Policy Shift Through Numerically-Driven Inferencing: An EPIC Experiment About When Base Rates Matter

Edward L. Munnich (munnich@uclink.berkeley.edu)[1]
Michael A. Ranney (ranney@cogsci.berkeley.edu)
University of California, Graduate School of Education
4533 Tolman Hall, Berkeley, CA 94720-1670 USA

Janek M. Nelson (jamin@socrates.berkeley.edu)
University of California, Department of Psychology
3510 Tolman Hall, Berkeley, CA 94720-1610 USA

Jennifer M. Garcia de Osuna (jmgdo@uclink.berkeley.edu)
Noli B. Brazil (n_brazil@uclink.berkeley.edu)
University of California, Graduate School of Education
4533 Tolman Hall, Berkeley, CA 94720-1670 USA

Abstract

Drawing on research areas such as estimation, innumeracy, attitude, scientific conceptual change, social cognition, and judgment and decision making, we offer results from a paradigm we call Numerically-Driven Inferencing (Ranney, Cheng, Nelson, & Garcia de Osuna, 2001). NDI includes observing the effects of presenting critical, germane, and credible base rates that are relevant to social policies; such data, we found, can catalyze changes in belief systems. Here, 130 college students first estimated quantities relevant to important policy issues (e.g., abortion rates), then stated preferences for these values. They next received the true values as feedback, and were again asked for their preferences. This EPIC (Estimate, Prefer, Incorporate-feedback, & Change-policy) method helps quantify relationships among one's understandings of base rates and policies. As some have noted, we too found that people are often poor at estimating base rates. Going beyond past research, we further found that many are quite surprised by the true base rates, and readily revise their numerical preferences after receiving them. Preference changes seem surprise-mediated and are often actual *policy shifts* (which go beyond the mere re-scaling of preferences in proportion to the feedback). The shifts suggest that conceptual changes among a network of propositions gave rise to belief revisions. We also found that abortion rates queried in different ways yielded notably different policies and policy changes. EPIC may be used to improve numeracy, so we also discuss an NDI curriculum that engages younger people; it may allow us to further consider how numerical cognition and preference co-develop.

Preferences are central to humans' mental lives, and are main "outputs" of our belief systems. Even novel preferences are largely readily available to us. For instance, one can choose a team to root for at one's very first rugby match; a person can even voice preference on the veracity of evolution (cf. Brem, Ranney & Schindel, in press). In the spirit of novelty, ask yourself this: What is the mean, SAT-I percentile for this year's new undergraduates at your college alma mater? Now, what would you prefer it to be *next* year? Most people have likely never asked themselves such questions, yet answers are readily elicited (Ranney et al., 2001, which the present study partly replicates and partly extends). You might prefer an increase or a decrease in the mean percentile for a variety of reasons (e.g., prestige or diversity) and in general, many propositions might inform our specific and relative social preferences (see Ranney & Schank, 1998). Now, would your preference change if your estimate were well off the true mean percentile? Suppose your estimate were 20% too high—or 5% too low. More broadly, what sort of numerical feedback might you receive that would call some basic assumptions into question, and lead to a different preference than you had originally articulated?

How we prefer, in the presence or absence of base rates, is an area in which cognitive science directly contacts worldly concerns. Imagine legislation geared toward affecting the selectivity of a state's public universities: Unaware of current base rates, a candidate or voter might take a stand that conflicts with what s/he would generally prefer. This was illustrated by Ranney et al.'s (2001) findings of "odd reversals" between pairs of some typical participants. For instance, one person might believe the annual U.S. legal immigration rate to be 20% of the current population, and prefer 15%; another might estimate 2%, and prefer 3%: Here, the "anti-immigration" person is seemingly advocating five times the rate the "pro-immigration" peer preferred! How would these two respond if they learned the true value? We will provide that value later, but for now suggest that people *should* want to know it.

The Numerically-Driven Inferencing (NDI) paradigm, introduced by Ranney et al. (2001), focuses on estimation, numerical preference, and policies involving changes in

[1] The order of the first two authors is alphabetical.

base rates like those above. To address NDI issues, Ranney and colleagues have developed novel methods, such as EPIC, in which people: (1) *Estimate* a quantity relevant to an issue (e.g., U.S. immigration), (2) indicate what they would *Prefer* the value to be, (3) receive the correct base rate as feedback to *Incorporate*, and (4) indicate any *Change* desired in policy by giving their preference again.

Of course, people may neither know nor care about a given base rate—one may want a campus to be more (or less) elite, whatever average SAT-I scores may be. But consider Ranney et al.'s (2001) immigration results: The median estimate for legal immigration was 10%, and the median initial preference was to keep the *status quo* (10%). Using a variety of metrics, it was found that the participants were highly overconfident in their estimates and *quite* surprised to learn that the legal immigration rate was only 0.3%. What policy should one adopt *now*? If one cares little about base rates, "0.3%" should not affect one's preference. However, if base rates have import for judgment—whether or not people commonly know them—EPIC offers a way to assess the degree to which base rates can cause policy shifts.

Our work builds on research from many fields (e.g., attitude, mathematical/scientific conceptual change, social cognition, and judgment and decision making) to which we cannot do justice here. But, clearly, work on base rates and estimation must be mentioned. Base rates have been central to a decision-making debate that focuses on how much people neglect them in making probability and/or frequency judgments (e.g., Tversky & Kahneman, 1974; Gigerenzer, 2000; Nb. Ranney et al., 2001, generally found overconfidence effects even using frequency formats). However, in contrast to this debate, NDI deals more directly with the base rates themselves, and is not (much) concerned with base rates regarding Bayesian analyses. Regarding estimation, NDI is similar to other approaches that solicit real-world estimates (e.g., Brown & Siegler, 2001, Huttenlocher, Hedges, & Prohaska, 1988). Beyond straightforward estimation, Brown and Siegler gave participants correct answers (seeds) for some estimates, to see how future estimates were affected; this work offered a way to quantify aspects of conceptual change, which NDI extends to preference and how that changes. It is notable that, although past estimation research has largely focused on numbers that are explicitly stipulated or cannot readily be changed (e.g., cities' latitudes), NDI considers the many base rates that people *can* have roles in modifying.

This paper (a) describes the paradigm's budding theory, (b) reports our group's latest NDI findings, and (c) proposes a practical extension: a curricular intervention that could offer insights into how policies shift more longitudinally.

What's Behind the Numbers Is What Counts

The Theory of Explanatory Coherence (TEC; e.g., Ranney & Schank, 1998; Thagard, 1989) describes ways in which arguments may or may not cohere. Conceptual change can be spawned by incoherence, as when people discover conflicts among their thoughts and attempt to revise their beliefs to bring about greater global coherence. Ranney and Thagard (1988) illustrated one aspect of TEC's belief revision account with the typical/composite participant, "Hal," who (a) believed that a pendulum bob released at a swing's apex would fall with a lateral (outward) motion, partly because he (b) believed that a child on a playground swing would laterally "fly off" the swing at its apex. Ranney and Thagard argued that a network of beliefs cohered with the generated trajectory prediction (a). Similarly, NDI suggests that one may generate an estimate that is based upon a complex of propositions (e.g., evidence and hypotheses). Since (a) was incorrect, as the bob was later seen to fall straight down, Hal quickly went on to restructure his beliefs—and to seriously (and appropriately) doubt (b)'s veracity. Returning to NDI, Ranney et al. (2001) noted similar revisions on the immigration topic: After receiving the feedback (0.3%), the median participant switched from the status-quo policy to wanting immigration to become *thrice* its current rate (i.e., 1%). This revision seems striking, as the intervention is "merely" the provision of a single number. NDI may, thus, be of interest to those studying learning and conceptual change as well as those who study judgment and decision-making. Similarly, we seek to better understand how numerical feedback can have knowledge-transforming effects, potentially yielding more globally coherent belief systems.

Consider an example used in our present experiment: Participants estimated and stated preferences for the number of legal abortions (which we defined for participants) carried out in the U.S., for every one million live births. (For a sense of the method, we urge the reader to both estimate and give a preference for the queried number.) Next, participants were given the actual number. (All feedback values were derived from reliable sources, which were cited for participants.) With that value in mind, they were again asked for preferences. To the extent that people shift their policies after feedback, we suggest that they are revising their beliefs to move towards greater global coherence with respect to the new evidence. Besides this abortion question, we asked three others that also involved issues (a) that were rather familiar to participants (by their own accounts), and (b) for which clear numerical preferences can be elicited.

What do the elicited numbers mean? One's understanding of a topic—in this paper, abortion, capital punishment, or college admissions—may be thought of as connected ideas that may include personal experiences, media information, religious opinions, and more. When one is asked to *estimate* an abortion rate, one rarely simply tries to remember the value. Instead, a person may activate a mental set of existing epistemic and experiential understandings about abortion that shape the number being generated. The estimate, then, represents the person's "working reality" for the abortion issue. Likewise, we might think of numerical *preference* as a projection of people's beliefs about society on a topic. Thus, a preference represents only the tip of a "reasoning iceberg"—that is, an output from an extensive complex of thought that lies below the surface of overt response. For example, based on one's knowledge—represented as an estimate—one may find one's assumed reality of abortion acceptable and simply reiterate the estimate *as* a preference (i.e., status quo). However, if one is shocked by the actual U.S. abortion rate, one's "reality" is challenged. Such a cognitive conflict may change a person's thinking about an issue, and lead one to conclude that prior reasoning about the issue was incorrect, or at least incomplete.

Due to this reflection, the "bulk" of this iceberg—the rest of one's belief network—may be transfigured by the impact of base rate feedback. The base rate can be thought of as a numerical representation of the "true reality" of an issue. We would, then, expect that one's understanding may undergo cascading shifts due to the feedback, with opinions or positions changing as various ideas and feelings about "what is happening in society" arise. In such cases, we do not expect simply a proportionate re-scaling of initial preferences, but rather a different kind of preference that, in turn, represents a qualitatively different view of the issue. While we lack the space here to address our work on the psychometrics of surprise, we hope it suffices to say that NDI's measures of shock and surprise correlate with the cognitive conflict spawned by the feedback. We would then expect that those most surprised by a rate would yield the most qualitatively changed preferences after feedback. Examining over a dozen topics, Ranney et al. (2001) did indeed find such effects, with those who were technically surprised by the immigration rate *four times* more likely to significantly change their positions than those less surprised.

Hypotheses About Policy and Policy Shift Patterns

We note that data priority, one of TEC's descriptively useful (e.g., Schank & Ranney, 1991) principles, is relevant to our budding theory: Evidence that is critical, germane, and credible should be weighted heavily in our belief systems. For this reason, we employ issue-critical, ideologically-neutral numbers in NDI experiments, and hypothesize that when such information becomes available, to the extent that it is surprising, it should lead to nontrivial belief revision (cf. Ranney, Schank, Mosmann, & Montoya, 1993). For example, here is the (above) abortion question's answer: At the time of the experiment, there were 335,000 legal abortions per million live births in the US. Depending on your estimate, you may find this value surprising, and your numerical preference may become quite different than it was just seconds ago. Indeed, Ranney et al.'s (2001) participants were both rather viscerally shocked by the number—per their written and oral comments—and measurably surprised: While providing estimates, NDI participants generated "non-surprise intervals" ("how high/low would the number have to be to surprise you?") and rated their confidence that the true rate would fall in their intervals. Ranney et al. (2001) found that true value for this question fell inside of participants' intervals only 21% of the time—roughly 3.5 times less often than one would expect based on participants' confidence ratings! This hypothesis, that surprise drives belief revision, gives rise to the following three predicted patterns of responses to base rates:

Pattern 1: On any given issue, some people may take an *absolute stance*. Based on Ranney et al. (2001), we expected that some participants would (likely consistently) prefer the elimination of abortions or executions, regardless of numerical feedback.[2] When rigidly extreme patterns are seen, the respondents seem to tell us that they do not accept the base rate as relevant to their belief systems, so they will not change their preferences after feedback. In connectionist terms, we might think of rigid stances in NDI as indications that some nodes are isolated, encapsulated, or activation-clamped (see Ranney & Schank, 1998).

Pattern 2: Alternatively, participants may sometimes merely *proportionately rescale* their preferences with respect to feedback; if one preferred halving the number of abortions initially, one might still prefer halving the actual abortion rate when it is unveiled. For such respondents, while a base rate is taken into account, it has at best a relatively shallow (proportionate) impact on their beliefs. Unlike Pattern 1, proportionate responses suggest that a rate is *relevant* to one's belief system. But the rate is not surprising enough to require a dramatic accommodation of the extant belief network. Rather, one assimilates the datum, and proportionally re-scales the output of the network.

Pattern 3: Finally, one might show a *policy shift* after feedback. For instance, if a person preferred halving the number of abortions initially, s/he might find the actual magnitude so surprising as to prefer something dramatically different than simply halving the true value after feedback. Such a change suggests a policy shift—a considerable, accommodative, belief revision arising from the feedback.

The following experiment was conducted to better understand which kinds of question-and-feedback items most elicited which of the above patterns from people. Prior research (Ranney et al., 2001) indicated that we should analyze participants categorically, based upon three aspects of their estimates and preferences: Those with a preference of zero on a topic were separated from the rest, as almost all do so rigidly and thus fall in Pattern 1; the rest were then categorized based upon (a) whether they shifted preference and/or policy, and then (b) whether they initially overestimated or underestimated the base rate (since a policy shift's direction will likely differ based on the direction of one's relative surprise). To measure policy shift, we first computed ratios for participants' *initial policies* by dividing their initial preferences by their estimates, and computed ratios for their *final policies* by dividing their final preferences by the feedback values. Finally, we computed a ratio for *policy shift*—the degree to which a policy changed after feedback—defined as the difference between one's *final policy* and *initial policy* as a proportion (here, percentage) of the initial policy. These ratios are summarized as follows:

$$\text{Initial Policy} = \text{Initial Preference} / \text{Estimate}$$

$$\text{Final Policy} = \text{Final Preference} / \text{Actual Number}$$

$$\text{Policy Shift} = \frac{\text{Final Policy} - \text{Initial Policy}}{\text{Initial Policy}}$$

[2] One can, of course, cling rigidly to any number. But, when people indicate that no number of abortions, executions, or murders is acceptable, they provide not only a number, but an *explicitly* absolute stance; thus, we have a unique opportunity to consider whether the stance is, in fact, absolute, or susceptible to feedback.

Note that a proportional re-scaling (e.g., halving initially and finally) yields no policy shift, as (0.5-0.5) / 0.5 equals zero.

Method

One hundred and thirty undergraduates from the University of California, Berkeley, voluntarily participated as part of their Introduction to Cognitive Science course.

Questions appeared on separate pages that students wrote answers on. On the first page, they estimated a value and offered its non-surprise interval; on the second, they gave their initial preference for the value (and various ratings); on the third, they were given the true value, and were asked for their final preference (and ratings). The questions were as follows—but please think of your own estimates and preferences for each item before reading the next section:

1a. What is your best estimate of the current number of legal abortions, per 1,000,000 live births in the U.S.? ____abortions. [Subset of participants]

1b. What is your best estimate of the number of legal abortions performed, per 1,000,000 fertile U.S. women (aged 15-44) for a single year? ____abortions. [Subset of participants]

2. For the past few years in the U.S., how many executions do you think were performed, compared to the number of murders committed? ____executed prisoners to ____murders.

3. What is your best estimate of the current, average SAT I percentile (from 1% as low to 99% as high) of undergraduates admitted to U.C. Berkeley from high school? ____percentile

Participants received either Question 1a or 1b, and both Questions 2 and 3, in varying orders. Pages with *estimate* and *initial preference* queries were given first, and the page with *feedback* and the *final preference* query came later. Participants were instructed not to return to the estimate and initial preference pages once they had finished them.

Results

Estimation Analyses

Overall, participants often showed systematic inaccuracies in estimating the mean SAT-I percentile and rates of abortions and executions (Table 1). For the abortions-per-births variant (n=28), the median estimate was 10,000 abortions per million live births—thirty[3] times lower than the actual value, 335,000. In fact, *all* students underestimated the rate. For the abortions-per-women variant (n=53)[4], the median estimate was (coincidentally) 10,000 abortions per million fertile women, while the actual number was 20,000.

For the capital punishment question (n=99) students generally overestimated the number of executions relative to murders: The median estimate was one execution per 50 murders, while the actual rate was roughly one per 250.

Finally, for the SAT-I question, students systematically underestimated the mean percentile of applicants admitted to their own university: The median estimate was the 85th percentile, while the true value was the 94th percentile.[5]

Policy Shift Analyses

With these (mis)estimations in mind, we turn to whether and how policies shifted after feedback. Since ratios can distort sample variance, we carried out nonparametric (Wilcoxon Signed Rank) tests between initial and final policies.

For the abortions-per-births variant, of the 28 who answered all of the item's parts, five consistently preferred zero abortions. This is in keeping with the prediction that those who initially prefer no abortions already have a stable policy (elimination), and would generally not shift it due to the feedback. The 23 "nonzero" students exhibited a reliable overall policy shift: from preferring an initial reduction in abortions to an even greater proportionate reduction after feedback (z = -2.973, p = .001; see Table 1 for median estimates, initial policies, and policy shifts after feedback).[6]

For the abortions-per-fertile-women variant, 11 of 53 responders initially preferred no abortions, and nine of them did so finally, too. This lack of change is again consistent with the predicted Pattern 1: Elimination responses are stable.[7] Of the 42 participants who gave nonzero preferences throughout, there was no overall policy shift (z = 0.3; Table 1). That is, as a group they (a) were significantly more accurate initially than those receiving the prior variant, and (b) advocated roughly the same (abortion reducing) policy before and after feedback.[8]

For the capital punishment topic, 33 of the 99 students offered an absolute response for either their initial or final preference, that is, of either no murders or no executions. Of the 33, 31 offered zero as their initial preference. Six of the 31 switched to a nonzero response, whereas 25 still

---

[3] With a larger sample, Ranney et al. (2001) reported a median student estimate of 5,000 for this query (i.e., off by a factor of 67).

[4] Participant totals do equal 130, as some received alternate questions that we cannot address, given space constraints; the results for other items were similar to those presented here. In addition, a small set of participants gave non-numerical responses, which are beyond our present inquiry's scope. Finally, a small set of participants did not complete all relevant parts for a given topic.

[5] Ranney et al. (2001) offered an account for this percentile gap.

[6] Eight participants actually preferred the same nonzero number of abortions at both times. It is plausible that these participants' policies only appeared to shift because their estimates diverged from actual values. We return to this in the Discussion. Even so, when considered separately, the 15 people who did shift preferences also reliably shifted policies (z = -1.922, p = .03; Table 1).

[7] Nb. The elimination policy is embraced by "strange bedfellows;" e.g., some want to ban abortion, while others want perfect birth control (Ranney et al., 2001); we would expect neither to be swayed from their categorical stance by numerical feedback.

[8] Of these 42, 17 did not show a change in absolute preference, so it may again be odd to think that they changed policies. However, even for the 25 who changed preferences, there was no reliable policy shift (z = 0.4; Table 1). On this abortion variant, there was a relatively even split between those underestimating (15) and overestimating the rate (9). Underestimators did not exhibit a policy shift (z = -0.6; Table 1), however overestimators exhibited a marginally significant policy shift (z = 1.4, p = .08; Table 1).

Table 1. A breakdown of initial estimates, feedback values and degree of policy shift by item and category of participant

Question	Participants	N	Median Estimate	Median Initial Policy	Actual Rate	Median Policy Shift (From Initial Policy)
Abortions per 1 Million Live Births	All Responses	28	10,000			
	All nonzero responses	23	10,000	0.50		-64% **
	Preference changers	15	10,000	1.00	***335,000***	-50% *
	Underestimators	15	10,000	1.00		-50% *
	Overestimators	0	--	--		--
Abortions per 1 Million Fertile Women	All Responses	53	10,000			
	All nonzero responses	42	10,000	0.29		0%
	Preference changers	25	5,500	0.38	***20,000***	0%
	Underestimators	15	2,000	0.33		0%
	Overestimators	9	200,000	0.38		+100% #
Executions per 1000* Murders (*scaled from the two-blank response form)	All Responses	99	20			
	All nonzero responses	66	20	4.44		+53% *
	Preference changers	43	14.3	4.00	***4***	0%
	Underestimators	13	1	20.00		-88% **
	Overestimators	28	50	2.00		+27% *
SAT-I Percentile of UC-Berkeley Undergrads	All Responses	120	85th			
	All nonzero responses	120	85th	1.00		-3% **
	Preference changers	92	85th	1.00	***94th***	-1% **
	Underestimators	83	80th	1.00		-2% **
	Overestimators	8	96th	0.98		+2% *

Note: ** denotes p<.01, * denotes p<.05, # denotes p=.08 (Wilcoxon Signed Rank tests).

preferred zero, and were joined by two who had given nonzero responses initially. These results are again largely consistent with our predicted Pattern 1: The vast majority of "absolute preferers" start with a stable policy, and are rarely swayed by feedback. The remaining 66 students exhibited a significant overall policy shift, initially preferring more than four times as many executions relative to murders (median initial policy = 4.44), and, after feedback, showing a policy shift towards an even higher relative number of executions (z = 2.016, p = .02; Table 1). As was the case for the prior topics, there was a minority of students (n=23) who offered nonzero preferences and kept these same preferences after feedback. An analysis for the 43 who did change preference shows no reliable policy shift (z = -0.5; Table 1). However (as with the abortions/fertile-women variant), there was a relatively even split between those underestimating (n=13) and overestimating (n=28) the base rate, suggesting that the two sets of people might have shifted policy in different ways. This was in fact the case, as we found a reliable policy shift for each group (underestimators: z = -2.652, p = .004; overestimators: z = 2.107, p = .02; see Table 1). To unpack this divergence, note that both groups preferred more executions relative to murders before feedback. Underestimators, upon learning that the base rate was closer to their preferences than they had thought, adopted a new policy calling for a smaller increase in executions than they had initially. In contrast, the overestimators, realizing that the value was further from their preferences (usually much lower) than they had thought, adopted a policy calling for an even greater relative increase in executions (see Table 1). In other words, both groups underwent net policy *shifts*–not the mere rescaling of preferences—in response to feedback.

Finally, results for the university admissions topic were similarly clear-cut: None of the 120 respondents preferred zero, which would be nonsensical, as there is no 0^{th} (nor 100^{th}) percentile. Participants' overall policy shift was reliable (z = -5.2, p < .001; Table 1), and although there were 28 students who kept the same preference throughout, the remaining 92 still showed a significant policy shift from the status quo to a (slightly) lower-than-actual preference (z = -3.5, p < .001; Table 1). Underestimators shifted from preferring a roughly status quo policy to a significantly lower-than-actual value (z = –4.1, p < .001). Overestimators were rare (n=8), but after initially preferring a slight decrease in perceived percentile, they preferred roughly the status quo after feedback (i.e., a significant policy shift: z = 1.8, p = .04). So, from this and the prior topic, we find particularly direct support for the prediction of surprise-mediated belief revision, as the direction of participants' surprises determined the direction of their policy shifts.

Discussion

This experiment provides broad and further support for the central hypothesis that surprising base rates can catalyze policy change (i.e., Pattern 3), since we noted statistically reliable shifts for three of our four topics. In other words, numerical information can indeed carry notable weight (cf. TEC's data priority principle) and lead to accommodative belief revision: When our participants were surprised by feedback, as was common with the abortions-per-live-births, capital punishment, and college admissions item, they produced preference changes that suggested deeper changes in their belief systems. Regarding the abortions-per-fertile-women variant, participants were largely more accurate and less surprised by feedback (e.g., the abortion variants' base rates were twice vs. 33.5-times the median estimates), and thus seemed to have assimilated the datum, largely only proportionately rescaling their preferences (Pattern 2)[9].

[9] This "per-fertile-women" variant was used because of Ranney et al.'s (2001) speculation that people so drastically underestimate on

Further, of those who offered preferences of zero (on either an abortion variant or the capital punishment topic), the vast majority retained this preference after feedback (Pattern 1), as such topics sometimes engage an absolute stance.

A Question That Helped Spawn Curricular EPIC

Why did some participants remain wedded to a preference after receiving a statistic that was contrary to their initial estimates? Beyond the possibility of "economic set-point" types of preferences, we conjecture that there may not have been enough time in this experiment for people to get used to the novelty of giving or modifying preferences. Were the experiment spread out over a semester, participants' greater familiarity with such questions might lead them to develop improved estimation skills that could transfer to estimating even more novel entities. Such a deeper, more intense, intervention might translate into less need to shift policies later. Moreover, as they respond to similar questions that draw on slightly different statistics (e.g., the two abortion variants in the present study), students may begin to show more consistent policies across those questions (e.g., if the number of abortions should be reduced by a particular amount for one variant, they would advocate a similar reduction for the other variant).

Based on our experiments, we are implementing secondary-school interventions with the goals of preparing young people to (1) seek out numbers relevant to issues they care about, (2) view each estimation opportunity from multiple perspectives, and (3) coherently integrate their data into preferences. We also hope that this curricular project will yield important longitudinal evidence regarding our developing theory, including a better understanding of those people who maintain the same nonzero preference, even in the face of rather surprising feedback.

Conclusions

In a democratic society, it seems critical that people make informed policy decisions, and this study shows that people can place considerable weight on base rates as their understandings evolve. It concerns us, though, that even the relatively well-educated and more numerate of our population: (a) have strong beliefs about some important societal issues, yet often little idea of the issue-critical base rates, and (b) exhibit considerable policy malleability across question wordings. Still, we were encouraged by our findings that when accurate base rates are received, they can be used productively in transforming our policies. Such findings yield promising implications for instruction that span from mathematics to civics.

the per-births question, partly because they tend to overestimate how often women are pregnant. This plausibly accounts for the divergent responses we elicited by the two questions: (a) Participants' estimates for the per-births variant were less accurate (and more surprising) than those for the per-fertile-women variant. (b) For the per-births variant, but not per-fertile-women variant, a significant overall policy shift was observed; this was presumably due to greater surprise—per-births participants captured the actual value in their non-surprise intervals reliably less often than did per-fertile-women participants (35% vs. 64%; $X^2_{(1)} = 5.2$, $p = .02$). Thus, even when people consider the same issue, they can arrive at notably different estimates and policies depending on how the same basic query is worded (cf. Schwarz, 1999).

Acknowledgments

We thank Laura Germine, Franz Cheng, Nick Lurie, Christine Diehl, Anna Thanukos, Patti Schank, Florian Kaiser, Lije Millgram, Ragnar Steingrimsson, Michelle Million, and the UCB Reasoning Group. This work was funded by a UCB faculty research grant, an AERA/IES Postdoctoral Fellowship, and an NSF graduate training grant.

References

Brem, S., Ranney, M., & Schindel, J. (in press). The perceived consequences of evolution: College students perceive negative personal and social impact in evolutionary theory. *Science Education*.

Brown, N., & Siegler, R. (2001). Seeds aren't anchors. *Memory & Cognition, 49,* 405-412.

Gigerenzer, G. (2000). *Adaptive thinking: Rationality in the real world.* New York: Oxford University Press.

Huttenlocher, J., Hedges, L.V., & Prohaska, V. (1988). Hierarchical organization in ordered domains: Estimating the dates of events. *Psychological Review, 95*, 471-488.

Ranney, M., Cheng, F., Nelson, J., & 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 Technologie* (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.

Schank, P., & Ranney, M. (1991). The psychological fidelity of ECHO: Modeling an experimental study of explanatory coherence. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 892-897). Hillsdale, NJ: Erlbaum.

Schwarz, N. (1999). How the questions shape the answers. *American Psychologist, 54*, 93-105.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12*, 435-502.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-31.