Known Knowns and Unknown Knowns:
Multiple Memory Routes to Improved Numerical Estimation

Dav Clark, Psychology Department, University of California, Berkeley, davclark@berkeley.edu
Michael Andrew Ranney, Graduate School of Education, University of California, Berkeley, ranney@berkeley.edu

Abstract: Conceptual change represents a crucial, challenging, learning component. This study hypothesized and observed evidence for two parallel forms of learning within the Numerically-Driven Inferencing (NDI) paradigm's rather minimalist intervention of providing direct feedback regarding a numerical estimate—feedback that yields remarkably robust cognitive alterations. The present experiment probed the nature of learning apropos recall or estimation improvements observed after participants (a) provided estimates, (b) received feedback, and (c) re-estimated after waiting for one day. The results show that improved estimation/recall was predicted by two independent elements—surprise at feedback and an explicit sense of episodic recall upon testing. This suggests at least two learning processes: (1) an explicit (though perhaps approximate) recollection of a quantity's magnitude and (2) a non-episodic semantic restructuring that correlates with surprise. Thus, even for concise, factual information, we educators might consider students' "unknown knowns"— knowledge that learners gain without any explicit understanding that they have done so.

Introduction

Various streams of cognitive research suggest that significant conceptual change is difficult to effect (Chi, 2005; diSessa & Sherin, 1998; Lord, Ross, & Lepper, 1979). Numerically-Driven Inferencing (NDI) procedures, however, have yielded notable and encouragingly long-lasting levels of conceptual change with quite minimalist interventions (e.g., by providing estimators with a single, critical, highly germane, feedback statistic; cf. Rinne, Ranney & Lurie, 2006). The EPIC procedure represents one such intervention that is relatively compact and well specified (EPIC and NDI were introduced by Ranney, Cheng, Garcia de Osuna, & Nelson, 2001). Notably, EPIC has been shown to induce long-lasting conceptual change (e.g., Ranney et al., 2008), as evidenced by increased accuracy on estimations up to 12 weeks after the procedure (Munnich, Ranney & Bachman, 2005). If one can determine the cognitive factors that drive the efficacy of this particular curricular intervention, one might then be able to target these factors in developing and refining other pedagogical strategies.

In the EPIC procedure, participants engage with real-world numerical facts that bear on a societal issue, such as abortion, criminal justice, the environment, etc. (e.g., Garcia de Osuna, Ranney, & Nelson, 2004; Munnich, Ranney, Nelson, Garcia de Osuna, & Brazil, 2003). An example item/quantity is "the ratio of murders committed to prisoners executed in the U.S." People often poorly estimate these quantities, and thus the true values are surprising to many individuals. For the example above, individuals estimated anywhere from *two* to *a million* murders per execution (Munnich et al., 2003). (This broad range of innumeracy might be compared to order-of-magnitude errors made in reasoning about physics problems, although crime and physics likely diverge in terms of the complexity of, and one's familiarity with, their respective causalities.) Such starkly diverging estimators might have quite different beliefs regarding the effectiveness of capital punishment as a deterrent to murder (given that deterrence is commonly used in reasoning by both pro- and anti-execution participants).

Successful classroom curricula have been developed for improving students' estimation abilities as part of our laboratory's research on NDI—both in the contexts of high school mathematics and elite graduate journalism classrooms (Munnich, Ranney, & Appel, 2004; Ranney et al., 2008). The EPIC procedure is one component of these curricula in which participants (1) provide an **Estimate** for each policy-relevant item, (2) state what they would **Prefer** each quantity to be, (3) receive actual quantities as feedback to **Incorporate** (as new "**Information")**, and (4) indicate whether their preferences have **Changed** upon receiving feedback. Regarding change, our laboratory found that, after learning that about 250 murders occur per executed U.S. prisoner, participants significantly changed their preferences about the ratio (Munnich et al., 2003).

A fundamental question in cognition concerns the nature of what is learned. Some well-established psychological learning and memory models (Nadel & Moscovitch, 1997) would predict that changes in estimation accuracy must ultimately be mediated by episodic memory (which could give rise to abstracted semantic memories via a process of consolidation). In this case, we would expect participants' explicit memory for feedback (the "I" in EPIC) to correlate well with improvements in estimation accuracy at subsequent testing. Recent evidence suggests, however, that the re-modeling of existing conceptual structures may not depend on episodic memory formation (Tse et al., 2007). Indeed, there is broad evidence that learning may often be subserved by multiple memory systems, perhaps acting in parallel (Clark & Ivry, in press). In this case, we might expect increases in estimation accuracy even when participants report no memory whatsoever for the quantity provided as feedback—particularly if participants had pre-existing knowledge to support such learning.

Surprise upon receiving feedback provides evidence for pre-existing knowledge—specifically an incorrect prior expectation regarding the true value. Thus, surprise may correlate with the kind of non-episodic re-modeling described above. However, surprise may also reflect the emotional impact of the information (Munnich, Ranney & Song, 2007; Thagard, 2006). Therefore, it is important to assess not only surprise, but also whether the surprise had an emotional or intense character. In other words, surprise may mediate improved episodic memory, or it may indicate the existence of prior knowledge that may facilitate changes in semantic memory. Of course, these routes to improved estimation may operate partly or wholly in parallel.

While the hypotheses that follow are certainly not exhaustive, they constitute a set of potential reasons for the efficacy of the EPIC procedure. Most generally, engaging in estimation with feedback may result in a general increase in one's estimation ability (as Ranney et al., 2008, observed following a one-week, EPIC-related/inspired curriculum); such an effect would be seen even for items that received no feedback. We additionally considered that learning would be driven by surprise (e.g., Munnich et al., 2005). Moreover, improvements in estimation could be driven by a direct (potentially approximate) episodic memory of feedback. If *all* improvement were driven by episodic memory, though, we would expect little extra power in a model that included both surprise and episodic memory ratings beyond the power of a model that only included one of these. If, however, a model including both surprise and memory provides a significantly better fit, this offers some indirect evidence that at least a portion of improvement is occurring through a non-episodic route.

We additionally considered the optimal timing for feedback: Immediate engagement with feedback following the generation of an incorrect estimation may (a) interfere with the successful learning of the new, correct, immediately provided information. However, another possibility regarding immediate feedback is that it may (b) help encourage an individual to both engage with and correct her beliefs. Thus, if we were to find that participants improve more with *delayed* feedback than immediate feedback, hypothesis (a) would be supported. Alternatively, if performance were superior with immediate feedback, (b) would seem more plausible.

Experimental Methods

The following experiment was designed to assess whether estimative improvement occurs even with respect to items for which no feedback was received—as was found in curricular NDI studies (e.g., Munnich et al., 2004; Ranney et al., 2008). The experiment addresses (1) the effects of the timing of feedback on subsequent improvements in numerical estimation—as well as (2) whether these improvements are necessarily mediated by episodic memory. A subset of the EPIC procedure was used to explore these issues; participants engaged only in estimation ("E") and feedback ("I"), leaving aside personal preference ("P" and "C").

Participants

Twelve people (seven female) participated (and 19 participated in a later experiment that replicated our main findings), including University of California, Berkeley, undergraduates and members of the general public recruited via online recruitment systems (RPP and RSVP). They received either course credit or $20 for their participation in two one-hour sessions over two consecutive days. Ages ranged from 18-56 years.

Materials

Numerical facts (106 of them) were selected from Ranney et al.'s (2008) collection (See the introduction for an example). Three statistical facts were set aside for the basis of example items (namely US population, world population, and US Gross National Income). Items ranged over a number of topics, and included politics, population dynamics, economics, the environment, education, crime etc. Most items were expressed in percentage form, with the rest being counts of dollars, people, events, or things. For numbers above 999, a comma was used, as in "13,600." For numbers in the millions, billions, or trillions, the appropriate word was used to indicate the order of magnitude (e.g., "300 million"). This was intended to minimize possible confusions about the exact value of the number.

Procedure

Custom software utilizing Vision Egg (Straw, 2008) presented all materials and collected responses (source code available upon request). Descriptions of numerical facts were presented in four or fewer lines of text (with fewer than 55 characters per line). A prompt for numeric entry was located below the description. Feedback concerning the veridical value was provided in a third location, between the description and the text-entry area.
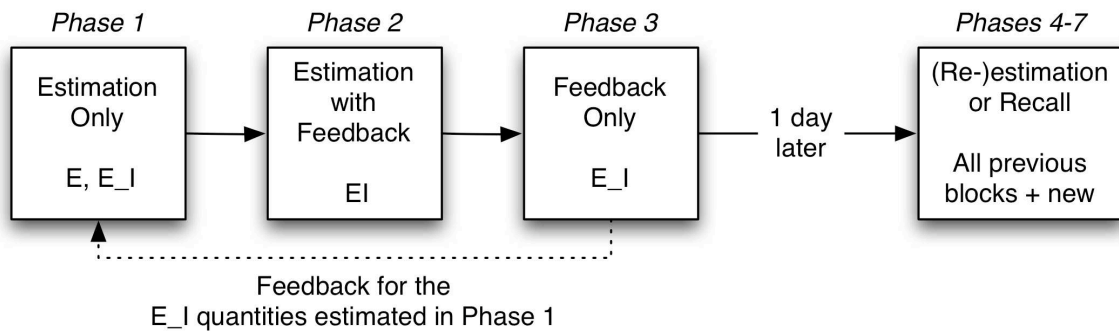
Blocks of Items

Items were randomly distributed into the following four kinds of blocks. Each of these blocks was involved in two or more phases over the course of the experiment. **E**: Participants *only* provided **Estimates** in a single phase. **EI**: Participants provided **Estimates,** followed *immediately* by correct numerical **Information** as feedback (i.e., feedback was provided in the same phase as the initial estimation). **E_I**: Participants provided **Estimates**, and then received correct numerical **Information** in a phase that was *well-separated* (but by less

than 45 minutes) from the phase in which they provided their Estimate (i.e., "_" signifies a temporal delay). **New**: A block of items was reserved to provide a gauge of false recognition or false recollection.

Experimental Phases

Participants engaged in a number of self-paced phases on each of the two consecutive days, as Figure 1 depicts. The structure of stimulus presentation and response collection was uniform across a given phase. During the first day, analogous to a "study" period, participants completed three partially similar phases of numerical estimation and/or informative feedback. (See Phases 1-3 in Figure 1.) The second day was analogous to a "test" period, in which participants' learning was assessed (Phases 4-7 in Figure 1).

Figure 1. A schematic of the experiment's seven phases. *Phase 1*: Estimates were obtained for the E and E_I blocks of items (23 each), randomly intermixed in one phase. *Phase 2*: Participants provided 23 Estimates that were immediately followed by Informing the participant of the correct value. *Phase 3*: Feedback (I) was provided for the 23 items from the E_I block that had been Estimated in Phase 1. *Phases 4-7*: Subjects estimated (or recalled) quantities and provided explicit memory ratings for all previous items as well as 34 new items.

During estimation (Phases 1 and 2, with 23 items each), subjects were given a textual description of an item's quantity, followed by a prompt to provide an estimate. For Phase 2, feedback was provided 500 milliseconds after each estimate was entered. For Phase 3 (with 23 items), the correct numerical value was provided *prior* to the textual description in order to minimize covert estimation.

In Phases 2 and 3 (thus, for blocks including "I"), participants provided a self-report on their subjective sense of surprise. Their responses were restricted to the following three choices: (1) Little or no surprise, (2) Genuine surprise, or (3) "Visceral" or intense surprise. The presence of any form of surprise is an indication of pre-existing memory structures, while the distinction between levels (2) and (3) may indicate a difference in the emotional impact of the feedback for that item. (Note that Munnich, Ranney, & Song, 2007, found that a prospective measure of surprise reliably predicted the sort of retrospective surprise ratings solicited here.)

On Day 2, trials were similar to the estimation-only trials in Phase 1 described above in that no additional feedback was provided. An additional 34 items from the "new" block were randomly intermixed with the items presented during study. Furthermore, participants rated their memory for the item according to the following four levels: (1) "The item is new to me," (2) "The item was presented yesterday, but I have no sense of the value provided as feedback," (3) "The item was presented yesterday, and I have some sense of the correct value," or (4) "The item was presented yesterday, and I have a fairly accurate recollection of the value."

Choice 1 indicates no recognition or recollection. This is equivalent to labeling the item as "new," and it is the correct response for items from the new block. As a group, choices 2-4 indicate that the item is "old," but with varying levels of familiarity and/or recall. These are correct responses for the E, EI and E_I blocks (although choices 3 and 4 entail a belief that the participant actually received feedback at study, and so might also be considered incorrect for the E block). Choices 2 and 3 indicate perceived recognition, but at least a partial failure in explicit recall. Choice 4 indicates a subjective sense of fairly complete recall.

Explicit recall is used herein in a somewhat different way than in most learning and memory studies. Indeed, these memory ratings can be viewed as a form of metacognition regarding the estimation process.

Analysis

We modeled improvement as a binomial outcome (as did Munnich et al., 2005). This allows for the treatment of items that have differing distributions within a unified framework. (It would be difficult modeling both percentages and values in, say, the billions, particularly given our sample size.) Items were labeled as to whether estimates improved or not. These labels were fit with a binomial generalized linear model, using the lme4 package in R. This treatment allows for a full multi-factorial mixed-effects analysis. Below, participants are

always included as a random effect, and other factors are treated as fixed effects. Contrasts were evaluated using the multcomp package, which controls for family-wise error rate (Hothorn, Bretz, & Westfall, 2008).

Unless otherwise noted, data were pre-processed to remove ties. This was done to allow for a null hypothesis that 50% of the remaining items randomly improved and 50% randomly worsened. If we counted ties as failures to improve, then random drift would end up spuriously suggesting the lack of an effect. Removing ties allowed for tests of whether estimates, on average, improved more than they worsened—both formally and when examining graphs. Otherwise, the removal had little effect on the results.

Results

Improvements in Participants' Estimation Accuracy and/or Numerical Recall

Figure 2 shows the proportions for the number of items showing improvement (to any degree), by condition. We can easily reject a null model (i.e., with all conditions modeled by the same mean) in favor of a model including the three feedback conditions ($\chi^2(2) = 25.9$, $p < 10^{-6}$). Post-hoc comparisons between each condition and chance levels, as well as between condition comparisons (as in a Tukey HSD test) were performed simultaneously. In the no-feedback condition (E), estimation improvement did not differ significantly from chance ($p = .39$), although improvement with Immediate (EI) and Delayed (E_I) feedback were clearly above chance ($p < 10^{-4}$). This may seem quite expected, but it *might* have been the case that improvements were at least partially driven by general improvements in estimation skill, and this would have led to at least some modest improvements even without feedback on test items. Indeed, this kind of estimation skill development was the successfully accomplished goal of various EPIC-based curricula (e.g., Munnich et al., 2004; Ranney et al., 2008). In the present, less extensive, experimental manipulation, though, we understandably elicit no such skill improvements. Thus, we assume that these improvements are driven almost entirely by item-specific learning.


Figure 2. Proportion of items from each block for which estimates improved on Day 2 (as compared to estimates on Day 1). Items with identical estimates on Day 1 and Day 2 have been removed; thus chance is .5. Asterisks note significant differences.

Contrary to part of our hypotheses, participants exhibited essentially equal proportions of estimation improvements for the EI and E_I blocks. While these treatments yielded significantly more improvement than for the no-feedback case (both p's $< 10^{-4}$), they did not differ significantly from each other ($p = .79$).

Recollection and Surprise

As is often the case, the participants' forced familiarity judgments appeared to be superior to their own assessments of their memory (Merikle, 2007). In participant debriefings, several individuals claimed to be uncertain as to whether items were old even from Phase 1 to Phase 3 for items in the E_I block—over an interval of less than 45 minutes! However, participants performed well at discriminating between old and new items a day later when given a forced choice; 76% of new items were identified as new on Day 2, compared to an average of less than 9% regarding previously seen items. This level of recognition accuracy is not surprising, given the considerable depth of processing involved—and the rich, pre-existing, memory structures available for scaffolding these episodes. (This may be another point of departure regarding items such as the likelihood of a murder ending in an execution, in contrast to the pre-existing memory structures K-16 students might hold regarding the physics of mechanics or electromagnetism. In both cases, though, it seems probable that multiple conceptual changes may result—e.g., "Wow, the high murders-per-execution ratio makes me think [a] that gangsters would rarely fear the death penalty, given [b] how rarely their fellow gang members are executed and [c] that death rows are basically blossoming geriatric wards, meaning [d] that gangsters would be more likely to bring guns." Such conceptual changes do not seem wildly different, in many important respects, than those required to understand that ballistic trajectories are curvilinear, rather than diagonally rectilinear.)
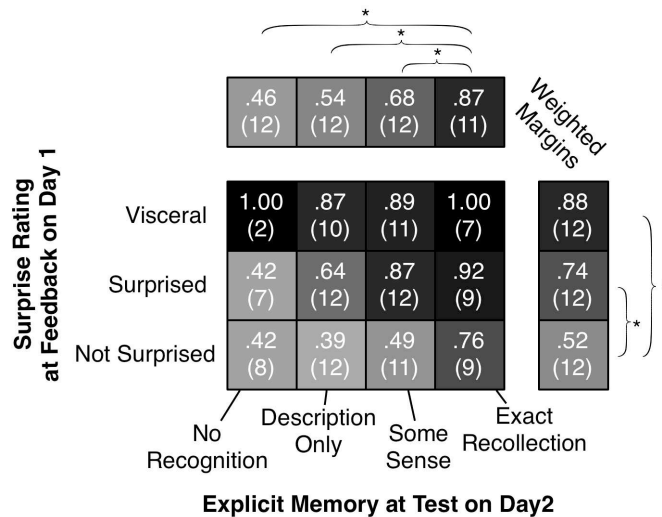
Figure 3. Proportions of items with improved estimation accuracy (or recall) on Day 2. The effects of surprise (Day 1) and explicit memory rating (Day 2) are shown independently via weighted margins, and (the lack of) an interaction via the central table. Asterisks note individual factors' significant differences; significant interactions were not observed. Only items with specific feedback are included (i.e., EI and E_I conditions). Proportions represent between-subject means, and the number of subjects contributing to each cell is shown in parentheses.

As the no-feedback condition (E) yielded non-significant changes in estimation accuracy, we hereafter consider only items from conditions including feedback ("I"; i.e., EI and E_I). These effects are depicted in Figure 3. We find that a model with both surprise and declarative memory responses cannot be rejected in favor of a reduced model excluding memory ($\chi^2(3) = 34.8$, $p < 10^{-7}$), and the comparison involving the removal of surprise is even more striking ($\chi^2(2) = 295.22$, $p < 10^{-16}$). An inclusion of an interaction term does not yield a significantly superior model ($\chi^2(6) = 2.85$, $p = .8$), nor does the inclusion of the feedback condition (EI vs. E_I; $\chi^2(1) = .8$, $p = .36$). There was a small but non-significant difference between surprise ratings for EI and E_I blocks: subjects rated 65% of EI block items as surprising ("2" or "3") vs. 59% for E_I. No straightforward trend was observed with explicit recollection. The timing of feedback may have some effect on estimation improvement that is mediated by surprise, but such issues seem best addressed in subsequent research. Below, we only consider comparisons within surprise level and memory ratings independent from one another.

"Exact recall" for the value of a given item (i.e., memory response "4")—as compared to other memorial impressions—was a highly significant predictor of improved estimation (p's < .001 for ratings "1" and "2;" $p = .01$ when compared with response "3"). No other comparisons between memory levels were significant. For surprise, items with moderate and visceral ratings were significantly more likely to exhibit improvement than those with a no-surprise rating ($p < .002$ in both cases), but did not differ significantly from one another. According to our earlier logic, this suggests more of an effect of prior knowledge than emotional impact, although emotional impacts may still play nontrivial roles. Note that participants provided the exact numerical figure given as feedback only 35% of the time when selecting choice 4. Even if we broaden this accuracy criterion liberally to items for which participants are within 15% of the true value, they were only correct about 74% of the time.

Finally, if we consider the link between surprise and explicit recall, there seemed to have been little relationship in the present study (as opposed to some other studies). The correlation of fixed effects between memory and surprise terms in our model was consistently smaller in magnitude than .1. This, combined with the lack of a significant interaction term, provides some evidence for independent learning processes.

Exclusions

As many as three items lacked estimates from some subjects or exhibited a clear lack of understanding in a response (e.g., a "10 million" response for a question soliciting a percentage) and these items were excluded from the analyses above. Due to a technical issue, one participant did not receive the standard E manipulation, but was included in memory and surprise-related analyses, as these analyses did not include E trials.

Discussion

Given the overall improvements in estimation ability evidenced in curricular studies by Munnich et al. (2004) and Ranney et al. (2008), it is of interest that we see no statistically significant improvement in items that didn't receive feedback (the "E" block). Nonetheless, it seems that learning in this considerably shorter present

experiment was largely item-specific and related to the integration of feedback. This lack of improvement in the present experiment may be due to a lack of time for reflection or the development of strategies—which were highlighted, taught, and fostered in the curricular studies. (Munnich et al., 2004, and Ranney et al., 2008, also focused, to a fair degree, on preferences and personalized policies.)

In addition, we had originally hypothesized that the EI block would yield the greatest overall improvement in estimation. We assumed that immediate feedback would be more memorable because participants would have the opportunity to directly compare their estimate with the actual value and consider its validity. This would have arguably led to a more tangible interaction with the true value than in the E_I condition. However, our data show that the E_I block, if anything yielded larger improvements. (Although EI and E_I conditions did not significantly differ, this pattern has held up in subsequent experiments as well.) This may be due to a number of reasons, warranting further experimental inquiry.

One possibility is that we are observing something akin to a distributed learning effect—because even though the feedback is provided only once, the size of the effect is reminiscent of the effects of well-spaced study periods reported by Pashler et al. (2007). Note that this is a major departure from standard spacing effects, as the correct information is presented *only once*. However, while EI items were only shown once before testing, item descriptions in the E_I condition were shown twice—initially in Phase 1 (when participants provided their initial estimates) and again in Phase 3 (when participants were given feedback). This may have enhanced those items' representations (e.g., with a bit more time-on-task), providing a richer or more stable context for incorporating feedback (as with Tse et al., 2007). On the other hand, these data might also suggest that we have not explored the most relevant timescales. For example, the optimal estimate-feedback delay may be much shorter, as with the work of Johnson and Siefert (1994). A final possibility is that the participants' assessments of surprise at receiving feedback may inappropriately draw their attention back to their incorrect prior knowledge. Thus, we might find generally enhanced learning and larger effects of feedback spacing, were we to encourage participants to engage more fully with the new information—perhaps as is done in the usual EPIC procedure, in which participants revise their preferences/policies based on the new information and the inferencing that such feedback triggers (as in Munnich et al., 2004, and Ranney et al., 2008).

Learning Without Recall

From the point of view of a memory theory, the most interesting result is perhaps the existence of learning even when participants claimed "no sense" of the numerical value provided at feedback—rather like a memorial analog to blindsight (Merikle, 2007). This argues against the notion that improvements in estimation are simply the result of explicit episodic memory. The result is reminiscent of extant *dual-process* memory models. For example, Davachi et al. (2003) suggest that successful recognition could occur through a process of recollection and/or a sense of familiarity. These processes moreover appear to be subserved by distinct sub-regions in the medial temporal lobe. In the present study, we see improvement in numerical estimation—which is perhaps most akin to a cued recall task for EI and E_I items—without full recall of the number presented on the previous day. Thus, the task here is perhaps more naturally expressed in the language of the *remember/know* distinction (Knowlton, 1998). That is, while participants appear not to *remember* a number from the previous day, there is still a sense in which they *know* the number better than they knew it the day before.

Based on the significance of the existing results, however, it seems reasonable to posit that a non-episodic form of learning undergirds some of the improvement in participants' abilities to estimate accurately. Further, the learning for improved estimation (or memory) seems to occur even without an explicit recollection of the feedback from the prior day. This argues for some implicit and/or rapidly semanticized learning in support of these improvements.

Educational Implications

We have provided support for the effectiveness of instruction that engages with a person's pre-existing knowledge. A finding of central importance to educators is the lack of improvement on E items. It seems that if the pedagogical goal is to improve numeracy, then simple engagement with a given set of quantities is not enough to enhance estimation abilities for other quantities. Item specific training could be augmented with more generalizable strategies, though, as has been demonstrated with other interventions (e.g., Munnich et al., 2004). Students' abilities to estimate quantities for which they *have* received feedback can be quite good, and it seems that there is at best a small effect of whether students think about their own beliefs immediately preceding or a day prior to receiving correct instruction. This effect is in line with results from elements of the distributed learning literature (Pashler et al., 2007)—perhaps with some form of priming for subsequent encoding occurring during one's initial estimation. Future studies will examine differing intervals for both feedback and retention. Given our results, though, it is clear that effective learning can occur over a panoply of semantic, statistical, items in a framework that includes engaging students' pre-existing understandings of the material. It is worth noting further that, unlike some other forms of conceptual change cited above, the nature of what was learned here may be of a more continuous nature than, say, "what determines the phases of the moon."

Such information may be more amenable to "implicit" forms of learning than concepts that have a more categorical, intricate, or non-linear structure. It bears considering, though, whether one can introduce more graded forms of learning into the logical structures of science and related fields.

In general, participants displayed fairly good recognition of items they had seen previously, distinguishing new items from ones presented the previous day. However, improvements occurred frequently even when participants claimed "no sense" of the feedback, particularly when they found the item surprising—in other words, when estimation improvement was predicted more by surprise than by explicit recall. In such cases, learners seem to have poor metacognitive inclinations—they do not "know what they know."

Combined with strategies for evaluating one's own estimates (e.g., those provided in curricula by Munnich et al., 2004, and Ranney et al., 2008), a learner could be encouraged to generate what they believe to be a "guess," but which might turn out to be well informed by what they have in fact learned. If encouraged to make such guesses, students should still spend time reasoning about their guess; prior work from our group (Ganpule, 2005) has indicated that good estimators spend more time representing a quantity before estimating it, consistent with work from the problem solving literature. Guessing too quickly is more often counterproductive. Indeed, this result encouraged the curricula of Munnich et al. (2004) and Ranney et al. (2008) to highlight the importance of self-critique and disconfirmation. (However, one of us anecdotally finds people, at times, to be markedly underconfident estimators. For instance, people frequently claim to have "no idea" of the population of California, yet often laugh out loud when asked if it could be "1,000.") Another idea is that students might infer something about the quality of their estimate by introspecting on whether they remember being surprised at a number during instruction. Such surprise could indicate an improvement of one's sense of magnitude.

More generally, participants were able to learn something in this experiment without being aware of their learning. This strongly suggests that non-episodic (i.e., implicit or rapidly semanticized) learning processes could be involved even in the development of "declarative" factual information. Thus, depending upon an instructor's particular pedagogical goals, one might proceed with instruction without concern for students' initial awareness of their improved sense of certain facts (perhaps as in "immersion" language-learning experiences). Over the course of an intervention, one would hope that students would eventually come to know and trust their new knowledge on the topic, but such metacognitive awareness needn't come lock-step with improvements in the more basic knowledge. This line of reasoning is reminiscent of studies of children applying abstract mathematical rules before they are aware of doing so (Siegler, 2000).

Of course there is also a clear role for explicit episodic recall in learning numerical information. In particular, when participants displayed zero or moderate amounts of surprise, improvements in estimation/recollection were likely only if they believed they could recall the number. Thus, when students lack sufficient schemas for incorporating the numerical knowledge and the information imparted is unsurprising, rote memorization may be virtually the only remaining route. Of course, a complete pedagogy could include the construction of knowledge structures that would then allow for the parallel recruitment of non-episodic learning.

Instructional materials that elicit surprise in students may allow such students to learn without conscious awareness that they have learned anything—at least in domains that are scaffolded by nontrivial preexisting knowledge. If the material is unsurprising, it appears that episodic encoding may be a critical step in successful improvement. It should be noted that "surprise" might be often used as too specific a notion. It may be that the relevant feature has more to do with general emotional, motivational, or inspirational, salience—or how interesting the material is to students. (See Kang et al., 2009, on connections between surprise, motivation, and curiosity, the latter two of which are likely enhanced by soliciting estimates—and even preferences—as in the EPIC procedure). Certainly, however, it seems that there are multiple routes to learning even relatively concise facts, and a successful pedagogy might usefully engage factors such as surprise and engagement with pre-existing knowledge to bolster more rote forms of learning.

References

Chi, M. T. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *The Journal of the Learning Sciences*, *14*(2), 161–199.

Clark, D., & Ivry, R. B. (in press). Multiple systems for motor skill learning. *Wiley Interdisciplinary Reviews: Cognitive Science*.

Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences*, *100*(4), 2157–2162. doi:10.1073/pnas.0337195100

diSessa, A. A., & Sherin, B. L. (1998). What changes in conceptual change? *International Journal of Science Education*, *20*(10), 1155–1191.

Ganpule, S. (2005). *Strategy use in numerical estimation: Investigating the effects of an EPIC curriculum.* Unpublished master's thesis, University of California, Berkeley.

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.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology Learning Memory and Cognition*, *20*, 1420-1420.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, *20*(8), 963-973. doi:10.1111/j.1467-9280.2009.02402.x

Knowlton, B. J. (1998). The relationship between remembering and knowing: A cognitive neuroscience perspective. *Acta psychologica*, *98*(2-3), 253-265.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098–2109.

Merikle, P. (2007). Preconscious processing. In M. Velmans & S. Schneider (Eds.), *The Blackwell companion to consciousness*. Malden, MA: Wiley-Blackwell.

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.

Munnich, E., Ranney, M. A., & Song, M. (2007). Surprise, surprise: The role of surprising numerical feedback in belief change. 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.

Munnich, E. L., Ranney, M. A., & Appel, D. M. (2004). Numerically-Driven Inferencing in instruction: The relatively broad transfer of estimation skills. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-sixth Annual Conference of the Cognitive Science Society* (pp. 987–992). Mahwah, NJ: Erlbaum.

Munnich, E. L., Ranney, M. A., & Bachman, M. L. N. (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.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*(2), 217–227.

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin and Review*, *14*(2), 187–193.

Ranney, M., Cheng, F., Garcia de Osuna, J., & Nelson, 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. A., Rinne, L., Yarnall, L., Munnich, E., Miratrix, L., & Schank, P. (2008). Designing and assessing numeracy training for journalists: Toward improving quantitative reasoning among media consumers. In P. Kirschner, F. Prins, V. Jonker, & G. Kanselaar (Eds.), *International Perspectives in the Learning Sciences: Proceedings of the Eighth International Conference for the Learning Sciences, Vol. 2* (pp. 2-246 to 2-253). International Society of the Learning Sciences, Inc.

Rinne, L. F., Ranney, M. A., & Lurie, N. H. (2006). Estimation as a catalyst for numeracy: Micro-interventions that increase the use of numerical information in decision-making. In S. Barab, K. Hay, & D. Hickey (Eds.), *Proceedings of the seventh international conference on Learning sciences* (pp. 571-577). Mahwah, NJ: Erlbaum.

Siegler, R. S. (2000). Unconscious insights. *Current Directions in Psychological Science*, *9*, 79-83.

Straw, A. D. (2008). Vision egg: an open-source library for realtime visual stimulus generation. *Frontiers in neuroinformatics*, *2*. doi:10.3389/neuro.11.004.2008

Thagard, P. (2006). *Hot Thought: Mechanisms and Applications of Emotional Cognition*. The MIT Press.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., et al. (2007). Schemas and memory consolidation. *Science*, *316*(5821), 76-82. doi:10.1126/science.1135935

Acknowledgments