

When Negative Evidence Increases Confidence: Change in Belief After Hearing Two Sides of a Dispute

CRAIG R. M. MCKENZIE*, SUSANNA M. LEE, and KAREN K. CHEN
University of California, San Diego, USA

ABSTRACT

Four experiments examined change in confidence after hearing two sides of a dispute. The results showed that a case independently judged to weakly support one side often increased confidence that the opposing side was correct. Furthermore, the stronger the first case, the more likely a subsequent weak case had a reverse impact. Traditional belief-updating models, which tend to focus on change in belief after individual pieces of evidence rather than entire cases, cannot account for these results, and a model that can account for them is introduced. In the new model, case strength is evaluated with respect to a relatively demanding (and malleable) reference point. A weak case can fall below this demanding reference point, resulting in a reverse impact on confidence. Cases must exceed relatively high strength thresholds in order to have their intended impact because they are expected to be biased summaries of evidence. When it is clear that a weak case is unbiased, it affects confidence in the intended direction. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS belief updating; opinion revision; confidence; uncertainty; dispute resolution

People often find themselves first hearing one side of a dispute, then later hearing the other side. Perhaps the most obvious example is found in the courtroom, where jurors first hear the plaintiff's case, then the defendant's case. More informally, a former colleague might tell you her account of why she was fired, and later you hear your boss's account. Or you might read allegations leveled against a political figure, who then responds to them publicly. In each example, one presumably updates confidence that one or the other version of the event in question is most accurate after hearing each side. A juror might feel strongly that a defendant is guilty after hearing the prosecution, but feel less so after hearing the defense.

The focus of this article is on change in confidence after each of two sides of a dispute is presented. One might think that such changes depend on the strength of the presented case in a straightforward way. A

*Correspondence to: Craig R. M. McKenzie, Department of Psychology, University of California, San Diego, La Jolla, CA 92093-0109.
E-mail: cmckenzie@ucsd.edu

Contract/grant sponsor: National Science Foundation.
Contract/grant numbers: SBR-9515030, SES-0079615.

neutral case might have no effect on confidence, a weak case might pull confidence weakly in the direction implied by the case, and so on. In other words, change in confidence might depend on a comparison between a case's strength and a neutral zero point. However, we present data indicating that the reference point against which a case's strength is compared is more demanding than neutrality. We refer to this more demanding reference point as a case's minimum acceptable strength. A weak case can fall below this level, resulting in a change in confidence opposite of that intended. We show further that the minimum acceptable strength of the second case is influenced by the strength of the first case. Specifically, the stronger the first case, the higher the minimum acceptable strength of the second case. Raising the level of minimum acceptable strength makes weak cases more likely to have a reverse impact and leads strong cases to have less impact. We also argue that the reference point for case strength is more demanding than neutrality because cases are usually expected to be biased samples of evidence.

The first part of the article discusses two classic belief-updating models, adding and averaging, highlighting their different reference points for evaluating new evidence. The former tends to use a neutral zero point as a reference, and the latter uses confidence prior to hearing the new evidence. These models have tended to focus on change in confidence following individual pieces of evidence. We subsequently introduce a new reference point, minimum acceptable strength, which results in a model that we argue is better suited for describing change in confidence when presented with two sides of a dispute. Results from four experiments are then reported that confirm several novel predictions and provide evidence against alternative accounts of our data. Finally, we discuss the implications of our findings and suggest future research.

TRADITIONAL BELIEF-UPDATING MODELS: ADDING AND AVERAGING

Hogarth and Einhorn (1992) proposed an anchor-and-adjust model for updating opinion in a proposition or hypothesis following presentation of evidence. Our current concern is with updating confidence that a particular side of a dispute is the true one after hearing each of two sides involved in the dispute. The following is the general form of Hogarth and Einhorn's model (p. 8) using notation modified in part to accommodate our interest in the presentation of just two sides of an account, or cases, presented sequentially:

$$c_{i,j} = c_{i,j-1} + w_j(s_j - r) \quad (1)$$

where $c_{i,j}$ corresponds to confidence that case i is the best account after hearing the case presented in the j th position, and $0 \leq c_{i,j} \leq 1$. Because there are two sides, or cases, and each can be presented either first or second, i and $j = 1$ or 2 . For example, $c_{1,1}$ corresponds to confidence in side 1 after hearing the first case, and $c_{1,2}$ corresponds to confidence in side 1 after hearing both cases. (The model does not specify which case is presented first.) Confidence in case i before hearing the case presented in the j th position is denoted by $c_{i,j-1}$. If $j = 1$, then $c_{i,j-1}$ is confidence before hearing either case; that is, it is the prior belief that case i is most accurate. Furthermore, s_j is the *independently judged* strength of the case presented in the j th position. (As will become clear later, it is important to distinguish between a case's strength and how the case changes confidence.) Bounds on s_j depend on the specific form the model takes and are therefore defined later. r is the reference point against which a case's strength is evaluated and will be discussed in more detail below. Finally, w_j is the adjustment weight for the case presented in the j th position, where $0 \leq w_j \leq 1$. It affects the extent to which a case changes confidence. However, our emphasis is on direction of change rather than amount of change, so for simplicity, let $w_j = c_{i,j-1}$ if $s_j \leq r$, and let $w_j = 1 - c_{i,j-1}$ if $s_j > r$. This means that the extent to which a case affects confidence depends in part on how much 'room' is left on the scale. For example, a strong case cannot affect confidence much if

one is already 95% confident in that side. The same case would be expected to result in a larger (absolute) change if one started out, say, 40% confident. (For a more general treatment of w , see Hogarth and Einhorn, 1992.)

According to this general model, confidence after being presented with a case is equal to prior confidence plus an adjustment based on the new case. Note that the direction of change in confidence depends on the sign of the difference between the strength of a case, s_j , and the reference point, r . If s_j is greater than r , confidence increases. If s_j is less than r , confidence decreases.

Most important for present purposes is the reference point, r . Hogarth and Einhorn (1992, p. 9) argue that r will be equal either to zero or to prior confidence, $c_{i,j-1}$. If $r = 0$, the result is

$$c_{i,j} = c_{i,j-1} + w_j s_j \quad (2)$$

For this form of the model, let $-1 \leq s_j \leq 1$: A case independently judged to support one side would receive a positive rating, a case judged to support the other side would receive a negative rating, and a neutral case would receive a rating of zero. Equation (2) is an adding model and makes clear that direction of change in confidence depends entirely on the sign of the case's strength, s_j (e.g. Carlson and Dulany, 1988; Edwards, 1968; Wallsten and Manley Sapp, 1977). A case supporting a particular side will pull confidence in the direction of that side.

For example, imagine a defendant standing trial on criminal charges. Assume that a strong prosecution case is presented first, leading to a moderately strong belief that the defendant is guilty ($c_{1,1} = 0.8$), and a weak defense is subsequently presented ($s_2 = -0.1$). According to the above model, confidence in guilt after being presented with the weak defense will be: $c_{1,2} = 0.8 + 0.8(-0.1) = 0.72$. Thus, the adding model predicts that the weak defense will decrease confidence in guilt.

Still using equation (1), now let $r = c_{i,j-1}$, or confidence prior to hearing the case presented in the j th presentation:

$$c_{i,j} = c_{i,j-1} + w_j (s_j - c_{i,j-1}) \quad (3)$$

For this form of the model, let $0 \leq s_j \leq 1$. A case independently judged to support one side would be rated greater than 0.5, a case judged to support the other side would be rated less than 0.5, and a neutral case would receive a rating of 0.5. According to equation (3), direction of change in confidence depends on the relation between the strength of the new case, s_j , and prior confidence, $c_{i,j-1}$. If s_j is greater than $c_{i,j-1}$, confidence increases. If s_j is less than $c_{i,j-1}$, confidence decreases. By rearranging terms in equation (3), it is easy to see that it is an averaging model:

$$c_{i,j} = (1 - w_j)c_{i,j-1} + w_j s_j$$

Thus, current belief is a weighted average of prior strength of belief and the strength of the new case (Anderson, 1981; Lopes, 1985, 1987; Ostrom, Werner, and Saks, 1978; Shanteau, 1970, 1972, 1975). Resulting confidence must lie somewhere between the two.

Consider again a defendant on trial, with a strong prosecution having resulted in moderately strong belief in guilt ($c_{1,1} = 0.8$) before a weak defense is presented ($s_2 = 0.4$; recall that 0.5 is a neutral case for this model). According to the averaging model, confidence in guilt after hearing the weak defense will be: $c_{1,2} = 0.2(0.8) + 0.8(0.4) = 0.48$. Thus, the weak defense is predicted to decrease confidence in guilt, just as with the adding model.

Together, adding and averaging models can account for an impressive amount of complex belief-updating behavior (see Hogarth and Einhorn, 1992), but, as mentioned, these models have tended to focus on change

in confidence following individual pieces of evidence rather than opposing sides of a dispute, which is the focus of our model. For reasons discussed throughout the rest of this article, we believe that this distinction is important.

A NEW REFERENCE POINT: MINIMUM ACCEPTABLE STRENGTH

In the criminal trial example, both adding and averaging models predicted that the weak defense would decrease confidence in guilt. According to the adding model, a case supporting one side should never increase confidence in the opposing side, no matter how weak the case. Furthermore, because the strength of the defense was less extreme than confidence in guilt before hearing the defense, the averaging model also predicted that the defense would decrease confidence. We believed, though, that a weak case might increase confidence that the *opposing* account was the best one. Such a finding would pose problems for both adding and averaging models as depicted above and imply the need for a new reference point, which we refer to as a case's minimum acceptable strength (MAS).

A natural way to present our proposed reference point is simply to insert it into Hogarth and Einhorn's general model (equation (1)), letting $r = m_j$, the MAS of the case presented in the j th position:

$$c_{i,j} = c_{i,j-1} + w_j(s_j - m_j) \quad (4)$$

Assume that confidence is being reported in terms of case 1 and it is presented first. Under such circumstances, s_1 and m_1 would be expected to range between 0 and 1, and s_2 and m_2 between -1 and 0, though we do not expect m_j to deviate far from 0 in most situations, including our own studies (for reasons we discuss later). In the above model, as with the adding and averaging models, confidence after being presented with a case is equal to prior confidence plus an adjustment based on the new information. The adjustment, however, is not based on a comparison of case strength to a neutral reference point, nor to prior confidence, but to the case's MAS. If s_j is greater than m_j , confidence increases. If s_j is less than m_j , confidence decreases. We propose that m_j is generally non-zero (i.e. is not equivalent to a neutral case). In particular, we propose that the reference point that determines change in confidence subsequent to hearing an entire case is more demanding than a neutral zero point. This leaves room for a weak case to fall short, resulting in a reverse impact on confidence.

Imagine again a defendant on trial, a strong prosecution has led to high confidence in guilt ($c_{1,1} = 0.8$), and a weak defense is then presented ($s_2 = -0.1$). Assume further that the defense case's MAS, m_2 , is -0.2 . The weak defense falls below minimum acceptability (i.e. is closer to the neutral value of 0), $s_2 - m_2$ is therefore positive, and confidence in guilt *increases*: $c_{1,2} = 0.8 + 0.2[(-0.1) - (-0.2)] = 0.82$.

We have illustrated the model using these particular circumstances—a strong first case followed by a weak second one—to distinguish it from traditional adding and averaging models. However, a strong first case is not necessary for a second case to have a reverse impact (although, as we argue below, it will help). Furthermore, equation (4) indicates that even a case presented first can have a reverse impact if it is sufficiently weak.

This article does not test the specific form of the model presented in equation (4). We present the model primarily as a tool for communicating the qualitative implications of our proposed reference point, and we will concentrate on testing qualitative differences between our account and competing accounts. Furthermore, some of the ideas tested here lie outside the scope of the model. Our goal is to establish three empirical claims:

Claim 1: The reference point that determines change in confidence following cases is more demanding than neutrality.

This idea was discussed above and is tested in Experiments 1–4.

Claim 2: The reference point for the second case is influenced by the strength of the first case.

If, as we argue, the reference point against which the strength of a case is compared is MAS, this suggests some arbitrariness on the part of the judge as to where to set the criterion. What level of strength is ‘minimally acceptable’? When a straightforward psychological landmark such as neutrality or prior confidence is not used, we suspect that the reference point can be manipulated. In particular, we examined whether the MAS of the second case is influenced by the strength of the first case. We hypothesized that the stronger the first case, the higher (i.e. the farther from 0) the MAS of the second. Raising the level of MAS implies that weak cases will be more likely to fall short and have a reverse impact and that strong cases will have a smaller impact. This notion is tested in Experiments 1–3. We modify equation (4) to capture the conditional nature of a case’s MAS:

$$c_{i,j} = c_{i,j-1} + w_j[s_j - (m_j | e)] \quad (5)$$

where e corresponds to everything known prior to hearing the case presented in the j th position. This article examines only the influence of the strength of the first case on the MAS of the second, but equation (5) allows for the possibility that the MAS of even the first case can be influenced by prior information.

Claim 3: The reference point for cases is more demanding than neutrality because cases are expected to be biased.

Establishing Claim 1 raises an obvious question: Why is the reference point for cases more demanding than neutrality? That is, why would a weak case increase confidence in the opposing account? We propose that it is because it is understood that cases generally do not represent random samples of evidence. Cases tend to be biased accounts of events or issues, and that is why the bar is raised for them. Claim 3, which lies outside the scope of our model, is tested in Experiment 4.

Our proposed explanation of why the bar is raised for cases also speaks to the issue of how much m_j will vary within and between disputes. Though the model can accommodate extreme m_j values, we mentioned earlier that we did not expect m_j to move very far from 0, at least under typical circumstances. MAS is assumed to offset bias, but gross bias is probably rare (perhaps because such bias would be overly transparent, leading the case to have a reverse impact). In the typical dispute, only a moderate amount of bias would be expected, and therefore m_j would also be moderate.

EXPERIMENT 1

Our first experiment presented participants with a scenario in which a defendant was standing trial on burglary charges. One group was presented with a strong prosecution followed by a weak defense, and another group was presented with the reverse order. Participants reported confidence that the defendant was guilty before hearing either side, then after each.

Two additional groups independently evaluated the strength of the defense and the prosecution. Because the proposed model distinguishes between a case’s strength and how the case affects confidence (a weak case

can have a reverse impact), direct measures of case strength are necessary. In particular, it has to be shown that the weak case is indeed seen as supporting the intended side.

Three hypotheses were tested. First, the weak defense was expected to have no effect on, or even increase, confidence in guilt, regardless of whether it was presented first or second (Claim 1). The other two hypotheses regarded the possibility that the strength of the first case influences the MAS of the second (Claim 2). If the strong prosecution presented first raises the MAS of the subsequent defense, the weak defense should increase confidence in guilt more when presented second rather than first. Finally, if the weak defense presented first lowers the MAS of the prosecution, the prosecution should also increase confidence in guilt more when it is presented second rather than first.

Method

Participants were 174 students at the University of California, San Diego, who received partial credit for psychology courses. A group of 93 participants were given a 5-page booklet, the first page of which asked them to imagine that they were jurors on a trial involving a warehouse that was burglarized in the middle of the night. The defendant was arrested the following day after police received an anonymous tip. About half of these 93 participants were told that they would be presented with a summary of the prosecution followed by the defense, and the other half were told the reverse order. They were to report confidence in the defendant's guilt after hearing each side. It was emphasized that each side was making the strongest case possible and that the lawyers for each side were equally competent and motivated. The second page asked for confidence in the defendant's guilt before being presented with either case. The third and fourth pages then summarized the prosecution and defense, asking for confidence after each. Case order depended on condition. Confidence was reported on a 21-point scale, with 0 = 'certain of innocence', 10 = 'guilt as likely as innocence', and 20 = 'certain of guilt'. The final page asked for demographic information.

The summary of the prosecution's case was designed to be rather strong:

- A witness identified the defendant in a police lineup.
- The type of burglary (commercial) is the same as that of the defendant's past convictions.
- The method of entry into the warehouse was unusual and is identical to that of the defendant's past convictions.

The summary of the defense's case was designed to be weak:

- A friend of the defendant's testified that he was with the defendant at his house at the time of the burglary.

The remaining participants were given the same context (defendant on trial for burglary) and rated the strength of the prosecution ($n = 40$) or defense ($n = 41$) on a 21-point scale, with $-10 =$ 'very strong evidence of innocence', $0 =$ 'neutral', and $+10 =$ 'very strong evidence of guilt'.

Results

Independent strength ratings

The independent ratings showed that the prosecution was viewed as rather strong ($M = 5.8$) and the defense as very weak ($M = -0.6$). Most (56%) rated the defense's strength as negative (indicating innocence), 20% rated it as neutral, and 24% rated it as positive (indicating guilt). Eliminating those rating the defense as neutral, 30% rated it as indicating guilt and 70% as indicating innocence.

Quantitative changes in confidence

Six of the 93 participants who were presented with both the prosecution and defense and reported confidence were excluded from the analyses because their absolute confidence or changes in confidence were more than

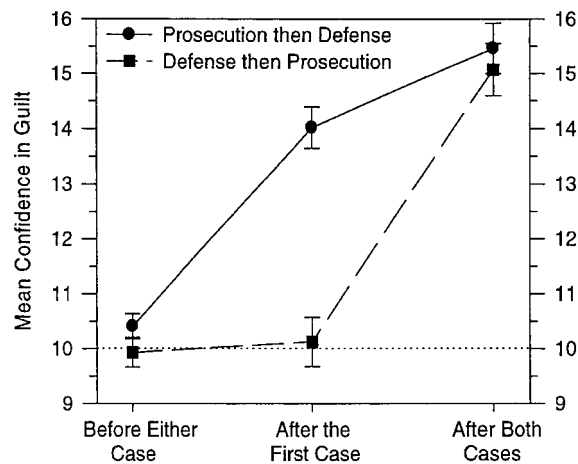


Exhibit 1. Experiment 1: Mean confidence in guilt for the prosecution-then-defense and defense-then-prosecution groups before either case was presented, after the first case, and after both cases. Confidence was reported on a 21-point scale, with 0 = 'certain of innocence', 10 = 'guilt as likely as innocence', and 20 = 'certain of guilt'. Standard error bars are shown.

3 SDs from their group's mean. Most of these eliminations were due to participants' apparently taking seriously the notion of 'presumed innocence' and reporting that they were certain that the defendant was innocent before being presented with either case (Ostrom *et al.*, 1978). These instances were fairly rare and extreme and we have few qualms about eliminating them for present purposes.

Exhibit 1 shows mean confidence in guilt for the prosecution-then-defense ($n = 46$) and defense-then-prosecution ($n = 41$) groups before being presented with either case, after the first case, and after both cases. Although the defense was generally independently rated as indicating innocence, the case nonetheless increased confidence in guilt slightly when presented first (defense-then-prosecution group), and did so even more when presented second, after the strong prosecution (prosecution-then-defense group). Furthermore, the prosecution increased confidence in guilt more when presented second, after the weak defense, rather than first.

A 2(Case Order: prosecution-then-defense versus defense-then-prosecution) \times 2(Case: prosecution versus defense) ANOVA was conducted on change in confidence, using Case as a repeated measure. Not surprisingly, there was an effect of Case, with confidence changing more following the prosecution than the defense ($F(1, 85) = 46.3, p < 0.001$). Most important is the interaction (both cases were expected to lead to larger increases when second), which was significant ($F(1, 85) = 6.4, p = 0.013$). Furthermore, contrasts showed that the prosecution increased confidence more when presented second than first ($t(85) = 2.07, p = 0.041$), and so did the defense ($t(85) = 2.24, p = 0.027$). Finally, change in confidence after the defense was not different from zero when the defense was first ($t(45) = 0.49, p = 0.63$), but it was when the defense was second ($t(40) = 3.78, p < 0.001$).

This pattern of results is as predicted. The defense was independently judged as weakly exonerating, but had no effect on confidence when presented first (because the reference point is more demanding than neutrality), and increased confidence in guilt when presented second (because the strong prosecution raised the bar for the subsequent defense). In addition, the prosecution increased confidence more when presented second (because the weak defense presented first lowered the bar).

Qualitative changes in confidence

To ensure that the results regarding the quantitative changes in confidence were not due to a few outliers, we examined qualitative changes as well. The defense-then-prosecution group was evenly divided between

decreasing, not changing, and increasing confidence in guilt after the defense (33%, 30%, and 37%, respectively). Eliminating those who did not change confidence, we compared the percentage of defense-then-prosecution participants who increased confidence after the defense (53%) to the percentage of non-neutral independent raters who rated the defense as positive (30%); the difference is marginally significant ($p = 0.053$, one-tailed Fisher's exact test). That is, more participants increased confidence in guilt following the defense presented first than independently judged the defense as indicating guilt. In addition, most participants (54%) in the prosecution-then-defense group increased confidence after the defense (20% decreased confidence and 27% showed no change). Again eliminating those who did not change confidence, we compared the percentage increasing confidence after the defense when it was second (73%) to the percentage increasing after it was first (53%); this difference is also marginally significant ($p = 0.083$). The defense was more likely to increase confidence in guilt when it came after, rather than before, the strong prosecution. Finally, the difference between the percentage of (non-neutral) participants rating the defense as positive (30%) and the percentage of (non-neutral) participants increasing confidence when the defense was second (73%) was highly significant, $p < 0.001$.

Discussion

The results were consistent with the three predictions made by our account. First, there was a qualitative difference between the independent strength rating of the defense's case and how it affected confidence. The defense was rated as weakly exonerating, but had no effect on confidence when presented first and increased confidence in guilt when presented second. This was true of both the quantitative and the qualitative changes in confidence. This supports our contention that, when reporting confidence following a case, the case's strength is evaluated relative to a non-neutral level of MAS (Claim 1). A case can have a reverse impact by falling short of this level. Second, the defense resulted in more of an increase in confidence in guilt when presented second rather than first. Third, the prosecution also resulted in a larger increase in confidence in guilt when presented second rather than first. Both of these latter effects are consistent with our position that the MAS of the second case is influenced by the strength of the first case in a predictable manner (Claim 2). The strong prosecution presented first raised the defense's MAS, making it more likely that the defense would fall short when presented second and increase confidence in guilt. Similarly, the weak defense presented first lowered the MAS of the prosecution's case, whose strength then exceeded the reference point by an even greater amount, causing the prosecution to increase confidence more when presented second.¹

With respect to Claim 1, other researchers have shown that evidence sometimes has an impact opposite to that implied by independent evaluations, but this is virtually always due to averaging the strength of prior and new information (Anderson, 1981; Hogarth and Einhorn, 1992; Lopes, 1985, 1987; McKenzie, 1994; Shanteau, 1970, 1972, 1975; for a non-averaging example, see McKenzie, 1999). However, because prosecution-then-defense participants were reasonably confident that the suspect was guilty before being presented with

¹The fact that there was no difference in final confidence between the groups (see Exhibit 1) might raise a concern for some readers. We are claiming that participants processed the same case differently depending on whether it was presented first or second (due to the different levels of MAS), but differences in processing are usually demonstrated by showing a difference in final confidence when order is manipulated. If our participants were processing the cases differently depending on order, as we claim, why was there no difference in final confidence? According to the MAS model, the reason is that the weak defense presented first lowered the bar for the subsequent prosecution, and the strong prosecution presented first raised the bar for the subsequent defense, and effectively canceled each other out. In other words, the boost in confidence in guilt following the prosecution as a result of the weak defense presented first was about equal to the boost following the defense as a result of the strong prosecution presented first. A lack of difference in final confidence for the strong and weak cases is perfectly reasonable according to the MAS model. The model does predict differences in final confidence when both cases are strong (primarily due to the assumed restricted range of m_j relative to s_j), and an experiment not reported here confirmed this prediction. Details of this experiment, along with related results from computer simulations, are available from the first author.

the defense, averaging predicts that the weak defense would decrease, not increase, confidence in guilt, and therefore cannot account for the present results.

However, there are alternative interpretations of some of our results. The two results that we argued indicated malleable MAS (Claim 2) could be explained by recency: Any case might have a larger impact when presented second rather than first. Such effects are common when confidence is updated after pro and con evidence in general (Hogarth and Einhorn, 1992), and in legal scenarios in particular (e.g. Anderson, 1959; Davis, 1984; Furnham, 1986; Walker, Thibaut, and Andreoli, 1972; Weld and Roff, 1938). It is therefore desirable to test Claim 2 under conditions where recency is not a viable alternative.

There is an additional alternative interpretation of the fact that the defense had no effect when presented first, but increased confidence when presented second. Though we believe that the defense was viewed independently as weakly exonerating, assume for the sake of argument that the defense's case is neutral.² Other researchers have demonstrated that participants sometimes use neutral information to support their current favorite hypothesis or preference (Wallsten 1976, 1981; see also Pitz, 1969; Russo, Medvec, and Meloy, 1996). Thus, perhaps when the defense was presented first, participants had no favorite hypothesis and saw the defense's case for what it was, namely, neutral. However, when the defense came after the prosecution, participants at that point were favoring the prosecution and might have used the neutral defense to bolster their preferred guilty hypothesis.

Fortunately, there is a simple way to distinguish between the current account and both alternatives.

EXPERIMENT 2

In Experiment 2, all participants were presented with the prosecution followed by the defense, and the defense was the same weak case used in Experiment 1. However, we manipulated the strength of the prosecution. One group was presented with the same strong prosecution used earlier, but a second group was presented with a weak prosecution. The strong-prosecution group was expected to behave similarly to the analogous group in Experiment 1, namely, increase confidence in guilt following the weak defense. Consider, however, the group presented first with the weak prosecution. Relative to the strong prosecution, the weak prosecution should lower the MAS of the subsequent defense, making it less likely that the strength of the defense will fall short. The weak defense should therefore be less likely to increase confidence in guilt for the weak-prosecution group than for the strong-prosecution group. Ideally, the same defense would lead to a decrease in confidence for the former group and an increase for the latter group.

Such a result could not be explained in terms of recency because it is the same case (the defense) presented in the same position (second) that is leading to different behavior. Furthermore, assuming that both the weak and strong prosecution lead participants to favor the guilty hypothesis before being presented with the defense, qualitatively different responses by the two groups to the defense could not be explained by participants' using the defense to support their current favorite hypothesis.

Method

There were 120 participants from the same population as in Experiment 1. The method and procedure for the strong-prosecution group was identical to that for the prosecution-then-defense group in

²To be sure that the defense case is independently considered exonerating, we asked an additional 60 participants to rate its strength, and the results were almost identical to those reported for Experiment 1. The mean rating was -1.0 , which is significantly different from zero ($t(59) = 1.84$, $p = 0.035$, one-tailed). Qualitatively, 53% of the participants rated the case as indicating innocence, 18% rated it as neutral, and 28% rated it as indicating guilt. It seems fair to say that the case is generally viewed as weakly exonerating.

Experiment 1. The only difference occurred for the weak-prosecution group, which read the following prosecution summary:

- The method of entry into the warehouse was unusual and is identical to that of the defendant's past convictions.

Results

Data for 5 participants were not analyzed for the same reasons given in Experiment 1, leaving 58 participants in the strong-prosecution group and 57 in the weak-prosecution group. Exhibit 2 shows mean confidence in guilt for the two groups before being presented with either case, after being presented with the prosecution, and after being presented with the defense. Note first that both the strong and weak prosecution cases led to increases in confidence. Both groups believed more strongly in the defendant's guilt rather than innocence after the prosecution. Most important is that there are clear differences in reaction to the same subsequent defense. The strong-prosecution group increased confidence in guilt after the defense, whereas the weak-prosecution group decreased confidence. These changes are different ($M_s = +0.7$ versus -1.9 , $t(113) = 4.5$, $p < 0.001$).

Qualitative changes in confidence revealed the same pattern. For the strong-prosecution group, the percentage of participants whose confidence after the defense decreased, remained the same, and increased, was 33, 24, and 43, respectively.³ The respective percentages for the weak-prosecution group were 68, 23, and 9. Eliminating participants who did not change confidence, we compared the percentage of strong-prosecution participants who increased confidence after the defense (57%) to the percentage of weak-prosecution participants who increased confidence after the defense (11%). This difference is significant, $\chi^2(1, N = 88) = 20.2$, $p < 0.001$.

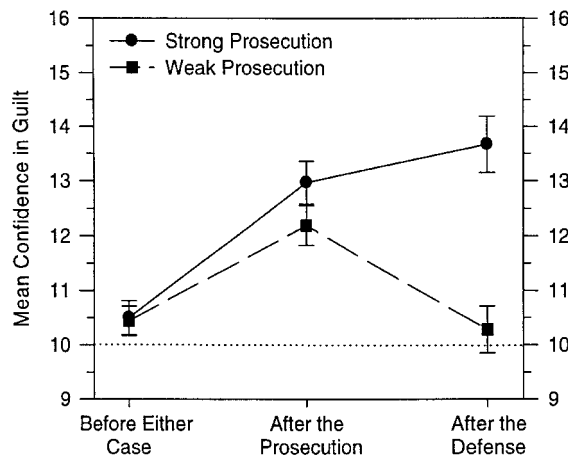


Exhibit 2. Experiment 2: Mean confidence in guilt for the strong-prosecution and weak-prosecution groups before either case was presented, after the prosecution, and after the defense. Confidence was reported on a 21-point scale, with 0 = 'certain of innocence', 10 = 'guilt as likely as innocence', and 20 = 'certain of guilt'. Standard error bars are shown.

³This is the smallest effect of the weak defense subsequent to the strong prosecution that we have witnessed in several experiments, shown in both the quantitative and the qualitative changes in confidence. Despite the relatively small increase in confidence, the direction of the effect is consistent with the other experiments and, for present purposes, the most important comparison is between the two groups in the current experiment.

Discussion

The weak defense tended to increase confidence in guilt when it followed the strong prosecution, but it decreased confidence when it followed the weak prosecution. The first finding replicates Experiment 1 and again supports our Claim 1 that change in confidence depends on a comparison between a case's strength and its MAS. The two current findings in conjunction support our Claim 2 that the MAS of the second case is influenced by the strength of the first. Relative to the strong prosecution, the weak prosecution lowered the MAS of the defense, making it more likely that the defense would exceed it and decrease confidence in guilt. Recency cannot account for these results because they concern the same case presented in the same position. Furthermore, because both groups were more confident in guilt than innocence before being presented with the defense, it cannot be argued that participants used the defense to support their favored hypothesis.

We want to highlight a notable aspect of the results. Exhibit 2 shows that strengthening the prosecution had a bigger impact on the influence of the subsequent defense than on the influence of the prosecution itself. The difference between the groups is small after being presented with the different prosecutions, but is large after being presented with the same defense. This is just as predicted by the model. The strong prosecution primarily influenced final confidence indirectly, by raising the bar for the defense, rather than by directly convincing participants of the defendant's guilt.

EXPERIMENT 3

This experiment employed a design similar to that of Experiment 1, where the order of a weak and a strong case was manipulated, and an additional group independently evaluated the strength of the weak case. However, rather than using a legal scenario, participants read about an interpersonal dispute between two former roommates, one of whose car was damaged while he was away, and at issue was whether the other roommate caused the damage.

This experiment served three major purposes. First, if the same pattern of results is found with an interpersonal dispute, this would generalize the current ideas beyond the legal scenario used in Experiments 1 and 2. It is possible that there is something special about a legal setting that led to our results.⁴ Relatedly, whereas the instructions in Experiments 1 and 2 emphasized that both sides were presenting the strongest possible case, no mention of this was made in Experiment 3. We believe that this is generally understood by third parties privy to a dispute—legal or otherwise—and that it is therefore unnecessary to make this explicit in order for the current claims to hold.

Second, we have been discussing the current effects in terms of a shifting reference point (m_j), while holding the strength of a given case (s_j) constant. But one could potentially view the effects in terms of a case's perceived strength changing, while holding the reference point constant. Maybe the strength of the first case changes the subjective strength of the second case. A strong first case might make a second case appear weaker than if it had been presented first. Such contrast effects are common and could potentially explain the qualitative reversals of evidence shown here. However, demonstrating a qualitative difference between a case's independent rating and how it changes confidence when presented *first* would bolster our account. Our account predicts such a result, but it is hard to see how it could be explained in terms of contrast effects since these two groups would be responding in a qualitatively differently manner to the same first case presented to them. Experiment 1 showed that the weak defense led more participants to rate it as exonerating than to decrease confidence in guilt when it was presented first, but the effect was only marginally significant

⁴An obvious legal issue is 'burden of proof', but it seems to us that the most natural interpretation of this is that one should have a relatively high confidence threshold for concluding that a defendant is guilty, not that evidence should be processed in a biased manner. Furthermore, any biased processing of the evidence favoring the defendant would lead to the opposite of our finding that a weak defense increases confidence that the defendant is *guilty*.

(and we eliminated from the analysis participants who did not change confidence). Showing a reliable difference is important because, if the current account is correct, clear evidence of MAS should show up for the first case as well as the second. To this end, participants in the current experiment responded to a forced-choice question regarding whether the case increased or decreased their confidence.

Finally, showing that there is a qualitative difference between how a weak case is independently rated and how it changes confidence when presented *first* would undermine yet another alternative explanation of why a weak second case can have a reverse impact. It is conceivable that participants' confidence after the first case reflects, to some extent, the anticipated strength of the second (still unheard) case. In other words, participants might hold back judgment somewhat before hearing the second side. If so, then if the second case is presented and turns out to be weaker than 'assumed' when confidence was reported after the first case, confidence in the opposing account could increase. (In order to explain the results of Experiment 2, it would also have to be assumed that the stronger the first case, the stronger the assumed strength of the second case.) The appeal of this alternative account lies in its ability to explain why a weak case presented second can have a reverse impact. It is difficult, however, for it to explain a similar effect when the weak case is presented first.

Method

There were 88 participants from the same population as those in Experiments 1 and 2. All participants were presented with the same background information:

Imagine that two friends of yours, Frank and Bob, shared an off-campus apartment last summer. Frank flew to visit his parents one weekend, leaving his car parked on the street. When he returned on Sunday, he saw that there was damage to the driver's side of the car. Fixing it will cost at least \$2,000. At issue is whether Bob caused the damage to Frank's car.

There were three groups. One (the independent judges) was presented only with Bob's account, designed to be weak:

- Bob denies driving Frank's car while Frank was away. He does not know how the damage occurred.

These independent judges were then asked a forced-choice question: Does Bob's account indicate that he did or did not cause the damage?

The other two groups read that they would be presented with both Bob's and Frank's accounts and report confidence after each account that Bob caused the damage to Frank's car. Both of these groups then reported confidence before hearing either side. Confidence was reported on a 21-point scale, with 0 = 'certain that Bob did not cause the damage', 10 = 'both equally likely', and 20 = 'certain that Bob did cause the damage'. One of these two groups was then presented with Bob's account on the next page and answered a forced-choice question: 'Does Bob's account increase or decrease your confidence that Bob caused the damage?' They then read a summary of Frank's account on the next page, designed to be rather strong:

- Two neighbors said they saw Bob driving Frank's car—often recklessly—much of Saturday.
- The next-door neighbor said that Bob appeared to be drinking rather heavily on Saturday.
- The object that caused the damage left some of its paint on Frank's car. This paint matches the color of another car that was reported to the police Saturday night as having been hit while parked (unoccupied) in the parking lot of the local liquor store.

Confidence that Bob caused the damage was again reported using the 21-point scale. The final group read Frank's account first, rated confidence, then read Bob's account and answered the forced-choice question.

Thus, all three groups responded to a forced-choice question regarding Bob's account, which is sufficient for our purposes. The independent judges responded whether it indicated that Bob did or did not cause the

damage; the Bob-then-Frank group responded whether it increased or decreased their confidence that Bob caused the damage when it came first; and the Frank-then-Bob group responded whether it increased or decreased their confidence that Bob caused the damage when it came second.

Results and discussion

Bob's weak account was independently judged to support his side: Only 4 of 30 independent judges (13.3%) reported that the account indicated that Bob caused the damage to Frank's car. However, when the group presented first with Bob's case was asked whether it increased or decreased their confidence that Bob caused the damage, 11 of 28 participants (39.3%) reported an increase. The difference between these percentages is reliable ($p = 0.025$, one-tailed Fisher's exact test). This effect is predicted by the current account, where the reference point is more demanding than neutrality for even the first case, but is problematic for a contrast-effect account of our data, as well as an alternative account in which participants are assumed to withhold judgment to some extent after hearing the first case. Furthermore, when Bob's case was presented second, 23 of 30 (76.7%) reported an increase in confidence that he caused the damage. The difference between the number of participants reporting an increase in confidence after Bob's case when it was first versus second was also reliable ($p = 0.004$). This was predicted because Frank's strong case presented first raised the bar for Bob's account, making it more likely that it would fall short. Note the large difference between the percentage of participants independently judging Bob's case to indicate that he caused the damage (13.3%) and the percentage of participants increasing confidence that he caused the damage after his case was presented second (76.7%). These results replicate those of the earlier experiments, despite using an interpersonal rather than a legal dispute, and despite eliminating all references to each side presenting the strongest possible case. The current ideas generalize beyond the legal scenario used in Experiments 1 and 2.

EXPERIMENT 4

The above experiments provided evidence for Claim 1 (the reference point that determines change in confidence after hearing a case is more demanding than neutrality) and Claim 2 (the reference point for the second case is influenced by the strength of the first case). We now address why Claim 1 is true. Why is the reference point for cases more demanding than neutrality? Claim 3 states that it is because cases are expected to be biased summaries of available evidence. Cases are generally not random samples and must therefore clear a higher hurdle in order to have their intended impact on confidence.

In Experiment 4, we manipulated whether the weak case was presented by one of the disputants (as in Experiments 1–3) or by a neutral third party. Presumably, the evidence would more likely be seen as biased in the former condition. The prediction was that the weak case presented by the disputant would be more likely to increase confidence in the opposing side.

Method

Participants were 155 UCSD students recruited from the same population as the previous studies. There were two groups, both presented with the same context used in Experiment 4 (the interpersonal dispute regarding the damaged car). One group was told that they were going to be presented with Bob's account, the person whose involvement was at issue. Before hearing Bob's weak account, participants reported their confidence that Bob caused the damage. They were then presented with Bob's account (he denies driving the car and does not know how the damage occurred) and responded whether their confidence that Bob caused the damage increased or decreased.

The other group was told that they were going to be presented with a summary of all the evidence gathered by Amy, a friend of both disputants. They read that Amy had no interest in showing that Bob was or was not responsible. She simply wanted to settle the matter by finding out the truth about what happened. Before being presented with Amy's summary, participants reported their confidence that Bob caused the damage. They then reported whether Amy's summary (the same information presented by Bob above) increased or decreased their confidence that Bob caused the damage.

Results and discussion

When Bob presented the weak case, 27 out of 77 participants (35.1%) increased confidence that Bob caused the damage, but when Amy presented the same information, only 14 out of 78 (17.9%) increased confidence ($\chi^2(1, N = 155) = 5.8, p = 0.02$). The weak evidence was almost twice as likely to have a reverse impact when Bob rather than Amy presented it. Note also that the first percentage replicates Experiment 3 (where 39.1% reported an increase when Bob's account came first) and the second value is similar to the percentage of participants in Experiment 3 who independently judged Bob's case as indicating that he caused the damage (13.3%). The similarity between the latter values suggests that unbiased evidence will change confidence in a manner consistent with independent evaluations, just as one might expect given our account.

We take these findings to support Claim 3: Weak cases can have a reverse impact because cases are generally expected to be biased, thereby raising the bar. Presenting the same information in an unbiased fashion makes it much less likely that it will have a reverse impact on confidence.

GENERAL DISCUSSION

The current results indicate that, when two sides of a dispute are presented sequentially, a case's impact depends not on its strength, but on its strength relative to its minimum acceptable strength (MAS). This reference point is more demanding than neutrality, leaving room for a weak case to fall below it and have a reverse impact (Claim 1). Four experiments showed that cases independently rated as weakly supporting one side often increased confidence in the opposing side. The implicit reasoning seems to be along the lines of 'If *that's* the best they can do, then I believe the other side (even) more'.

In addition, the MAS of the second case appears to be influenced by the strength of the first case. The stronger the first case, the higher the MAS of the second (Claim 2). Increasing MAS makes it more likely that weak cases will fall short and have a reverse impact on confidence, and that strong cases will have a smaller impact. Consistent with this, Experiments 1 and 3 showed that a weak case was more likely to increase confidence in the opposing side when it came after rather than before the opposing side's strong case. Similarly, the strong case in Experiment 1 led to a larger increase when it followed rather than preceded the opposing weak case. In addition, Experiment 2 showed that the same weak case presented second increased confidence when preceded by a strong opposing case, but decreased confidence when preceded by a weak case. Relative to the strong first case, the weak first one lowered the MAS of the second case, making it more likely that its strength would exceed this level and therefore have its intended impact. Taken together, we believe that this is persuasive evidence that MAS is malleable.

It is not hard to imagine how MAS might be influenced by the strength of the preceding case. Consider, for example, the fact that the weak case in Experiments 1 and 3 was more likely to increase confidence in the opposing side when presented second rather than first. Participants might have expected at least some of the points made by the first side to be addressed by the second side, and it might have been salient that this did not occur. This, of course, could not happen when the weak case came first. More generally, a strong first case might evoke a commonly expressed sentiment with respect to an upcoming second case: 'They've got a lot of explaining to do.'

The results are also consistent with our claim that MAS is more demanding than neutrality because judges expect cases to be biased samples of evidence (Claim 3). Experiment 4 showed that a weak case was less likely to increase confidence in the opposing side when it was presented by a neutral third party rather than by a disputant. Indeed, when presented by a neutral third party, the weak case increased confidence in the opposing side about as rarely as it was independently judged to support the opposing side.

The present experiments can be seen as examining how people process biased information. Other researchers have also examined this and concluded that people adjust insufficiently for bias (Brenner, Koehler, and Tversky, 1996; Hamill, Wilson, and Nisbett, 1980; Nisbett and Ross, 1980; Ross, Amabile, and Steinmetz, 1977). Most relevant is research conducted by Brenner *et al.* (1996), who presented some participants with just one side of a legal dispute and asked them to predict how many fully-informed participants (those presented with both sides) would favor a particular side. They found, for example, that those presented with only the plaintiff's case overestimated the number of fully informed participants who sided with the plaintiff, indicating that participants were not adjusting enough for the fact that they were hearing only one side. These results are not inconsistent with our own. We provided evidence that participants adjusted for bias (Experiment 4 provided the most direct evidence), but not that the adjustment was optimal (it is not even obvious in our studies what that amount would be). Similarly, Brenner *et al.* (1996) did not argue that participants did not adjust at all for bias, only that the adjustment was insufficient.

Brenner *et al.* (1996) proposed that, to adjust for hearing only one side of a dispute, participants might consider how strong the opposing side was likely to be. Indeed, the authors found that having participants consider the likely strength of the opposing side increased the adjustment significantly. Our experiments indicate, however, that people naturally adjust for bias in a different manner: A case is compared to a demanding standard (minimum acceptable strength) that depends (in part) on the amount of bias expected for that case. This is a reasonable, albeit perhaps unexpected, strategy for dealing with bias. Whether or not the strategy results in normatively defensible confidence after hearing only one side will depend on whether the expected bias (i.e. the degree to which m_j differs from 0) is appropriate in that instance. It is possible, though, that, in general, MAS is not sufficiently demanding and that therefore bias is not sufficiently adjusted for after hearing only one side of a dispute.

Along similar lines, we see in hindsight that there was a plausible alternative way to process biased evidence: It could simply be discounted. One way to capture this idea formally is to multiply s_j by a parameter, k , ranging between 0 and 1 (and letting $r=0$). When $k=0$, evidence would simply be ignored, perhaps because it is expected to be completely biased and therefore worthless, and when $k=1$, evidence would have full impact, perhaps because the evidence is seen as completely unbiased. However, such a view could not explain our results because biased information was not merely discounted or watered down; it sometimes had a reverse impact. Rather than multiplying evidence strength by a constant, our effects are algebraically akin to having a constant subtracted from it—which is how we modeled the process. There might be other situations, though, in which the evaluation of biased evidence is best described as a discounting process.

Recall our rationale behind the claim that the strength of the first case influences the MAS of the second. We noted that what constitutes minimally acceptable strength has no straightforward answer and would therefore be malleable. However, another reason that the strength of the first case might influence the MAS of the second is that the strength of the first case might indicate how large the entire body of available evidence is. Perhaps the stronger the first case, the larger the body of relevant evidence, generally speaking. If there are many facts associated with a dispute, it might be reasonable to assume that it would be easy to select evidence to build an impressive case—and that failure to do so would be all the more revealing.

We pointed out earlier that traditional adding and averaging models cannot explain the data reported here. We also argued that several alternative accounts of our data did not fare well, including recency (Experiment 2), using neutral data to support a favored hypothesis (Experiment 2), an account whereby the anticipated strength of the second (unheard) case is taken into account when reporting confidence after the first case (Experiment 3), and contrast effects (Experiment 3). Relatedly, Pennington and Hastie (1993)

have made a compelling argument that jurors' decisions in criminal trials depend on whether the prosecution's or defense's account results in a more coherent story based on presented evidence. For example, presenting information in the same temporal order of the event in question facilitates the listener's ability to construct a coherent story and therefore influences perceived case strength and verdict decisions (Pennington and Hastie, 1988). Perceived causal links between facts, as opposed to just the facts themselves, are important. Pennington and Hastie (1992) compared their story model to algebraic anchor-and-adjust models, like those discussed in this article (see also Pennington and Hastie, 1981). The anchor-and-adjust models described behavior well when participants were forced to update confidence after each bit of evidence, but not when asked for confidence only after hearing both sides. The story model performed best in the latter situation, which is presumably more natural for jurors. Our focus, however, is not on final confidence after hearing both sides (where the story model performs well), nor on change in confidence after each piece of evidence (where traditional belief-updating models perform well), but on change in confidence after hearing each side. The story model primarily addresses the factors underlying case strength and final decisions, whereas we are interested how cases of *given* strengths change confidence when presented sequentially. Because the story model does not address belief updating, it neither predicts nor accounts for our findings.

It is perhaps worth mentioning that, though our main findings have been discussed in terms of negative evidence increasing confidence in the opposing side, this is somewhat arbitrary. The results can just as easily be framed in terms of positive evidence decreasing confidence by simply reversing the poles of the scales (e.g. reporting confidence in innocence rather than guilt). There is nothing asymmetric about the current findings.

Though we have (necessarily) sampled only a small number of possible situations, we believe our view of belief updating is relevant to any situation in which a third party is presented sequentially with two sides of a dispute: in the courtroom, in formal debates, when public figures respond to allegations, when academics publish replies, in interpersonal and professional disagreements. The experiments reported here are best viewed as providing evidence that the two reference points suggested by Hogarth and Einhorn (1992) do not exhaust the possibilities. Classic adding and averaging models do account for much belief-updating behavior, but they do not describe updating that takes place after hearing both sides of a dispute. Under these conditions, the current account provides a better explanation of change in belief.

ACKNOWLEDGEMENTS

This research was supported by National Science Foundation Grants SBR-9515030 and SES-0079615. The authors thank Jerome Busemeyer, Tim Rickard, Tom Wallsten, Paul Windschitl, John Wixted, Mike Ziolkowski, and several anonymous reviewers for helpful discussions and comments. Portions of the research were presented at the 38th Annual Meeting of the Psychonomic Society and the 1998 Annual Meeting of the Society for Judgment and Decision Making.

REFERENCES

- Anderson, N.H. (1959). Test of a model for opinion change. *Journal of Abnormal and Social Psychology*, **59**, 371–381.
- Anderson, N.H. (1981). *Foundations of Information Integration Theory*. New York: Academic Press.
- Brenner, L.A., Koehler, D.J., and Tversky, A. (1996). On the evaluation of one-sided evidence. *Journal of Behavioral Decision Making*, **9**, 59–70.
- Carlson, R.A., & Dulany, D.E. (1988). Diagnostic reasoning with circumstantial evidence. *Cognitive Psychology*, **20**, 463–492.
- Davis, J.H. (1984). Order in the courtroom. In D.J. Müller, D.E. Blackman, & A.J. Chapman (Eds.). *Psychology and Law: Topics from an International Conference* (pp. 251–265). New York: Wiley.
- Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.). *Formal Representation of Human Judgment* (pp. 17–52). New York: Wiley.

- Furnham, A. (1986). The robustness of the recency effect: Studies using legal evidence. *Journal of General Psychology*, **113**, 351–357.
- Hamill, R., Wilson, T.D., & Nisbett, R.E. (1980). Insensitivity to sample bias: Generalizing from atypical cases. *Journal of Personality and Social Psychology*, **39**, 578–589.
- Hogarth, R.M., & Einhorn, H.J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, **24**, 1–55.
- Lopes, L.L. (1985). Averaging rules and adjustment processes in Bayesian inference. *Bulletin of the Psychonomic Society*, **23**, 509–512.
- Lopes, L.L. (1987). Procedural debiasing. *Acta Psychologica*, **64**, 167–185.
- McKenzie, C.R.M. (1994). The accuracy of intuitive judgment strategies: Covariation assessment and Bayesian inference. *Cognitive Psychology*, **26**, 209–239.
- McKenzie, C.R.M. (1999). (Non)Complementary updating of belief in two hypotheses. *Memory and Cognition*, **27**, 152–165.
- Nisbett, R., & Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice Hall.
- Ostrom, T.M., Werner, C., & Saks, M.J. (1978). An integration theory analysis of jurors' presumptions of guilt or innocence. *Journal of Personality and Social Psychology*, **36**, 436–450.
- Pennington, N., & Hastie, R. (1981). Juror-decision making models: The generalization gap. *Psychological Bulletin*, **89**, 246–287.
- Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**, 521–533.
- Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology*, **62**, 189–206.
- Pennington, N., & Hastie, R. (1993). The story model for juror decision making. In R. Hastie (Ed.), *Inside the Juror: The Psychology of Juror Decision Making* (pp. 192–221). Cambridge: Cambridge University Press.
- Pitz, G.F. (1969). An inertia effect (resistance to change) in the revision of opinion. *Canadian Journal of Psychology*, **23**, 24–33.
- Ross, L., Amabile, T.M., & Steinmetz, J.L. (1977). Social roles, social control, and biases in social-perception processes. *Journal of Personality and Social Psychology*, **35**, 485–494.
- Russo, J.E., Medvec, V.H., & Meloy, M.G. (1996). The distortion of information during decisions. *Organizational Behavior and Human Decision Processes*, **66**, 102–110.
- Shanteau, J.C. (1970). An additive model for sequential decision making. *Journal of Experimental Psychology*, **93**, 63–68.
- Shanteau, J. (1972). Descriptive versus normative models of sequential inference judgment. *Journal of Experimental Psychology*, **85**, 181–191.
- Shanteau, J. (1975). Averaging versus multiplying combination rules of inference judgment. *Acta Psychologica*, **39**, 83–89.
- Walker, L., Thibaut, J., & Andreoli, V. (1972). Order of presentation at trial. *Yale Law Journal*, **82**, 216–226.
- Wallsten, T.S. (1976). A note on Shanteau's 'Averaging versus multiplying combination rules of inference judgment'. *Acta Psychologica*, **40**, 325–330.
- Wallsten, T.S. (1981). Physician and medical student bias in evaluating diagnostic information. *Medical Decision Making*, **1**, 145–164.
- Wallsten, T.S., & Manley Sapp, M. (1977). Strong ordinal properties of an additive model for the sequential processing of probabilistic information. *Acta Psychologica*, **41**, 225–253.
- Weld, H.P., & Roff, M. (1938). A study in the formation of opinion based upon legal evidence. *American Journal of Psychology*, **51**, 609–628.

Authors' biographies:

Craig R.M. McKenzie received his PhD in psychology from the University of Chicago in 1994 and is now Associate Professor in the Department of Psychology at the University of California, San Diego. His research interests include inference, uncertainty, and choice.

Susanna M. Lee and **Karen K. Chen** are former research assistants at the University of California, San Diego, where they received their BA degrees in psychology.

Authors' addresses:

Craig R.M. McKenzie, **Susanna M. Lee** and **Karen K. Chen**, Department of Psychology, University of California, San Diego, La Jolla, CA 92093-0109, USA.