(2007). In D. S. McNamara & G. Trafton (Eds.) *Proceedings of the Twenty-ninth Annual Conference of the Cognitive Science Society* (pp. 503-508). Mahwah, NJ: Erlbaum.

# Surprise, Surprise: The Role of Surprising Numerical Feedback in Belief Change

**Edward Munnich (emunnich@usfca.edu)**
University of San Francisco, Psychology Department, 2130 Fulton Street, San Francisco, CA 94117

**Michael Andrew Ranney (ranney@cogsci.berkeley.edu)**
University of California, Graduate School of Education, 4533 Tolman Hall, Berkeley, CA 94720-1670

**Mirian Song (mirian@berkeley.edu)**
University of California, Graduate School of Education, 4533 Tolman Hall, Berkeley, CA 94720-1670

## Abstract

What percentage of U.S. residents is incarcerated? If you now learned the amount and it surprised you, would it be more memorable than if it were not surprising? Our past research documented conceptual changes related to policy issues when one receives a single, critical number. In the present study, Experiment 1 uses a Numerically-Driven Inferencing (NDI) paradigm method in which participants estimated policy-relevant quantities, learned the true quantities, and rated their surprise regarding that feedback. When asked to recall the quantities either eight or 84 days post-feedback, participants improved the most over their original estimates on items that surprised them the most. In Experiment 2, we found that a measure of prospective surprise ("shock"; Ranney, Cheng, Nelson, & Garcia de Osuna, 2001)—derived from an interval in which participants believed the number fell, and participants' confidence that the number fell in that interval—reliably predicted retrospective surprise ratings like those in Experiment 1. We conclude that surprise is a rather stable construct about which people have considerable metacognition. Future work in this area may suggest how leaders, voters, and consumers can best employ their emotional responses to numbers and enhance cognitive strategies that help shape effective policies.

**Keywords:** Belief Revision; Conceptual Change; Cognition & Emotion; Decision Making; Estimation; Reasoning.

Please write down an answer to the following: How many pounds of garbage are produced per capita, per day, in the U.S.? Include all trash produced by individuals directly and by businesses (except for waste from construction sites). Reflect briefly on how you reached this estimate. Now, think of what you would *prefer* the quantity to be. The true quantity is 4.5 lbs. (Table 1). To the extent that this statistic surprised you, will the surprise make this statistic more memorable days or months from now?

Surprise has been identified as one of the six basic, universal emotions (Ekman, 1992), as it is associated with a distinct bodily reaction over widely divergent cultures. Thagard (2006) highlights the particular adaptive value of surprise, in that it is the emotional response yielded when one receives information that does not cohere with one's current representations. As such, surprise leads one into a cycle of questioning and, possibly, discovery. For instance, one might estimate garbage production based on what one hauls to the curbside each week. This strategy would likely lead to a low estimate, as it leaves out the waste produced by businesses, and leads to surprise when one learns how high the true amount is. Alternatively, one might picture dumpsters lined up outside of businesses and landfills in one's community and not fully grasp how many people are responsible for producing that much trash; this would lead to overestimation of per-person garbage production, and might yield surprise when one sees that the true amount is much lower. In each of these cases, the surprise brought on by receiving the true quantity may lead to a reexamination of one's assumptions, and could ultimately lead to a greater coherence across one's network of beliefs about garbage and related concepts (e.g., the economy).

Wertheimer (1945) proposed that surprising feedback (e.g., a recalcitrant badminton opponent) leads to a restructuring of one's beliefs, resulting in better solutions to problems. In terms of cognitive development, Piaget (1977) argued that highly surprising information leads to accommodative change, whereas information that is minimally surprising leads to assimilation into extant schemas. More recent conceptual change research describes discrepant events—outcomes that run counter to expectations—which redirect attention and may lead to the restructuring of networks of established beliefs (Clement & Steinberg, 2002; Ranney & Schank, 1998).

Literature on discrepant events often focuses on science activities, in which surprising outcomes direct students' attention to beliefs that need revision, and this leads to deeper understandings of scientific principles. For instance, Clement and Steinberg (2002) discussed a discovery process "Susan" exhibited in building electrical circuits. She initially conceived of electrical current as flowing from a battery to a bulb to light the bulb. This assumption works for simple circuits, but fails for more complex ones: When a capacitor (two metal plates separated by an insulator) was placed between battery and bulb, Susan predicted that the bulb would not light, reasoning that current could not traverse the capacitor's insulation. She was surprised, therefore, when the bulb did light—leading her to develop a model in which current may originate from any piece of metal (e.g., the aluminum plate of the capacitor nearest the bulb), not just from batteries. As Susan saw ever more sophisticated circuits, she ran into conflicts, was surprised, and then made relevant revisions to her belief network about circuits. In terms of the ECHO model of reasoning (e.g., Ranney & Thagard, 1988), which represents hypotheses and evidence as propositional nodes within a connectionist network, discrepant events may yield more coherent networks by (a) adding new nodes, (b) adding excitatory or inhibitory links among nodes, and/or (c) changing weights on existing links. Discrepant events are, thus, useful both in pedagogy and in belief revision research.

Related research on knowledge of policy-relevant numbers has illustrated ways in which knowing a particularly salient single number can restructure one's beliefs about an issue, and have both qualitative and quantitative consequences for one's preferences. Munnich, Ranney, Nelson, Garcia de Osuna, and Brazil (2003) found that the same underlying quantity (the number of abortions in the U.S.) can yield dramatically different levels of estimation accuracy. Depending on how the question is addressed, receiving feedback on actual quantities can either provoke (a) *policy shifts*, which are non-proportional changes in numerical preferences (e.g., going from preferring half as much of what a quantity is thought to be, to preferring one-tenth as much of what the true quantity is), or (b) a mere *proportional rescaling* of prior preferences (e.g., although feedback changes what one believes the quantity to be, one maintains a preference that it be halved). As an illustration of this, when asked the ratio of abortions to *births* in the U.S., participants were quite inaccurate and showed large policy shifts upon learning the true value; in contrast, regarding the ratio of abortions to *fertile women*, participants were much closer to the correct answer, and largely showed only proportional rescaling of their preferences. Moreover, Rinne, Ranney, and Lurie (2006) found that shocking numerical feedback on the death rates of various diseases led undergraduates to provide funding allocations that more closely tracked numerical feedback on relative disease mortality rates. Furthermore, Garcia de Osuna, Ranney, and Nelson, (2004) found that surprising feedback led many participants to articulate quite different kinds of rationales for their numerical preferences than they had given prior to receiving numerical feedback.

In the present paper, we extend this body of work by further considering the role of surprising numbers over time. In particular, we focus on the relationship between surprise and *memory* for policy-relevant numbers.

## Numerically Driven Inferencing

The Numerically Driven Inferencing paradigm (NDI; Ranney et al., 2001) considers whether, when, and how numerical evidence can catalyze belief revision. Within the NDI paradigm, Ranney and colleagues have developed various methods, including EPIC (Estimate, Prefer, Incorporate, & Change), in which participants: (1) *estimate* a quantity that is relevant to an issue, (2) state what they *prefer* the quantity to be, (3) receive correct base rate feedback to *incorporate* (e.g., that 507 of every thousand 18-24 year olds are registered to vote), and finally, (4) state what they now prefer the quantity to be—exhibiting whether their preferences *changed* after learning the true number.

This paper again addresses one of NDI's central hypotheses about the mechanism of numerically driven conceptual change: Belief revision is a function of how surprised one is by feedback. That is, receiving a single, surprising number, as in the kinds of discrepant events described above, may lead one to reconfigure networks of hypotheses and evidence to achieve a higher degree of coherence. The overall hypothesis of this paper is that surprise is a rather stable construct that relates to changes in numerical beliefs. Specific hypotheses are as follows:

**Hypothesis 1:** Surprising values highlight incoherence in one's belief network, leading to belief revisions that increase coherence. One source of evidence of this would be that recall accuracy would improve as a function of how surprising a number was to a participant.

**Hypothesis 2:** Surprise is a rather stable construct about which people have considerable metacognition, such that the extent to which participants' pre-feedback expectations deviate from the actual quantities (i.e., "shock," a measure of prospective surprise), is strongly related to how surprised they actually feel upon receiving the value (i.e., retrospective surprise).

Table 1: Questions Given to Students in Experiment 1. True values were the most recent values available when the experiment was conducted in 2003-2004.

| Question | True Value | Retest Time |
|---|---|---|
| **US Legal Immigration per 1000 Residents** | 3 | **8 Days Post-FB** |
| **UC Berkeley Annual Tuition** | $18 K | |
| **College Degrees per 1000 Adults** | 275 | |
| **US Voter Registration per 1000 Young Adults** | 507 | |
| **Toyota Camry Price** | $20 K | |
| **Mean US Sleep per Night** | 6.9 hrs. | |
| **Mean US One-Way Commute** | 25.5 mins. | |
| **Households with TV(s) per 1000 US Households** | 980 | |
| **US Incarceration per 1000 Residents** | 7 | **12 Weeks Post-FB** |
| **Mean US Male Athlete Salary** | $2.5 M | |
| **Inflation: $1K in '62 = __ in '02** | $5.8 K | |
| **Mean Garbage Production per day per US resident** | 4.5 lbs. | |
| **US Cars per 1000 Drivers** | 1.2 K | |
| **Female K-12 Teachers per 1000 Teachers** | 833 | |
| **US Computers per 1000 Households** | 510 | |
| **Mean Non-Diet Soda Calories** | 150 | |

# Experiment 1

As a test of Hypothesis 1, we examined the extent to which participants' surprise upon learning a number predicts the accuracy of their recall of that number.

## Method

Participants were 95 eighth-grade Algebra I students from three consecutive class periods at a San Francisco Bay Area middle school. All 95 received four "common" items (Voter Registration, Immigration, Incarceration, Athlete's Salary), and four of 12 other items that were each given to one-third of the students (see Table 1 for a complete list of items).

Table 2: The Main Steps of Experiment 1's Procedure

| |
|---|
| *Estimate* |
| Out of 1,000 U.S. Citizens between the ages of 18 and 24, estimate the number of people who were **registered** to vote in the presidential election in the year 2000. |
| *Preference* |
| *[Preference regarding what the quantity ought be was elicited; these data are not addressed herein]* |
| *Incorporate Feedback* |
| According to the U.S. Census Bureau, 507 out of every 1,000 U.S. citizens between the ages of 18 and 24 were **registered** to vote in the 2000 presidential election. |
| *Rate Surprise* |
| Rate how surprised you are by this number: 1 = Not At All … 5 = Extremely Surprised |
| *Change Preference?* |
| *[See Preference above]* |
| ***Retention interval: 8 days/12 weeks*** |
| *Recall Feedback* |
| Out of 1,000 U.S. Citizens between the ages of 18 and 24, try your best to remember the number of people who were **registered** to vote in the 2000 presidential election |
| *Change Preference?* |
| *[See Preference above]* |

Two items in the EPIC format were presented for a few minutes each day over a four-day period. (See Table 2 for the procedure; note that it includes all steps, including ones involving preference that are not examined in this paper; see Munnich, Ranney, & Bachman, 2005, for discussion of these steps.) For each item, students first estimated a quantity. Next, they indicated their preferences for the estimated quantity. Participants then received feedback (the true value) to incorporate, after which they indicated on a 1-5 scale how surprised they were by the actual number; the students then had the opportunity to revise their preferences in light of feedback.

Beginning eight days after the first EPIC items were presented, an "RP" extension of EPIC was administered—in which students were asked to *recall* the two numbers they received on the first day and to again indicate their *preferences*. The following (tenth) day, students recalled, and offered preferences for, the two numbers they received on the second EPIC day. Eleven weeks after that, students recalled the feedback for, and offered preferences for, the remaining four items (again, over two days). Thus, four EPIC-RP item sequences were completed after eight days' delay, and four more after 84 days.

We developed a metric of improvement from one's initial estimate to the number one recalled either eight days or 12 weeks after receiving feedback, based on Brown and Siegler's (e.g., 2001) Order of Magnitude Error. Improvement in accuracy was computed as follows[1]:

*EstimationError* =
Estimate / TrueNumber {if Estimate ≥ True}
TrueNumber /Estimate {if True > Estimate}
*RecallError* =
RecalledNumber / TrueNumber {if Recalled ≥ True}
TrueNumber / RecalledNumber {if True > Recalled}
*Improvement in accuracy* = EstimationError / RecallError

## Results and Discussion

As Munnich et al. (2005) reported with regard to the same data set, participants' numerical understandings improved between estimation and recall, over both the eight-day and 12-week intervals. Our Hypothesis 1, though, concerns the extent to which improvement in accuracy (numerical understanding) reflects the surprise that participants reported upon seeing the true numbers.

Table 3: Improvement in Recall, in Descending Order of Surprise Ratings (nb. when Improvement > 1, one's recall is more accurate than was one's initial estimate)

| | Question | Mean Surprise | Mean Improvement[2] |
|---|---|---|---|
| **8- day** | **Immigration** | **3.5** | **18.82[3]** |
| | **Tuition** | **3.3** | **1.19** |
| | **College** | **2.7** | **0.62** |
| | **Vote** | **2.5** | **0.70** |
| | **Camry** | **2.4** | **0.79** |
| | **Sleep** | **2.2** | **1.01** |
| | **Commute** | **2.1** | **0.93** |
| | **TV** | **2** | **0.60** |
| **12- wk** | **Incarceration** | **3.3** | **1.79** |
| | **Athlete** | **3** | **1.34** |
| | **Inflation** | **2.8** | **0.87** |
| | **Garbage** | **2.8** | **1.70** |
| | **Cars/Driver** | **2.7** | **0.45** |
| | **Teacher** | **2.6** | **0.98** |
| | **Computers** | **2.5** | **0.64** |
| | **Soda** | **2.4** | **0.78** |

[1] We computed all statistics using absolute values of log ratios to preserve the pattern of variance, but we present simple ratios here for expository purposes. Estimation and Recall Errors are computed to preserve the *magnitude* of error; in other words, if the estimate is twice the true value, it is treated as equally erroneous as an estimate that is half the true value.

We predicted that surprise would be associated with the long-term recall of numerical feedback, so we tested whether the quantity participants recalled was closer to the feedback number than it was to the original estimate. (See Table 3 for summary data.) A linear model was constructed in which participants and items were treated as random factors. Using restricted maximum likelihood estimation, we computed z-scores for regression coefficients assuming asymptotic normality. Surprise ratings did not reliably interact with delay of recall, so the interaction term was removed from the model. Surprise reliably predicted improvement from estimation to recall ($z = 2.34$, $p = .02$), but delay of recall did not reliably predict improvement ($z = 0.46$, n.s.).

Experiment 1 indicated that the more surprising the feedback on a question, the more participants improved over the retention interval (i.e., how close recall was to the feedback value relative to the proximity of initial estimates to that feedback number). This held for both eight-day and 12-week intervals, and both results are evidence in support of Hypothesis 1.

## Experiment 2

This experiment was performed to further our understanding of the nature of surprise—how surprised one expects to be prior to feedback and how surprised one feels upon receiving feedback. To the extent that there is a relationship between anticipated surprise and the surprise one indicates upon learning the true quantity, that relationship would suggest that surprise is a relatively stable construct, about which people have notable metacognition (Hypothesis 2).

### Method

We analyzed data from our laboratory (McGlothlen, 2003) collected from 14 high school students who were asked to give oral explanations while responding to nine EPIC items (Table 4). For this analysis, we were concerned with the extent to which students' non-surprise intervals and confidence ratings (prospective surprise; Ranney et al., 2001), predicted their surprise ratings after they received the true number (retrospective surprise). To indicate their prospective surprise, after students provided estimates, they were asked 1) how high the true number would have to be to surprise them, 2) how low would the true number would have to be to surprise them, and 3) how confident they were, on a scale from 55% to 95%, that the true number would fall within the non-surprise interval indicated in steps 1 and 2. Based on these steps, we calculated a 1-10 "shock" scale (Ranney et al., 2001): If subjects captured the value in their non-surprise interval, then:

$$\text{Shock} = 11 - ((\text{Confidence percentage} + 5) / 10).$$

If subjects did not capture the actual value in their non-surprise interval, then:

$$\text{Shock} = (\text{Confidence percentage} + 5) / 10$$

Therefore, a "1" on the shock scale corresponds to hardly being surprised at all—being 95% confident that one would capture the number in one's non-surprise interval and successfully capturing the number. "5" and "6" correspond to moderate shock—respectively, 55% confidence and capturing the number and 55% confidence *without* capturing the number. Finally, a shock of "10" corresponds to maximal surprise—that is, having been 95% confident but not capturing the number in one's non-surprise interval.

Table 4: Questions in Experiment 2. True values were the most recent values available when the experiment was conducted in 2003.

| Question | True Value |
|---|---|
| **College Degrees per 1000 Adults** | 275 |
| **US Legal Immigration as % of Total Population** | 0.3% |
| **California Population** | 35.1 M |
| **Mean Age in U.S.** | 36.6 yrs. |
| **Median College Vs. HS Grad Wages** | 1.8 : 1 |
| **Median US Net Worth (< 35 yrs.)** | \$11.6 K |
| **Median US Net Worth (65-74 yrs.)** | \$176 K |
| **Murders per 1 Million Residents** | 55 |
| **Deaths in Auto Crashes per 1 Million Residents** | 148 |

### Results and Discussion

To analyze the results, we constructed a linear model (in which participants and items were treated as random factors) and tested whether students' shock values predicted their surprise ratings post-feedback. Results show this to be the case, as prospective surprise on the shock scale was a reliable predictor of retrospective surprise ($F(1,103) = 12.10$, $p < .001$). Further analyses examined the relationship between shock and surprise across items, subject-by-subject. Nine out of 14 students exhibited correlations of .64 or

[2] It is reasonable to ask why some mean improvements are, in fact, decrements (i.e., <1), indicating that participants performed less accurately for some items, at the recall stage, than they performed originally at estimating. It appears that (a) many students had forgotten less surprising feedback numbers by the time they were retested, and were actually re-estimating rather than recalling, and (b) spontaneous student comments suggested that that students were not trying as hard at the recall stage (when the task had lost some of its novelty). So, their re-estimates would not be expected to be as accurate as their original estimates. That said, note that for all four items in which mean surprise was three ("moderately surprised") or greater—immigration, tuition, incarceration, and athlete salaries—students showed clear improvement at the recall stage. This is consistent with Hypothesis 1, in that when feedback was surprising it improved students' memories for the numbers.

[3] Since mean improvement for the immigration item was considerably greater than for other items, a separate analysis excluded this item—yet still showed that surprise reliably predicted improvement ($z = 2.44$, $p = .02$).

higher (all $p$'s < .03). Two additional students exhibited marginally significant correlations ($r$ = .56, $p$ = .06; $r$ = .49, $p$ = .09), perhaps suggesting that more items would have yielded even more reliable relationships. (If these *were* reliable relationships for the two students, their correlations suggest that their shock values would respectively explain 31% and 24% of the variance of their surprise ratings.) One of the remaining three students consistently captured the true numbers in the student's non-surprise intervals with confidence ratings of 75% and higher, so the individual's nonsignificant correlation may have been due to restricted range. Finally, two students exhibited non-significant correlations (although one was $r$ = .38, $p$ = .16).

In Experiment 2, we found that students' pre-feedback expectations reliably predicted their surprise upon receiving feedback, which lends support to Hypothesis 2: Surprise does appear to be a relatively stable construct, about which people have notable metacognition. Variation in the predictive value of the shock index for individual students' surprise suggests that, although the two scales correlate, they may tap partially dissociable aspects of a broader notion of surprise. In the General Discussion, we will return to the question of how each scale might be independently inform prediction of how well a student will ultimately remember a statistic.

The present findings suggest that people may have reasonably apt metacognition in that prospective surprise involves one's ability to anticipate one's own surprise. An alternative explanation is that offering a non-surprise interval and a confidence rating ahead of time may lead participants to deliberately provide post-feedback surprise ratings that are consistent with values elicited before feedback (i.e., reflecting how surprised they said they would be, not how surprised they actually were). We do not think this is a likely explanation, since the prospective (i.e., non-surprise intervals and confidence) and retrospective surprise questions were phrased quite differently, did not appear on the same page, and were separated by several other questions—so it would not have been obvious to participants that the two measures served the same purpose. Below, we propose follow-up experiments that manipulate the inclusion of both types of surprise measure in *recall* experiments, so as to determine better how independent the measures are.

## General Discussion

The present work provides a new perspective on numerical understanding to the body of work on "hot cognition" (e.g., Thagard, 2006)—the intersection of cognition and emotion. We draw two main conclusions from the present findings: First, there is a reliable relationship between the surprisingness of a single piece of numerical feedback and the accuracy of participants' subsequent recall of that number. Second, individuals seem to have rather good metacognition regarding their surprise response, as projected surprise was a good predictor of retrospective surprise. Ultimately, we hope this work might include physiological measures that may provide neural perspectives on the surprise reaction.

Our findings build on research on learning from discrepant events, given surprise's role among the basic emotions. Unlike fear, surprise is not clearly negative (i.e., not a withdrawal emotion), so one does not remember what was surprising just to avoid an event in the future; indeed, surprise is often enjoyable. Unlike happiness, a clearly positive (i.e., approach) emotion, surprise may provoke us to restructure our world-views, so that we are not surprised again. That is, surprise is not so jarring that we reflexively avoid an event in the future, but just jarring enough for us to redirect our attention and reorganize our conceptual understandings.

This research suggests two directions for future work on the role of surprise in numerical understanding. First, we think it important to turn to a more detailed qualitative analysis of what happens to people's explanations of how they arrive at predictions—both immediately after surprising feedback, as well as over time (Ranney & Thagard, 1988; Ranney, Schank, Mosmann, & Montoya, 1993). The dynamics of the role of surprise in estimating and recalling numbers could also be more finely modeled. Thagard's (2006) HOTCO model employs a connectionist network that incorporates emotion in achieving coherence during argumentation. In HOTCO, beliefs are connected via excitatory and inhibitory links (as with ECHO), and surprise arises following the introduction of new information (as in discrepant-events experiments) or when a discovery is made about the inconsistency of aspects of one's extant network. Surprise then triggers re-weighting of the present links and, as necessary, the introduction of new belief nodes. The present paper shows that surprising feedback can lead to the long-term retention of numbers that are relevant to social issues; understanding the dynamics of how this unfolds presents an important challenge for the future.

A second possible direction for future research is to explore how hindsight bias moderates the feeling of being surprised by new numbers. Hindsight bias (e.g., Hawkins & Hastie, 1990) arises when one has knowledge of an outcome and is overconfident about how well one would have anticipated that outcome earlier. In terms of numerical outcomes, Rinne et al. (2006) found that estimating a disease's mortality rate before feedback increased the influence of the feedback on one's later preference for research funding. Such participants more often explained their decisions by making reference to the actual values as well. An attractive explanation for this increased reliance on feedback numbers is that having just made an estimate, and faced with evidence of one's misunderstanding of the data involved with an issue, the possibility that one would show hindsight bias is reduced. Awareness of our numerical misunderstandings seems to spawn greater surprise at the feedback numbers, and, thus, greater belief revision.

Consider how estimating a number and providing an *a priori* non-surprise interval relates to hindsight bias, and to the surprisingness of numerical feedback. The present paper points to the impact of that surprisingness on the long-term retention of numbers: To the extent that surprise is accentuated when one "puts one's cards on the table" by specifying an estimate, we would expect that estimating before learning a quantity's true value would lead to improved retention of feedback numbers. In this paper, we observed a large, but not complete, overlap between the prospective shock values and the retrospective surprise ratings. In the future, manipulating the asking of prospective and retrospective surprise questions as a part of long-term numerical recall studies may illuminate both shared and separate components of these two scales. It is possible that eliciting a non-surprise interval lessens hindsight bias, thereby accentuating a sense of surprise, which might then drive the long-term retention of numerical feedback.

In this paper, we considered two data sets that, respectively, indicate that numerical surprise 1) has a long-term impact on how accurately people retain numbers, and 2) can be reliably predicted using metacognitive self-assessment. As our project moves forward, we hope to understand better the dynamics of numerical cognition. We also hope to further connect other domains involving discrepant events that contribute to surprise, which increases the coherence of an individual's representations. Findings in this research area should interest both educators and others involved in public policy, as they aim to help people attain more coherent understandings of what numbers mean in the contexts of learning and citizenship.

## Acknowledgements

We thank Mandy Bachman, Jake Disston, Lilian McGlothlen and their students, Myles Crain, Barbara Ditman, Andrew Galpern, Jacob Levernier, Tania Lombrozo, Nick Lurie, Luke Miratrix, Janek Nelson, Patricia Schank, Maureen O'Sullivan, Sophia Rabe-Hesketh, Luke Rinne, and full UCB Reasoning Group for their skills and comments. This work was funded by UCB faculty research grants and an AERA/IES Postdoctoral Fellowship.

## References

Brown, N. & Siegler, R. (2001). Seeds aren't anchors. *Memory & Cognition, 29*(3), 405-412.

Clement, J. & Steinberg, M. (2002). Step-wise evolution of mental models of electric circuits: A "learning-aloud" case study. *Journal of the Learning Sciences 11*(4), 389-452.

Ekman, P. (1992), An argument for basic emotions. *Cognition & Emotion 6*, 169-200.

Garcia de Osuna, J., Ranney, M., & Nelson, J. (2004). Qualitative and quantitative effects of surprise: (Mis)estimates, rationales, and feedback-induced preference changes while considering abortion. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 422-427). Mahwah, NJ: Erlbaum.

Hawkins, S. & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107*(3), 311-327.

McGlothlen, L. (2003). *High school students reasoning with numbers: Interviews of students using the estimate, predict, incorporate, and change (EPIC) method*. Unpublished MA Paper. U.C. Berkeley.

Munnich, E., Ranney, M., & Bachman, M. (2005). The longevities of policy-shifts and memories due to single feedback numbers. In B.G. Bara, L. Barsalou, & M. Bucciarelli (Eds.) *Proceedings of the Twenty-seventh Annual Conference of the Cognitive Science Society* (pp. 1553-1558). Mahwah, NJ: Erlbaum.

Munnich, E., Ranney, M., Nelson, J., Garcia de Osuna, J., & Brazil, N. (2003). Policy shift through Numerically-Driven Inferencing: An EPIC experiment about when base rates matter. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society* (pp. 834-839). Mahwah, NJ: Erlbaum.

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures* (A. Rosin, Trans.). New York: Viking. (Original work published 1975).

Ranney, M., Cheng, F., Nelson, J., and Garcia de Osuna, J. (2001). *Numerically driven inferencing: A new paradigm for examining judgments, decisions, and policies involving base rates.* Paper presented at the Annual Meeting of the Society for Judgment & Decision Making

Ranney, M., & Schank, P. (1998). Toward an integration of the social and the scientific: Observing, modeling, and promoting the explanatory coherence of reasoning. In S. Read & L. Miller (Eds.), *Connectionist models of social reasoning and social behavior* (pp. 245-274). Mahwah, NJ: Erlbaum.

Ranney, M., Schank, P., Mosmann, A., & Montoya, G. (1993). Dynamic explanatory coherence with competing beliefs: Locally coherent reasoning and a proposed treatment. In T.-W. Chan (Ed.), *Proceedings of the International Conference on Computers in Education: Applications of Intelligent Computer Technologies* (pp. 101-106).

Ranney, M., & Thagard, P. (1988). Explanatory coherence and belief revision in naive physics. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 426-432). Hillsdale, NJ: Erlbaum.

Rinne, L., Ranney, M., & Lurie, N. (2006). Estimation as a catalyst for numeracy: Micro-interventions that increase the use of numerical information in decision-making. *Proceedings of the Seventh Annual International Conference of the Learning Sciences.*

Thagard, P. (2006). *Hot thought: Mechanisms and applications of emotional cognition.* Cambridge, MA: MIT.

Wertheimer, M. (1945). *Productive thinking.* Oxford, UK: Harper.