Experiment 3 choosing to Wager on the Republicans winning the House of Representatives.

House lower than those in the marginal condition, $M = 64.8$, $t(96) = 2.1$, $p < 0.05$. To assess the effect of political affiliation on likelihood judgments we performed further analysis (Appendix A). Finally, as in Experiment 1, the evidence was judged to be slightly probability-raising; all of the probability-raising judgments were at or above the scale-mid-point and the mean was significantly greater than the scale mid-point, $M = 4.9$, $t(19) = 4.5$, $p < 0.001$.

5. Experiment 4: everyday causation

The goal of Experiment 4 was to replicate the weak evidence effect with a larger number of items not drawn from the public policy arena. We created materials based on 12 themes inspired by everyday events. We also wanted to assess the relation between conditional and marginal judgments and the causal power of the cause (the probability of the outcome given the weak cause alone). If participants completely neglect alternatives, causal power judgments should be identical to conditional judgments. If they sometimes consider alternatives, but not sufficiently, causal power judgments should be lower than conditional judgments. We therefore collected causal power judgments in addition to conditional, marginal, and probability-raising judgments. The four questions for one of the themes are shown in Table 2 (see Appendix B for the full set of stimuli).

5.1. Methods

The conditional, marginal and causal power questions were divided into three questionnaires such that each questionnaire had four of each question type. Six filler items were added for a total of 18 questions per questionnaire. The 12 probability-raising questions were all included in a fourth questionnaire, also with six filler items. The dependent measures and instructions were identical to Experiment 1.

Seventy-three Brown University undergraduates participated for psychology course credit. They were assigned at random to one of the four questionnaires and completed it in approximately 15 min.

5.2. Results and discussion

The results are shown in Table 3. As in Experiment 1, conditional judgments (Mean = 46.4) were lower than marginal judgments (Mean = 52.2) when collapsed over themes, $t(53) = 2.2$, $p < 0.05$, Cohen's $d = 0.6$, and when collapsed over participants, $t(11) = 2.4$, $p < 0.05$, Cohen's $d = 1.4$.

Again, the causes were judged probability-raising. Probability-raising judgments were significantly higher than the scale mid-point of 4, Mean = 5.0, $t(18) = 10.4$, $p < 0.001$, Cohen's $d = 2.4$, and the means of all 12 themes were above the mid-point. Of the 228 judgments only 11 were below the mid-point.

Conditional judgments were significantly higher than causal power judgments (Mean = 38.1) when collapsed over themes, $t(53) = 2.8$, $p < 0.01$, Cohen's $d = 0.8$, and when collapsed over participants, $t(11) = 2.9$, $p < 0.05$, Cohen's $d = 1.7$, suggesting that participants did not completely neglect alternatives. Inductive inferences rely on retrieval from semantic memory (Dougherty, Gettys, & Ogden, 1999) and this retrieval is sometimes driven by the strength of the relation between the cue and associated memory structures (Quinn & Markovitz, 1998). We hypothesize therefore that alternative causes may sometimes come to mind when they are highly available leading

Table 3
Means and standard errors by theme for Experiment 4.

	Conditional	Marginal	Causal power
Likelihood Judgment	46.4 (2.1)	52.2 (2.6)	38.1 (2.3)

Table 2
The four questions for one of the themes in Experiment 4.

Question type	Wording
Conditional	A man buys a half-gallon of milk on Monday. The power goes out for 30 min on Tuesday. How likely is it the milk is spoiled a week from Wednesday?
Marginal	A man buys a half-gallon of milk on Monday. How likely is it the milk is spoiled a week from Wednesday?
Probability-raising	A man buys a half-gallon of milk on Monday. The power goes out for 30 min on Tuesday. Does that raise or lower the likelihood that the milk is spoiled a week from Wednesday?
Causal power	A man buys a half-gallon of milk on Monday. The power goes out for 30 min on Tuesday. How likely is it that the power going out for 30 min on Tuesday causes the milk to be spoiled a week from Wednesday?

to the observed pattern of judgments. This interpretation rests on the assumption that participants understand "how likely is it that X causes Y" to be a request for causal power, the likelihood that the focal cause successfully brings about the effect.

6. Experiment 5: decision-making about everyday causation

Experiment 5 was intended to replicate the decision-making effect of Experiment 3 with a theme drawn from everyday causation. We chose the "spoiled milk" example because it is amenable to testing in order to determine a payout based on the gamble.

6.1. Methods

One-hundred and fifty-nine participants were recruited via Internet message boards and participated online. Upon logging onto the experiment, they read the same instructions as Experiment 2. Participants were then randomly assigned to one of the conditions. In the conditional condition they read the following:

On Monday we will buy a half-gallon of milk. On Tuesday the power will go out for 30 min.
Will the milk be spoiled a week from that Wednesday?

Please Choose One Option:

1. You get 10 dollars no matter what happens.
2. You get 30 dollars if the milk is spoiled, otherwise nothing.

The order of response options was randomized. The marginal condition was identical except the sentence about the power outage was omitted. After data collection, we tested the gamble described in each condition by purchasing a half-gallon of milk on a Monday and checking its freshness the following Wednesday, either cutting power for 30 min on Tuesday or not depending on condition and not interfering otherwise with the milk. Both experiments yielded unspoiled milk. We then paid out prizes to a random subset of participants based on the options they chose.

6.2. Results and discussion

As predicted, a smaller proportion of participants in the conditional condition (21.1%) chose to gamble on spoilage than in the marginal condition (36.4%), $z = 2.1$, $p < 0.05$.

7. General discussion

Two experiments identified a weak evidence effect in judgment. When participants predicted an outcome conditioned on a weak cause for that outcome, they gave lower judgments than when predicting the outcome without any mention of the cause, despite the fact that the causes were separately judged as probability-raising. Two additional experiments established the effect in a decision-making paradigm; participants were less likely to gamble on an outcome when given weak positive evidence for the outcome.

Evidence against pragmatic accounts of the phenomenon come from two sources: First, when judging the marginal prior to judging the conditional, participants raise or do not change their judgments suggesting they do not interpret the weak evidence as negative relative to an expectation (Experiment 2). Second, because participants display the effect when gambling on an outcome they cannot be conflating the conditional probability judgment with causal power (Experiments 3 and 5). This conclusion also obtains some support from the fact that causal power judgments in Experiment 4 were lower than marginal or conditional judgments.

7.1. Mechanism and related phenomena

We attribute the weak evidence effect to the process by which people use their causal knowledge to predict effects from their causes. People do so by retrieving relevant causal variables and embedding them in a mental model that supports forward inference via simulation (Kahneman & Tversky, 1982). No judge can be expected to consider every relevant cause. Instead, people tend to restrict attention to a single mechanism (Fernbach et al., 2011). When reasoning about a conditional probability, people focus on the conditioned-on cause leading to low judgments. When judging a marginal probability however, people begin at a different point, by retrieving more available causes, leading to higher judgments.

A similar logic explains why unpacking a hypothesis into atypical constituents decreases judgment. Unpacking the description of an event, like "death from disease", into constituents, like "death from heart disease or some other disease" usually increases the judged probability of the event (Tversky & Koehler, 1994). However, unpacking with atypical constituents (e.g. "death from pneumonia or some other disease") can reduce judgment suggesting that people neglect constituents that are not mentioned (Sloman, Rottenstreich, Wisniewski, Hadjichristidis, & Fox, 2004). This is analogous to the present case in which mentioning a weak cause leads to neglect of alternative causes. Rottenstreich and Tversky (1997) showed that a causal partition leads to a greater unpacking effect than a temporal partition, consistent with the claim that causes crowd one another out.

Causal conjunction fallacies are, in a sense, a reverse case. The conjunction fallacy occurs when a conjunction of two events is judged more probable than one of the events alone. Causal conjunction fallacies are a specific case where the conjunction of a cause and its effect is judged more likely than the effect alone. Kahneman and Tversky (1983) give the following example where a is judged more likely than b:

(a) An earthquake in California sometime in 1983, causing a flood in which more than 1000 people drown.
(b) A massive flood somewhere in North America in 1983, in which more than 1000 people drown.

Errors like this occur when the marginal outcome probability is low and the causal power is fairly high, precisely the converse of the conditions that facilitate the weak evidence effect. One explanation for this is that people focus too much on the mechanism connecting the cause and the effect when assessing the conjunction (Ahn & Bailenson, 1996). Focusing on the strength of the causal relation leads people to neglect the base rate and judge the conjunction fairly high. In the marginal case however, the absence of readily available causes leads to low judgments.

What all these cases have in common is the requirement to predict an effect from causal knowledge. Similar phenomena sometimes emerge in domains that do not have this structure but they may have different explanations. McKenzie, Lee, and Chen (2002) have shown that when reasoning in the context of an argument with opposing sides, weak evidence of innocence will sometimes increase belief in guilt. They argue that this phenomenon emerges because sides in a dispute are motivated to provide the strongest possible case; a weak case implies an inability to amass strong evidence. Evaluating evidence relative to the strength of an expectation is often called for, but our results cannot be explained in this way. The judgments and decisions we asked people to make were not presented in the context of an argument that supports expectations about the strength of evidence. Moreover, as discussed above, Experiment 2 speaks against the related pragmatic possibility that the effect is driven by the mention of a weak cause implying the absence of stronger causes that are present under ordinary circumstances.

"Reverse belief updating" sometimes emerges even in the absence of an adversarial context. For instance, Lopes (1985) reports that in a Bayesian updating task, observing weak evidence favoring a hypothesis after having just seen strong evidence leads some people to (incorrectly) adjust their judgment downward (for evidence concerning a related phenomenon, the 'dilution effect,' see Nisbett, Zukier, & Lemley, 1981; Shanteau, 1975). This could reflect a general tendency to evaluate evidence with respect to comparisons in the immediate environment rather than with respect to its absolute value. The analogy to the current findings is tenuous though because the weak evidence effect emerges from considering a single piece of evidence and not from integrating over multiple samples.

7.2. Implications

Anyone in the business of convincing others (e.g. marketers, politicians, scientists) should take heed that positive evidence will not always benefit persuasion. In line with this idea, Simonson, Carmon and O'Curry (1994) have shown that adding a feature to a product can reduce choice probability even if the feature is not seen as reducing the value of the product. This may be because undue focus on the feature directs attention away from other beneficial product attributes. Voting behavior is another such area that Experiment 3 speaks to directly. Decisions to vote are in part a function of how likely people think their vote is to matter (Quattrone & Tversky, 1984). If weak positive evidence influences people's predictions about election outcomes in the wrong direction, it could lead to unwarranted decisions to abstain. This suggests that weak positive evidence used as a tool to induce people to vote could actually be deleterious.

Conversely, awareness of the weak evidence effect may help people avoid being persuaded when it is used as a rhetorical tool. For instance, opponents of a public policy initiative might attempt to diminish support for the initiative by focusing attention on particular aspects of it. A 15-cent increase in the minimum wage may be a beneficial part of a larger economic stimulus bill, but focusing attention on that part of the plan makes it seem unlikely to work. This may be one reason that people react negatively to complex, sweeping policy initiatives while expressing support for each of the pieces individually as seen in recent polls (CNN, 2010).

8. Conclusion

The law of total probability implies that if event A raises the probability of event B, the probability of event B must be higher when A is present than when it is unknown. The weak evidence effect is a violation of this basic norm of probability theory. This violation arises because people focus on what they perceive in their immediate environment and neglect other information, a tendency that is ubiquitous in human cognition. It arises when people reason (Evans, Over, & Handley, 2003), test hypotheses (Doherty, Chadwick, Garavn, Barr and Mynatt, 1996), understand language (Keysar, Lin, & Barr, 2003), troubleshoot (Fischhoff, Slovic, & Lichtenstein, 1978), and make categorical judgments (Ross & Murphy, 1996). Such focus may often be a reasonable approximation strategy, but it sometimes leads to error.

Acknowledgments

Appendix A. Analysis of Likelihood judgments from Experiment 2 as a Function of Political Affiliation

Only three participants identified as affiliating with the Tea Party so we added them to the "Other" category for the subsequent analysis. Likelihood judgment means by affiliation are shown in the table below. The likelihood judgments were entered into a two-way ANOVA with condition (conditional vs. marginal) and political affiliation as between-participant factors. Demonstrating the weak evidence effect, there was a main effect of condition, $F(1, 97) = 4.4$, $p < 0.05$, partial $\eta^2 = 0.05$; conditional judgments were lower than marginal judgments. There was also a main effect of political affiliation, $F(1, 97) = 5.6$, $p < 0.001$, partial $\eta^2 = 0.16$; Republicans were most confident of a House takeover by Republicans, followed by Democrats, Others and Independents. There was no interaction; all groups evidenced the weak evidence pattern.

Judged likelihood of a republican takeover of the house of representative by political affiliation and condition.

Affiliation	Percent of sample (%)	Conditional	Marginal
Democrats	34.7	54.4	65.2
Republicans	16.3	74.4	80.7
Independent	30.6	48.2	59.6
Other	18.4	55.6	61.3
All groups	100.0	56.3	64.8

Appendix B

Conditional questions from Experiment 4. Alternative question forms (marginal, casual power and probability-raising) were generated as in the example in Table 2 in the main text.

Theme	Conditional
Cell phone	A woman is a 35 year old whose parents live in a different state. She loses her cell phone on April 1st. How likely is it she does not talk to her parents in April?
Beer company	A beer company owns a leading light beer. The company increases the advertising budget for its light beer by 3%. How likely is it the beer gains market share in the next year?
Vineyard	A California vineyard specializes in French style wine. The vineyard imports topsoil from France. How likely is it that the wine scores well in a blind taste test by French critics?
Probiotic diet	A man is a 20-year-old university student. He is on a probiotic diet. How likely is it he goes a year without the flu?

Appendix B (continued)

Theme	Conditional
House flipper	A house flipper is looking to sell a property he acquired one year ago. He repaints all of the bedrooms in the house. How likely is it he realizes at least a 2% profit when he sells?
College	A young man is applying to colleges and trying to improve his application. He volunteers for the big brother program. How likely is it he gets into a top 100 college?
Jacket	A young man is a healthy high school student. He goes out during a heavy rain without a jacket. How likely is it he gets a cold sometime this winter?
Gasoline	A woman has 2003 Honda. She uses the lowest grade of gasoline. How likely is it the car has mechanical problems in the next year?
Milk	A man buys a half-gallon of milk on Monday. The power goes out for 30 min on Tuesday. How likely is it the milk is spoiled a week from Wednesday?
Smoking	A 30-year-old woman wants to quit smoking. She goes to hypnosis sessions. How likely is it she no longer smokes in 1 year?
Baseball	A baseball player hit 20 homeruns in the 2009 season. After the season he used a computer program twice a week to train his visual acuity. How likely is it he hits more than 20 homeruns in the 2010 season?
Tourist	A tourist is taking a picture of the statue of liberty from the deck of the ferry. There is a breeze at the moment he takes the picture. How likely is it the photo comes out blurry?

References

Ahn, W., & Bailenson, J. (1996). Causal attribution as a search for underlying mechanisms: An explanation of the conjunction fallacy and the discounting principle. *Cognitive Psychology, 31*, 82–123.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367–405.

CNN, 2010. Opinion research poll, February 12–15, 2010 [Retrieved from CNN.com on October, 13, 2010].

Dawes, R. M. (2001). *Everyday irrationality: How pseudoscientists, lunatics, and the rest of us fail to think rationally.* Boulder, CO: Westview Press.

Doherty, M. E., Chadwick, R., Garavan, H., Barr, D., & Mynatt, C. R. (1996). On people's understanding of the diagnostic implications of probabilistic data. *Memory & Cognition, 24*(5), 644–654.

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). Minerva-DM: A memory process model for judgments of likelihood. *Psychological Review, 106*(1), 180–209.

Evans, J. St. B. T., Over, D. E., & Handley, S. J. (2003). A theory of hypothetical thinking. In D. Hardman & L. Maachi (Eds.), *The psychology of reasoning and decision making.* Chichester: Wiley.

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2010). Neglect of alternative causes in predictive but not diagnostic reasoning. *Psychological Science, 21*(3), 329–336.

Fernbach, P. M., Darlow, A. & Sloman, S. A. (2011). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology*, in press.

Fischhoff, B., Slovic, P., & Lichtenstein (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human, Perception, and Performance, 4*, 330–344.

Grice, P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.). *Syntax and semantics, Speech acts* (Vol. 3). New York: Academic Press.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1983). Extensional versus intuitive reasoning: The conjunction Fallacy in probability judgment. *Psychological Review, 90*(4), 293–315.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25–41.

Lopes, L. L. (1985). Averaging rules and adjustment processes in Bayesian inference. *Bulletin of the Psychonomic Society, 23*(6), 509–512.

McKenzie, C. R. M., Lee, S., M. & Chen, K. K. (2002). When negative evidence increases confidence: Change in belief after hearing two sides of a dispute. *Journal of Behavioral Decision Making, 15*, 1–18.

Nisbett, R. E., Zukier, H., & Lemley, R. (1981). The dilution effect: Nondiagnostic information. *Cognitive Psychology, 13*, 248–277.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon mechanical Turk. *Judgment and Decision-Making, 5*(5).

Quattrone, G. A., & Tversky, A. (1984). Causal versus diagnostic contingencies: On self-deception and the voter's illusion. *Journal of Personality and Social Psychology, 46*, 237–248.

Quinn, S., & Markovitz, H. (1998). Conditional reasoning, causality and the structure of semantic memory: Strength of association as a predictive factor for content effects. *Cognition, 68*, B93–B101.

Ross, B. H., & Murphy, G. L. (1996). Category-based predictions: Influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 736–753.

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

Shanteau, J. (1975). Averaging versus multiplying combination rules of inference judgment. *Acta Psychologica, 39*, 83–89.

Simonson, I., Carmon, Z., & O'Curry, S. (1994). Experimental evidence on the negative effect of product features and sales promotions on brand choice. *Marketing Science, 13*, 23–40.

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

Tversky, A., & Koehler (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101*(4), 547–567.