Measuring Learning in Informative Processes*

Robert C. Luskin
University of Texas at Austin

Ariel Helfer
Michigan State University

Gaurav Sood

April, 2016

Many communication processes are broadly *informative,* exposing their participants to some goodly quantity and wide variety of information on at least one sizable topic, like policy or electoral choices. The category includes both natural processes like election campaigns and ad hoc but naturalistic ones like Deliberative Polls (DPs). The information exposure need not be the point. Often, it is not. In election campaigns, it is incidental, a by-product of persuasion or mobilization. In DPs and other deliberative fora, it is intrinsic but instrumental—not an end in itself but an ingredient in more fully considered, better founded opinions. But in whatever fashion, for whatever reason, the exposure must occur, and at least some participants must learn something from it, or the process would be only degenerately "informative."[1]

Almost always, we are interested in this learning—sometimes in the learning of this or that particular bit, affecting this or that particular attitude (as in Gilens 2001), but more often in the amount of topic-wide learning, affecting a wider range of attitudes, as well as vote intentions, levels of engagement, and the like. In the aggregate, this is also normatively important, speaking to the educative value of election campaigns and the deliberative quality of deliberative fora.

But how best to measure it? The obvious choice is observed knowledge gain: the increase in the proportion of knowledge items answered correctly. For aggregate description, *faute de mieux,* this may serve. But individual-level explanatory analyses—of learning as a function of causally prior variables or of causally subsequent ones as a function of learning—are another story. There observed knowledge gain often yields underwhelming results. But why? Professors may glumly reflect that learning can be hard to induce. But is it really, to the extent it occurs, so haphazard and inconsequential? No, the problem seems to be one of measurement. Observed knowledge gain turns out to be a surprisingly weak indicator of true knowledge gain.

This paper confirms the problem, examines its roots, and suggests one simple escape.

Following a real-world illustration, we show mathematically that (1) observed knowledge gain can be *negatively* correlated with true knowledge gain; (2) observed post-process knowledge, by contrast, is always positively correlated with it; and (3) the first correlation, even when positive, can fall short of the second. We derive a necessary and sufficient condition for (1) and several sufficient conditions for (3), then use numerical simulations to show that, mathematical possibilities aside, observed knowledge gain is in fact only weakly correlated with true knowledge gain—but observed post-process knowledge quite strongly so (wherein lies the escape).

Empirical Motivation

Informative processes often bring attitude change, as the aggregate results of both Deliberative Polling (Luskin, Fishkin, and Jowell, 2002; Luskin, Hahn, and Fishkin 2012) and statistical simulations (Bartels 1996, Delli Carpini and Keeter 1996, Althaus 2003) suggest. But to what extent is it the learning, as opposed to other dynamics, that is responsible? Individual-level analyses based on observed knowledge gain often yield fainter than anticipated effects.

Consider, for instance, a simple linear model of attitude change as a function of learning and small-group influences in a DP. Let the *i*th individual's attitude toward some policy or policy proposal at time t be P_{it}, the mean attitude (excluding the individual him- or herself) in his or her small group be G_{it}, and his or her true knowledge gain be Δ_i^*. Let $t = 1$ and $t = 2$, alternatively denoted as t_1 and t_2, denote the beginning and the end of the process. Taking the observed P_{it} and thus G_{it} at face value while leaving open the question of how to proxy Δ_i^*, we may write:

$$P_{i2} - P_{i1} = \gamma_0 + \gamma_1 \Delta_i^* + \gamma_2(P_{i1} - G_{i1}) + w_i,$$

where the γ's are parameters and w_i is a disturbance. The small group coefficient γ_2 should be negative, in keeping with the individual's tending to narrow the gap between his or her initial opinion and those of the people around him or her, while the learning coefficient γ_1 should carry the same sign as that of the sample-wide mean attitude change $\bar{P}_2 - \bar{P}_1$, if those learning the most are in fact mostly responsible for the mean attitude change (Luskin, Fishkin, and Jowell 2002).

Illustratively, we estimate this model for two national DPs, on Australia's constitutional referendum in 1999 and the issue of crime in Britain in 1994 (Luskin et al. 2000; Luskin, Fishkin, and Jowell 2002), comparing the results when true knowledge gain Δ_i^* is proxied by observed knowledge gain Δ_i versus observed post-event knowledge x_{i2} (see Appendix A for the knowledge items involved). Using Δ_i, the results, in Table 1, are disappointing. The estimated γ_1 is substantively scrawny and statistically insignificant in five of the six estimations. Things look up, however, when we try substituting x_{i2}. The estimated coefficients get larger, and five of the six, all bearing the right sign, are statistically significant. Why this contrast?

(Table 1 about here)

"Observed" versus "True" Knowledge

The explanation begins with the differences between "observed" and "true" knowledge—and, derivatively, "observed" and "true" knowledge gain.

Definitions

The knowledge we are used to thinking about is "observed." It is, by definition, all we ever see. Conventionally, the *i*th person's *observed knowledge* at time *t*—call it x_{it}—is the proportion of the questionnaire's knowledge items the person answers correctly at time *t* (normally based on an index of $J \geq 2$ items, with $J = 1$ included as a special case). More precisely, $x_{it} = n(C_{it})/n(Q)$, where Q denotes the questionnaire's knowledge items, C_{it} the subset the *i*th person answers correctly at time *t*, and $n(Q) = J$ the number of elements in Q.

We are much less used to thinking about "true knowledge"—what "observed knowledge" is trying to observe. Its depths—what it means to *know* something—can be left to philosophy and neuroscience. Nearer the surface, however, we may define the *i*th person's *true knowledge* at time t—call it X_{it}—as the proportion of all knowable, relevant items he or she knows at time t. I.e., $X_{it} = n(K_{it})/n(U)$, where U ($\supset Q$) is the set of knowable, relevant items, K_{it} the subset the *i*th person knows at time t, and $n(K_{it})$ and $n(U)$ ($= R$) the numbers of items in K_{it} and U.

In these terms, *observed knowledge gain* is $\Delta_i \equiv x_{i2} - x_{i1}$, and *true knowledge gain* $\Delta_i^* \equiv X_{i2} - X_{i1}$. We call these "gains" because they are generally ≥ 0. So long as the process is not actually *dis*informative, Δ_i^* should always ≥ 0. So, with occasional, mostly small exceptions (mainly from guessing that happens to be luckier at t_1 than at t_2), should Δ_i.

Details and Commentary

- We leave the knowable bits or "items" in U unspecified, taking them only to be relatively granular.[2] Illustratively, they may include ascriptive, descriptive, definitional, nesting, or causal propositions: e.g., in the political domain, that the Republican party is right-of-center, that Arizona's immigration bill (SB 1070) requires state law enforcement officers to try to determine immigration status during lawful stops, detentions, or arrests; that the Affordable Care Act is what is commonly called "Obamacare," that the Republicans are a political party, that greenhouse gas emissions are changing the earth's climate.
- U includes everything that is relevant and knowable—what God knows—whether any mortal knows it or not. That, at any rate, is the cleanest, least arbitrary, and most defensible definition. Even the greatest experts always have more to learn. Confining U to what some person or other already knows would leave no room for new knowledge, for scientific discovery. It would also make no difference to our proofs and negligibly little to our simulations. On topics like policy issues, let alone politics as a whole, $n(U)$ must be at least in the hundreds of thou-

sands, if not orders of magnitude higher.

- Thus X_{it} never gets much above 0. Let X_{mt} and X_{Mt} be the attained, as distinct from theoretical or attain*able*, minimum and maximum—the lowest and highest X_{it} in the population (not just the sample participating in the process). Clearly, $X_{mt} = 0$. Some few people know absolutely nothing about politics. But X_{Mt} is also very low. Those following politics as vocation or avocation may know thousands, even tens of thousands ot items, but $n(U)$ is vastly larger. Thus it is hard to see how X_{Mt} can be higher than, say, .001. Probably it is far lower.

- Since very few people follow politics very closely, X_{it}'s distribution must also be severely right-skewed, with most people far closer to 0 than to X_{Mt}. That makes X_{it}'s variance exremely low. Its *co*variance with other variables (including a good x_{it}) can still be high.

- By contrast, x_{it} typically ranges all the way from 0 to 1. Q is a very small, very easy subset ("item sample") of U. Question writers do not know enough to ask the items known only by policy experts, let alone those no one knows. Of those they know, they favor relatively easy ones. Otherwise, hardly anyone would score much above 0, except by lucky guessing.

- Q's biased subsetting of U makes x_{it} much less right-skewed than X_{it}—indeed possibly not right-skewed at all, if the subsetting is biased enough. In our observation, for what it is worth, x_{it} does usually remain right-skewed, just much more mildly than X_{it}.

- The same contrast of ranges holds for Δ_i vis-à-vis Δ_i^*: Δ_i can be sizable (averaging above .2 in some DPs), while Δ_i^* must be extremely small.[3] Though typically larger than $n(K_{i1})$, $n(K_{i2})$ can still only be a tiny fraction of $n(U)$. At t_2 as at t_1, there is vastly more to know. Of course the *proportional* true gain Δ_i^*/X_{i1} may still be sizable. A Δ_i^* of .0005 may be a 20 or 30% increase.

- Accepting X_{it}'s 0 to near-0 range is humbling, not unlike contemplating one's place in the universe from the deep countryside on a clear, moonless night. Evaluatively, a more approachable standard, confining U to information that is reasonably salient (readily encountered), in the spirit of Lupia (2006), may be preferable. That would narrow the gap between the never-seen mean X_{it}, and the observed mean x_{it}, casting the latter—typically not tiny, but less than impressively large—in a more flattering light. It is no help, however, in understanding Δ_i's relation to Δ_i^*. For that, the denominator must include whatever can be learned—much of which, especially in hothouse processes like DPs, may be useful but far from initially salient.

- x_{it} differs from X_{it} in both its domain and what it tallies: Q (just the questionnaire's items) versus U (all relevant, knowable items) and C_{it} (items answered correctly) versus K_{it} (items known). Correspondingly, the measurement error, $e_{it} = x_{it} - X_{it}$, is of two sorts. The *response-production error* e_{it}^r lies in the responses to given items. The jth item may sometimes be answered correctly ($x_{ijt} = 1$) by those who do not know the answer ($X_{ijt} = 0$) or incorrectly ($x_{ijt} = 0$) by those who do ($X_{ijt} = 1$). Guessing, if lucky, yields $x_{ijt} = 1$ despite $X_{ijt} = 0$. Reticence may yield $x_{ijt} = 0$ (from a DK response) despite $X_{ijt} = 1$. At the index level, e_{it}^r is the mean, over j, of $x_{ijt} - X_{ijt}$—or, equivalently, $x_{it} - X'_{it} = [n(C_{it}) - n(K'_{it})]/n(Q)$, where $n(K'_{it})$ is the number of items in Q the ith person knows at time t, and $X'_{it} \equiv n(K'_{it})/n(Q)$ thus a version of X_{it} confined to Q. The *item-sampling error* e_{it}^s, the difference between the fractions of the questionnaire items and of the relevant universe the ith person knows—$X'_{it} - X_{it} = n(K'_{it})/n(Q) - n(K_{it})/n(U)$—lies in Q's subsetting of U. Note that $e_{it} = x_{it} - X_{it} = (x_{it} - X'_{it}) + (X'_{it} - X_{it}) = e_{it}^r + e_{it}^s$.[4]

- We make these distinctions to be clear about what we are including and excluding. The item-sampling error is important to understanding Δ_i's weakness as an indicator of Δ_i^*, as we shall see; the response-production error, not so. From here on, therefore, we exclude the latter, setting $C_{it} = K'_{it}$, $x_{it} = X'_{it}$, and thus $e^r_{it} = 0$. This simplifies the math at minor cost.

- The *item-sampling bias* $E(e^s_{it})$ is the mean item-sampling error over hypothetically repeated, independent draws of $n(Q)$ items from U. For the ith person at time t, X_{it} is a constant, but $n(K'_{it})$ and thus X'_{it} and $e^s_{it} = X'_{it} - X_{it}$ vary from item-sample to item-sample.[5] Absent response production error (given $e^r_{it} = 0$), we may take $E(x_{it})$ ($=E(X'_{it})$) $= X_{it}$ when $X_{it} = 0$. Those knowing none of the items in Q answer none correctly. Setting $E(x_{it}) = X_{it} = 0$ aside, however, Q's privileging of easy items implies that $E(X'_{it}) > X_{it}$ and thus that $E(e^s_{it}) = E(X'_{it}) - X_{it} > 0$: the item sampling bias is positive (and, we submit, gigantic). It also implies $E(x_{it}) > X_{it}$.
- Implicitly, we have been taking the item-level x_{ijt} and X_{ijt} as binary—$x_{ijt} = 1$ (correct) or 0 (incorrect), $X_{ijt} = 1$ (known) or 0 (not known)—a close and harmless approximation.[6] We could wrinkle the definitions to allow X_{ijt} and x_{ijt} to vary more graduatedly from 0 to 1, but our math, simulations, and conclusions concern the proportions X_{it} and x_{it}, which already do just that (for $J > 1$, in the case of x_{it}).

Theory

Now consider the relations between these variables. The learning is $\Delta^*_i = X_{i2} - X_{i1}$, the measurements x_{i1}, x_{i2}, and $\Delta_i = x_{i2} - x_{i1}$. We take X_{i2} to be a function of X_{i1} and x_{i1} and x_{i2} to be functions of X_{i1} and X_{i2}, respectively. Indirectly, that also makes Δ^*_i, x_{i2}, and Δ_i functions of X_{i1}. For simplicity, we continue to assume $e^r_{it} = 0$, regard X_{it} and x_{it} as continuous over the [0, 1] interval (in practice, the [0, X_{M1}] interval for X_{it}), and take the functions relating X_{i2} to X_{i1} and x_{i1} and x_{i2} to X_{i1} and X_{i2} (from which those relating Δ^*_i and Δ_i to X_{i1} derive) to be continuous and differentiable over the [0, 1] interval.[7]

In this light, the knowledge and learning should have the following properties:

L1. *The participants learn*: $E(X_{i2}|X_{i1})$ is a strictly increasing function of X_{i1}, with slope $\delta E(X_{i2}|X_{i1})/\delta X_{i1} > 1$. Those starting at $X_{i1} = k > 0$ end at $E(X_{i2}|X_{i1} = k) > k$.

L2. *The more knowledgeable learn more*: $E(\Delta_i^* |X_{i1})$ is also a strictly increasing function of X_{i1}. More new information sticks when there is more previously cognized information for it to stick to (e.g., Recht and Leslie 1988; Eckhardt, Wood, and Jacobvitz 1991; Cooke, Atlas, Lane, and Berger 1993; Hambrick 2003).[8]

L3. *Those who know nothing learn nothing*: $E(X_{i2}|X_{i1} = 0) = 0$.[9] At $X_{i1} = 0$, there is nothing for new information to stick to. In any given unit (as distinct from item) sample nobody may actually have X_{i1} = exactly 0. People of that description are isolated, cognitively incapable, or both and thus unlikely to be interviewed or participate in an informative process. But they would be expected to emerge from such a process, could they be roped into one, at $E(X_{i2}|X_{i1}) = X_{i1} = 0$.

The *observed* knowledge and learning should have these properties:

M1. *Those who know nothing show no knowledge, while those who know the most show as much as the measure permits*: $E(x_{it}|X_{it} = 0) = 0$, $E(x_{it}|X_{it} = X_{Mt}) = 1$.[10] Absent lucky guessing and reticence (both assumed away as part of $e_{it}^r = 0$), those for whom $X_{it} = 0$ say DK to all the items in Q, yielding $E(x_{it}|X_{it} = 0) = 0$, while those for whom $X_{it} = X_{Mt}$, who know the answers to questions much harder than anything in Q, answer them all correctly. $E(x_{it}|X_{it})$ may possibly reach 1 for X_{it} well shy of X_{Mt} but at least reaches 1 by (and probably long before) $X_{it} = X_{Mt}$.

M2. *The more knowledgeable appear to know more, but the less so, the more they know*: $E(x_{it}|X_{it})$ is an increasing function of X_{it}, $\delta E(x_{it}|X_{it})/\delta X_{it}$ a decreasing function of it. The first clause says that x_{it} is neither useless ($\delta E(x_{it}|X_{it})/\delta X_{\text{it}}$ always = 0) nor perverse ($\delta E(x_{it}|X_{it})/\delta X_{it}$ ever < 0) as an indicator of X_{it}, the second that a higher X_{it} and thus a higher $E(x_{it}|X_{\text{it}})$ leave less room

$(1 - E(x_{it}|X_{it}))$ for further increase. Thus $\delta E(x_{it}|X_{it})/\delta X_{it}$ is initially steep but then flattening, approaching 0, as X_{it} nears X_{Mt}.[11] This is a "ceiling effect."

M3. *People appear to know more than they do*: $E(x_{it}|X_{it}) > X_{it}$ (for $X_{it} > 0$). Item-sampling bias means that no real-world questionnaire can yield $x_{it} < X_{it}$.

M4. *The more knowledgeable appear to learn more, then, beyond a point, to learn less*: $\delta E(\Delta_i|X_{i1})/\delta X_{i1} > 0$ for $X_{i1} < V_\Delta$, $= 0$ for $X_{i1} = V_\Delta$, and < 0 for $X_{i1} > V_\Delta$, where V_Δ is the turning point $(0 < V_\Delta < X_{M1})$. Both those who know nothing ($X_{i1} = 0$) and those who know the most ($X_{i1} = X_{M1}$) appear to learn the least. The former actually learn nothing; the latter, despite actually learning the most, cannot show any learning, having started at $E(x_{it}|X_{it} = X_{Mt}) = 1$. In-between, increasing X_{i1} increases $E(x_{it}|X_{it})$ and $E(\Delta_i^*|X_{i1})$ but also decreases the room $(1 - E(x_{i1}|X_{i1}))$ left for $E(\Delta_i|X_{i1})$. Up to V_Δ, the net effect is to increase Δ_i; beyond V_Δ, to decrease it. This too is a species of ceiling effect, with the difference that the curve does not merely plateau as X_{i1} approaches V_Δ but declines once X_{i1} passes it.

Model and Assumptions

Together, L1-L3 and M1-M4 suggest that Δ_i may not be a very good indicator of Δ_i^*. The higher the X_{i1}, the higher the expected Δ_i^* (L2), but, beyond a point, the lower the expected Δ_i (M4), obviously sapping Δ_i's correlation with Δ_i^*. The correlation could even be negative. The same properties also suggest that x_{i2} may do better. Increasing X_{i1} increases both the expected Δ_i^* and the expected x_{i2}, the latter via X_{i2} (L2, M1). Those ending with a high x_{i2} tend to have a high Δ_i^*—either observably, if starting at relatively low X_{i1} and thus low x_{i1}, or unobservably, if starting at relatively high X_{i1} and thus high x_{i1}.

But let us state and reason this more precisely. L1-L3 and M1-M4 can be concretized in learning and measurement equations, expressing X_{i2} as a function of X_{i1} and x_{i1} and x_{i2} as functions of X_{i1} and X_{i2}, respectively—and in the implied equations for x_{i2}, Δ_i^* and Δ_i as functions strictly of X_{i1}.

The functional forms need particular care. As proportions, X_{i2}, x_{i1}, x_{i2}, Δ_i^*, and Δ_i are all "limited"—bounded in theory by 0 and 1 (assuming $x_{i2} \geq x_{i1}$, so that $\Delta_i \geq 0$) and in practice by 0 and 1 for x_{i1} and x_{i2}, by 0 and X_{M2} for X_{i2} and Δ_i^*, and by 0 and $1 - x_{i1}$ for Δ_i. Strictly speaking, therefore, linear equations are inapt. This is not especially important for X_{i2}'s dependence on X_{i1}. We are concerned with increases—with knowledge *gains*—and X_{i2}'s practical ceiling, X_{M2}, is nowhere near 1. Thus we take that equation as linear. But x_{i1} and x_{i2} do frequently near or equal 1, and their consequently decreasing slopes on X_{i1} (M2's "ceiling effect") are a major part of the story. The logistic form is relatively intractable here, so we make the measurement equations quadratic—stipulating that they be concave and that the vertices occur at $X_{i1} = X_{M1}$ (and $X_{i2} = X_{M1}$) thus confining the variation in X_{i1} (and X_{i2}) to the upslope. Then both $\delta E(x_{i1}|X_{i1})/\delta X_{i1}$ and $\delta E(x_{i2}|X_{i1})/\delta X_{i1} > 0$, with both $\delta^2 E(x_{i1}|X_{i1})/\delta X_{i1}^2$ and $\delta^2 E(x_{i2}|X_{i1})/\delta X_{i1}^2 < 0$ (up to $X_{i1} = X_{M1}$, at whch point $\delta E(x_{i1}|X_{i1})/\delta X_{i1} = \delta E(x_{i2}|X_{i1})/\delta X_{i1} = \delta^2 E(x_{i1}|X_{i1})/\delta X_{i1}^2 = \delta^2 E(x_{i2}|X_{i1})/\delta X_{i1}^2 = 0$).

More precisely, we let

(1) $X_{i2} = \beta X_{i1}$

(2) $x_{i1} = bX_{i1} + c\,X_{i1}^2 + u_{i1}$

(3) $x_{i2} = bX_{i2} + c\,X_{i2}^2 + u_{i2} = b\beta X_{i1} + c\beta^2 X_{i1}^2 + u_{i2}$

implying

(4) $\Delta_i^* = X_{i2} - X_{i1} = (\beta - 1)X_{i1}$

(5) $\Delta_i = x_{i2} - x_{i1} = (\beta - 1)bX_{i1} + (\beta^2 - 1)c\,X_{i1}^2 + u_{i2} - u_{i1}$,

and the conditional expectations:

(6) $E(X_{i2}|X_{i1}) = \beta X_{i1}$

(7) $E(x_{i1}|X_{i1}) = bX_{i1} + c\,X_{i1}^2$

(8) $E(x_{i2}|X_{i1}) = b\beta X_{i1} + c\beta^2\,X_{i1}^2$

(9) $E(\Delta_i^*|X_{i1}) = (\beta - 1)X_{i1}$

(10) $E(\Delta_i|X_{i1}) = (\beta - 1)bX_{i1} + (\beta^2 - 1)c\,X_{i1}^2$.

We make the equations intercept-less, in keeping with L3 and M1; adopt the textbookish assumptions that $E(u_{i1}) = E(u_{i2}) = E(v_i) = 0$, that u_{i1} and u_{i2} are independent of X_{i1}, X_{i2}, and one another, and that v_i is independent of u_{i1}, u_{i2}, and X_{i1} to get from (1)-(5) to (6)-(10); and make the learning equation (1) exact (undisturbed) and take the measurement coefficients b and c to be the same at t_1 and t_2 (leaving the difference between x_{i1} and x_{i2} a function only of the differences between X_{i1} and X_{i2} and between u_{i1} and u_{i2}) to streamline the math;[12] We also assume, innocuously, that $V(X_{i1})$ be strictly > 0 (not everyone starts with exactly the same true knowledge).

More centrally, we stipulate that $\beta > 1$, consistent with L1-L2; that $b > 0$ and $c < 0$, to make (2) and (3) and thus (7) and (8) concave, in keeping with M2 and M4; and take the vertices of (2) and (3) and thus (7) and (8) to occur at Xi1 = XM1 (or, equivalently,Xi2 = XM2 = $\beta XM1$). We also set $E(\Delta_i|X_{i1} = 0) = 0$, in keeping with M1, adding, symmetrically, that $E(\Delta_i|X_{i1} = X_{M1})$ also = 0.

Figure 1 shows the curves traced by $E(x_{i1}|X_{i1})$, $E(x_{i2}|X_{i1})$, and $E(\Delta_i|X_{i1})$. The vertices occur at $X_{i1} = -b/2c$, $-b/2c\beta$, and $-b/2(\beta + 1)c$, respectively. Call these values V_1, V_2, and V_Δ. Since $b > 0$, $c < 0$, and $\beta > 1$, $0 < V_\Delta < V_2 < V_1$. As X_{i1} increases, $E(\Delta_i|X_{i1})$ peaks first, then $E(x_{i2}|X_{i1})$, then $E(x_{i1}|X_{i1})$. The associated ordinates are $-b^2/4c$ for both V_1 and V_2 but $-(\beta - 1)b^2/4(\beta + 1)c$ for

(Figure 1 about here)

V_Δ, making $E(\Delta_i|X_{i1})$'s peak lower than $E(x_{i2}|X_{i1})$'s = $E(x_{i1}|X_{i1})$'s. As the graph also suggests, $E(\Delta_i|X_{i1})$ peaks when X_{i1} is half its attained maximum: $V_\Delta = X_{M1}/2$.[13]

All this is simplification, of course. The functional forms may not be exactly right. X_{i2} may depend on more than X_{i1}, x_{it} on more than X_{it}. The assumptions about u_{i1}, u_{i2}, and v_i may not all hold. Their means may not be 0; they may not be independent of X_{i1} or each other; and their variances cannot be constant, given the limited nature of X_{i2} , x_{i1}, and x_{i2}. But all models and assumptions are, at best, approximations, and this relatively simple set-up heads us toward some illuminating results, as we are about to see.

Mathematical Implications

Our concern is with the correlations between the indicators Δ_i and x_{i2}, on the one hand, and the concept Δ_i^*, on the other:

$$(12) \qquad \rho_{\Delta\Delta^*} \equiv C(\Delta_i, \Delta_i^*) / \sqrt{V(\Delta_i)V(\Delta_i^*)}$$

$$(13) \qquad \rho_{x_2\Delta^*} \equiv C(x_{i2}, \Delta_i^*) / \sqrt{V(x_{i2})V(\Delta_i^*)},$$

where we use subscripted ρ's for correlations and V's and C's for variances and covariances. A glance at Figure 1 suggests that $\rho_{\Delta\Delta^*}$ can be negative and, even when positive, exceeded by $\rho_{x_2\Delta^*}$. But let us see where the math (detailed in Appendix B) takes us.

The variances and covariances in (14)-(18) can be shown to be:

$$(19) \qquad C(\Delta_i, \Delta_i^*) = (\beta - 1)^2[(\beta + 1)cC(X_{i1}^2 X_{i1}) + bV(X_{i1})]$$

(20) $C(x_{i2},\ \Delta_i^*) = b\beta(\beta - 1)V(X_{i1}) + c\beta^2(\beta - 1)C(X_{i1}^2, X_{i1})$

(21) $V(\Delta_i^*) = (\beta - 1)^2 V(X_{i1})$

(22) $V(\Delta_i) = (\beta^2 - 1)^2 c^2 \mathrm{V}(X_{i1}^2) + (\beta - 1)^2 b^2 \mathrm{V}(X_{i1}) - 2bc(\beta^2 - 1)(\beta - 1)C(X_{i1}^2, X_{i1}) + V(u_{i2})$
$+ V(u_{i1})$

(23) $V(x_{i2}) = b^2\beta^2 V(X_{i1}) + c^2\beta^4 V(X_{i1}^2) + 2bc\beta^3 C(X_{i1}^2, X_{i1}) + V(u_{i2})$

Several further results immediately follow:

R1. Either $\rho_{\Delta\Delta^*}$ or $\rho_{x_2\Delta^*}$ (or both) can be negative.

Note that for $X_{i1} \geq 0$, as it is here, $C(X_{i1}^2, X_{i1}) > 0$. And, by assumption, $b > 0$, $\beta > 1$, and $V(X_{i1})$ strictly > 0, while $c < 0$. Thus, specifically,

R2. $\rho_{\Delta\Delta^*} > 0$ iff $bV(X_{i1}) > -(\beta + 1)cC(X_{i1}^2, X_{i1})$

R3. $\rho_{x_2\Delta^*} > 0$ iff $bV(X_{i1}) > -c\beta C(X_{i1}^2, X_{i1})$

Note that R2 ⇔ $\rho_{\Delta\Delta^*} > 0$ iff $\gamma < -b/(\beta + 1)c$, and R3 ⇔ $\rho_{x_2\Delta^*} > 0$ iff $\gamma < -b/\beta c$, where $\gamma = C(X_{i1}^2, X_{i1})/V(X_{i1})$ is the slope of the linear, bivariate, population regression of X_{i1}^2 on X_{i1}. And since $-b/(\beta + 1)c$ obviously $< -b/\beta c$ (both > 0),

R4. $\rho_{\Delta\Delta^*} > 0 \Rightarrow \rho_{x_2\Delta^*} > 0$,

refining and elaborating on R1: $\rho_{\Delta\Delta^*}$ and $\rho_{x_2\Delta^*}$ can both > 0 (when $\gamma < b/(\beta + 1)c$); they can both < 0 (when $\gamma > -b/\beta c$); $\rho_{\Delta\Delta^*}$ can < 0, but $\rho_{x_2\Delta^*} > 0$ (when $-b/(\beta + 1)c < \gamma < -b/\beta c$); but $\rho_{x_2\Delta^*}$ cannot < 0 if $\rho_{\Delta\Delta^*} > 0$. If $\rho_{\Delta\Delta^*}$ is positive, so is $\rho_{x_2\Delta^*}$. Thus $\rho_{x_2\Delta^*}$ must > 0 at least as frequently as $\rho_{\Delta\Delta^*}$

Beyond this, the correlations are less scrutable. Plugging (19)-(23) into (12) and (13) yields nonlinear, nonadditive functions of β, b, c, $C(X_{i1}^2 X_{i1})$, $V(X_{i1})$, $V(X_{i1}^2)$, $V(u_{i1})$, and $V(u_{i2})$. A jointly sufficient condition for $\rho_{\Delta\Delta^*} > \rho_{x_2\Delta^*}$ is that $C(x_{i2}, \Delta_i^*) > C(\Delta_i, \Delta_i^*)$ and $V(\Delta_i) > V(x_{i2})$. But neither half of that condition necessarily obtains.

Numerical Simulations

So R1-R4 reveal only so much. Both $\rho_{\Delta\Delta^*}$ and $\rho_{x_2\Delta^*}$ can be negative, $\rho_{\Delta\Delta^*}$ more often than $\rho_{x_2\Delta^*}$. But how often, in each case, does that tend to be? Does $\rho_{x_2\Delta^*}$ tend to exceed or fall short of $\rho_{\Delta\Delta^*}$, and by what margin? How large does each tend to be? Each could be positive but weak. To address such questions, we turn to numerical (as distinct from statistical) simulations to get a sense of $\rho_{\Delta\Delta^*}$ and $\rho_{x_2\Delta^*}$ for plausible combinations of β, b, c, $C(X_{i1}^2 X_{i1})$, $V(X_{i1})$, $V(X_{i1}^2)$, $V(u_{i1})$, and $V(u_{i2})$.[14] We sketch the steps here, leaving further details to Appendix C.

The parameters β, b, c, $C(X_{i1}^2 X_{i1})$, $V(X_{i1})$, $V(X_{i1}^2)$, $V(u_{i1})$, and $V(u_{i2})$, governing the t_1 distribution of knowledge, the learning from the process, and the measurement of knowledge, vary across populations, topics, processes, and measurements. The maximum attained t_1 knowledge X_{M1} may be greater for some topics and for some populations than others (greater, e.g., about food in France than in Britain or about nuclear energy in Japan than in Rwanda). The expectable learning may be greater for some processes than others, depending on the quantity and accessibility of the information provided, the incentives to take it aboard, and the extent of helpful interactions between learners. DPs may typically induce more learning than public relations campaigns, and some DPs more than others. The knowledge index may entail greater or lesser item-sampling bias. Etc.

Realistically, these parameters can plausibly occupy only certain ranges. Since the exact ranges are inherently debatable, we make them generously wide, bounded at numbers that stretch the meaning of "plausible." effects of shifting them higher or lower. We denote the lower and

(Table 2 about here)

upper bounds by the subscripts m and M, appended to β, b, c, $C(X_{i1}^2 X_{i1})$, $V(X_{i1})$, $V(X_{i1}^2)$, $V(u_{i1})$, and $V(u_{i2})$ (and adopt the same convention for other quantities figuring in our reasoning, notably including X_{M1}). Table 2 summarizes the parameter ranges, as well as the ancillary settings to be introduced as we proceed (α, π, λ, q_m, q_M, ψ_m, ψ_M, φ_{1m}, φ_{1M}, φ_{2m}, and φ_{2M}).[15]

Numerical Scenarios

We draw the parameters in the sequence β, b, c, $C(X_{i1}^2 X_{i1})$, $V(X_{i1})$, $V(X_{i1}^2)$, $V(u_{i1})$, and $V(u_{i2})$. Here is the reasoning by which we bound them:

β. From (7), people starting at X_{i1} can be expected to gain $(\beta - 1)X_{i1}$, making $\beta - 1$ the proportional increase, Δ_i^* / X_{i1}. We take 20% ($\beta_m = 1.2$) to 100% ($\beta_M = 2$) as the plausible range.

b. As noted, $E(x_{i1}|X_{i1})$'s vertex is at $X_{i1} = X_{V1} = -b/2c$.[16] We set $X_{M1} = -b/2c$.

(6) $\quad X_{M1} = -b/2c\beta \;\Leftrightarrow$

(7) $\quad c = -b/2\beta X_{M1} \Leftrightarrow$

Now, set (2′) to pass through (0,0) and (X_{M1}, 1), so that $E(x_{i1}|X_{i1}) = 0$ at $X_{i1} = 0$ and = 1 at $X_{i1} = X_{M1}$. Then

$$1 = bX_{M1} + c\,X_{M1}^2$$

$$\Leftrightarrow \quad bX_{M1} + c\,X_{M1}^2 - 1 = 0$$

$$\Leftrightarrow \quad bX_{M1} + (-b/2\beta X_{M1})\,X_{M1}^2 - 1 = 0 \text{ (plugging in } -b/2\beta X_{M1} = c\text{, from (7))}$$

$$\Leftrightarrow \quad bX_{M1} + (-b/2\beta)X_{M1} - 1 = 0$$

$\Leftrightarrow \quad (b - (b/2\beta))X_{M1} - 1 = 0$

$\Leftrightarrow \quad (2\beta b - b)/2\beta))X_{M1} = 1$

$\Leftrightarrow \quad ((2\beta - 1)b)/2\beta))X_{M1} = 1$

$\Leftrightarrow \quad ((2\beta - 1)b))X_{M1} = 2\beta$

$\Leftrightarrow \quad b = 2\beta/(2\beta - 1)X_{M1}$

***b*2.** Since $b_1 > \beta b_{2,} \Leftrightarrow b_2 < b_1/\beta$, we set $b_{2M} = b_1/\beta$, using the previously drawn b_1 and β. Now b_{2m}. From (3), $X_{M2} = \beta X_{M1} + v_M$, where v_M denotes the v associated with X_{M1} (as distinct from the maximum v). Let the maximum v_M be $v_{MM} = \lambda\beta X_{M1}$ ($\lambda > 0$), so that $X_{M2M} = \beta(1 + \lambda)X_{M1,}$ i.e., X_{M2} can be up to 100λ% larger than expectable from X_{M1}. Then, taking $x_{M2} = 1$ and b_2 thus $= 1/X_{M2}$, $b_{2m} = 1/X_{M2M} = 1/\beta(1 + \lambda)X_{M1} = b_1/\beta(1 + \lambda)$. We set $\lambda = .2$, yielding $b_{2m} = b_1/1.2\beta$.

***V*(*X*i1).** Theoretically, $V(X_{i1})$ can range from 0 to .25—in the first case, when everyone has the same true knowledge; in the second, when half the population is at 1, and half at 0, with nobody in-between. In reality, however, 0 is too low, and .25 vastly too high. For high q, the stylized distribution with 100q% of the population at 0 and 100(1 – q)% uniformly distributed over $[0, X_{M1}] = [0, 1/b_1]$, captures X_{i1}'s severe right skew, and its readily calculated variance, of $[X_{M1}^2(1 - q)(1 + 3q)]/12 = [(1 - q)(1 + 3q)]/12b_1^2$,[17] suffices to give a rough sense of $V(X_{i1})$.[18] For now, we set $q_m = 2/3$ and $q_M = 3/4$, implying $\sigma_{X_1m}^2 = 13/192b_1^2 \cong .06771/b_1^2$ (for $q = 3/4$) and $\sigma_{X_1M}^2 = 1/12b_1^2 \cong .08333/b_1^2$ (for $q = 2/3$). Since b_1 is very large, both $\sigma_{X_1m}^2$ and $\sigma_{X_1M}^2$ are very small. At t_2, $\sigma_{X_2}^2$ depends on σ_v^2 as well as $\beta^2\sigma_{X_1}^2$. Let σ_v^2 be 100ψ% of $\beta^2\sigma_{X_1}^2$ ($\psi > 0$) implying that $\sigma^2_{X_2}$ = (1 + ψ) $\beta^2\,\sigma^2_{X_1}$ and making the ρ^2 (population R^2) for (3) $\beta^2\,\sigma^2_{X_1}$ /(1 + ψ) $\beta^2\,\sigma^2_{X_1}$ = 1/(1 + ψ). Since (3)'s variables are conceptual, and the rich reliably get richer, we set ψ_M = 3/7 (ρ^2 = .7) and ψ_m = 1/9 (ρ^2 = .9), implying $\sigma^2_{X_2m}$ = (1 + ψ_m) $\beta^2\,\sigma^2_{X_1}$ = (10/9) $\beta^2\,\sigma^2_{X_1}$ and $\sigma^2_{X_2M}$ = (1 + ψ_M) $\beta^2\,\sigma^2_{X_1}$ = (10/7) $\beta^2\,\sigma^2_{X_1}$, using the previously drawn β and $\sigma^2_{X_1}$.

***V*(x_{i2}).** Item-sampling bias makes x_{i1} and x_{i2} much less right-skewed than X_{i1} and X_{i2} (if right-skewed at all). Thus $\sigma^2_{x_1}$ and $\sigma^2_{x_2}$ dwarf $\sigma^2_{X_1}$ and $\sigma^2_{X_2}$. Let $\sigma^2_{u_t}$ be 100 φ_t % of $b_t^2\,\sigma^2_{X_t}$, where 0 < $\varphi_{tm} < \varphi_t < \varphi_{tM}$, so that $\sigma^2_{x_t}$ = (1 + φ_t) $b_t^2\,\sigma^2_{X_t}$. The reliability $\rho_{x_t x_t'}$ = 1 – ($\sigma^2_{u_t}/\sigma^2_{x_t}$) = $b_t^2\sigma^2_{X_t}/\sigma^2_{x_t}$ is then 1/(1 + φ_t). We take $\rho_{x_1x_1'}$ as ranging from .4 to .75 (on the high side for attitudes, but about right for knowledge). Much lower $\rho_{x_1x_1'}$'s tend to be disqualifying (the measure gets revised or discarded), much higher ones rare. Thus we set φ_{1m} = 1/3 ($\rho_{x_1x_1'}$ = .75) and φ_{1M} = 3/2 ($\rho_{x_1x_1'}$ = .4), implying $\sigma^2_{x_1m}$ = (4/3) $b_1^2\,\sigma^2_{X_1}$ and $\sigma^2_{x_1M}$ = (5/2) $b_1^2\,\sigma^2_{X_1}$. Given some learning-based reduction in noise, $\rho_{x_2x_2'}$ should tend to be slightly > $\rho_{x_1x_1'}$ and φ_2 thus slightly < φ_1. Thus we set φ_{2m} = 1/4 ($\rho_{x_2x_2'}$ = .8) and φ_{2M} = 11/9 ($\rho_{x_2x_2'}$ = .45), implying $\sigma^2_{x_2m}$ = (5/4) $b_2^2\,\sigma^2_{X_2}$, and $\sigma^2_{x_2M}$ = (20/9) $b_2^2\,\sigma^2_{X_2}$, using the previously drawn values of b_1, b_2, $\sigma^2_{X_1}$, and $\sigma^2_{X_2}$.[19]

Mechanics

The simplest distribution of each parameter would be uniform, but our reasoning suggests some skew. Within the ranges just sketched, b_1 and b_2 should generally be closer to b_{1M} and b_{2M} than to b_{1m} and b_{2m} (their distribution distinctly left-skewed), while β, $\sigma^2_{X_1}$, $\sigma^2_{X_2}$, $\sigma^2_{x_1}$, and $\sigma^2_{x_2}$ should generally be closer to β_m, $\sigma^2_{X_1m}$, $\sigma^2_{X_2m}$ $\sigma^2_{x_1m}$, and $\sigma^2_{x_2m}$ than to β_M, $\sigma^2_{X_1M}$, $\sigma^2_{X_2M}$, $\sigma^2_{x_1M}$, and $\sigma^2_{x_2M}$ (their distribution distinctly right-skewed). The half-standard-normal, $|N(0,1)|$, decidedly but not grotesquely skewed, seems a reasonable embodiment of this thinking.

More precisely, we draw suitable translations of b_1, b_2, β, $\sigma^2_{X_1}$, $\sigma^2_{X_2}$, $\sigma^2_{x_1}$, and $\sigma^2_{x_2}$ from within the corresponding translations of their ranges. The translations "reflect" b_1 and b_2 (making high values low and vice versa); translate b_{1M}, b_{2M}, β_m, $\sigma^2_{X_1m}$, $\sigma^2_{X_2m}$, $\sigma^2_{x_1m}$, and $\sigma^2_{x_2m}$ to 0; and let b_{1m} and b_{2m} be undershot and β_M, $\sigma^2_{X_1M}$, $\sigma^2_{X_2M}$, $\sigma^2_{x_1M}$, and $\sigma^2_{x_2M}$ be overshot only 100α% of the time.[20] Here we set $\alpha = .01$. Once having drawn b_1, b_2, β, $\sigma^2_{X_1}$, $\sigma^2_{X_2}$, $\sigma^2_{x_1}$, and $\sigma^2_{x_2}$, we derive $\sigma^2_{u_1}$, $\sigma^2_{u_2}$, and σ^2_v using (19)-(21); evaluate Conditions 1a and 3a; and derive $\rho_{\Delta\Delta^*}$ and $\rho_{x_2\Delta^*}$ using (11)-(15).[21] For further details (and other settings), see Appendix C.

Results

Table 3 displays the results from these million draws within these plausible ranges. These are generally bad news for Δ_i. True, Condition 1a is never met; $\rho_{\Delta\Delta^*}$ is never actually < 0. But it is also never very high, topping .3 only about one quarter of the time, almost never topping

(Table 3 about here)

.4, and averaging only .247—low even for a reliability (a correlation between indicators of the same concept) and still lower for a correlation between *concept* and indicator.[22]

By contrast, $\rho_{x_2\Delta^*}$ is almost never *less than* .5. It usually exceeds .6 (about two-thirds of the time) and not infrequently exceeds .7 (nearly 20% of the time). It averages .634. And although the sufficient Condition 3a is met only about a quarter of the time, $\rho_{\Delta\Delta^*}$ *always* < $\rho_{x_2\Delta^*}$. In short, Δ_i fares poorly, x_{i2} much better. These results are quite robust, moreover, as can be seen in Appendix C, which varies the distribution, the α, and the parameter ranges.

Reflections

Empirical examples, mathematics, numerical simulations—everything here is of a piece:

Δ_i is not a good measure of learning; x_{i2} is much better. More precisely, $\rho_{\Delta\Delta^*}$ is rarely very large and almost always less than $\rho_{x_2\Delta^*}$. We interrogate these results in a moment. Taking them as given, however, let us first note some interesting side implications.

The first is that Converse's (1990) oft-quoted description of political knowledge as having "low mean and high variance" is incorrect. X_{it} has low mean and *low* variance. Everyone sits at or very near 0. Normed against the extremely low mean, as in the squared "coefficient of variation" (the standard deviation divided by the mean), the variance would look somewhat higher but still very low. What is true (and what Converse probably meant) is that X_{it} is a variable of low mean and high *skew*. The same is generally, though not inevitably, true of x_{it},. Item-sampling bias notwithstanding, many people congregate at the low end of the scale. The mean and variance are much higher than for X_{it} but, almost always, still very low. Across seven whole pre-deliberation DP samples we have examined with this question in view, the variance ranges from .026 to .059, averaging .039. In the 1988, 2000, and 2008 ANES, whose knowledge items are easier, the variances are .080, .062, and .077. These are in fact *lower* as proportions of their logical maximum (.25) than the corresponding means are as proportions of theirs (1.0).[23]

Second, x_{it} must hugely overerstimate X_{it}, mainly thanks to the item-sampling bias $E(e_{it}^s)$. The response-production bias $E(e_{it}^r)$ may also be positive. Yes, some knowledge may be buried in DK and partially correct responses (Mondak 1999, Mondak and Davis 2001, Krosnick, Lupia, DeBell, and Donakowski 2008, Gibson and Caldeira 2009), and, yes, respondents may not always have enough time or incentive to retrieve fugitive knowledge (Prior and Lupia 2008). On the other hand, lucky guesses (on closed-ended items) outnumber, several-fold, the incorrect and DK responses hiding knowledge (Luskin and Bullock 2011). But these are much smaller errors, at least partially canceling each other out. So, even if $E(e_{it}^{r}) < 0$, it must be far smaller in magnitude than $E(e_{it}^{s})$. Finally, the *unit* sampling bias, too, is generally positive; surveys tend to underrepresent the less knowledgeable. In sum, x_{it} must overestimate X_{it}, even setting item-sampling bias aside—and all the more so, with item-sampling bias taken into account.

Third, there nonetheless remains the issue of how to evaluate x_{it}, given the smallness of X_{it}. Everything relevant and knowable is an unapproachable standard. For evaluation, it may therefore make sense to exclude the insufficiently *salient* (readily encountered) or *useful* (capable of changing attitudes or choices). At least if this qualifying criterion is salience, its imposition should brighten our characterization of x_{it}. More salient items are likelier to be known.

Usefulness, on the other hand, is a very different criterion. Many items are salient but not useful ("soft news," anyone?). Many more are useful but not salient (where to begin?). Salience is confined to the media spotlight, usefulness far widerspread.[24] And the correlation between them, given the media's and most citizens' predilection for fluff, may be weak, even negative. Thus narrowing by usefulness may not brighten, indeed may darken our impression of x_{it}.

The distinction is especially worth making because it is usefulness that is normatively important. Ignorance of what is salient but not useful lacks ready excuse but does little harm; ignorance of what is useful but not salient is excusable but damaging. Individuals should want to align their policy preferences and votes with their values and interests, and the value of majority rule is maximized by everyone's doing so. Knowledge of the useful helps; knowledge of the salient, if not also useful, does not.

Again, such re-normings are just ways of thinking better (or worse) of x_{it}'s distribution, regarding it as higher or lower than it is, by sy applying a more approachable or more normatively discriminating standard. Willy-nilly, often unconsciously, we all do something of the sort.

Yet it is important to recognize how rough, arbitrary, and thus debatable any such adjustment must be: who is to say what is salient or useful enough for inclusion? Or how big an adjustment is therefore appropriate (and in which direction, if by usefulness)?

Reconsiderations

Can Δ_i somehow be *vindicated* (shown to be a good measure after all)? No, not really. Theoretically, Q could be made large enough to equal U. Then e_{it}^{s} and thus $E(e_{it}^{s})$ would = 0, since K'_{it} would then = K_{it}. There would be no item sampling, hence no item-sampling bias. $E(x_{it}|X_{it} = X_{Mt})$ would = X_{Mt} (still far < 1), and $b_1 = b_2 = 1$.[25] That would preclude $\rho_{\Delta\Delta^*} < 0$, making $(\beta - 1)(b_2\beta - b_1)\sigma^2_{X_1}$ in (11) positive instead of negative. Intuitively, it should also greatly increase $\rho_{\Delta\Delta^*}$ and reduce $\rho_{x_2\Delta^*} - \rho_{\Delta\Delta^*}$, since there would no longer be any x_{it} anywhere near 1, unable to show much gain. Or if Q were still $\subset U$ but randomly sampled from U, $E(n(K'_{it}))$ would $= [n(Q)/n(U)]n(K_{it})$, and $E(e_{it}^{s})$ would again = 0. The salutary consequences for Δ_i, averaging across item samples, would be the same. But these are neverland scenarios. U is always immense, and Q always a very small, highly nonrandom subset of it, implying $E(e_{it}^{s})$ far > 0.

Outside our domain, artificially small topics might contain only a handful of knowable bits. In that case, Q could equal or approach U—not by making Q unrealistically large but because U would be so small. Again this would reduce or eliminate item-sampling bias, making $E(x_{it}|X_{it} = X_{Mt}) = X_{Mt}$, and $b_1 = b_2 = 1$, although in this case X_{Mt} would presumably = 1. That again should make $\rho_{\Delta\Delta^*}$ much larger and generally > $\rho_{x_2\Delta^*}$. Or the process, regardless of topic size, could simply be uninformative, i.e., $\alpha = 0$ and $\beta = 1$ in (5), implying that nobody systematically gets richer. Intuitively, that too should greatly increase $\rho_{\Delta\Delta^*}$ and reduce $\rho_{x_2\Delta^*} - \rho_{\Delta\Delta^*}$, by undercutting the negative relationship between actual learning and the room available to show it. But again both uninformative processes and fly-speck topics are outside our domain.

More generally, the assumptions underlying our proofs are inequalities expressing tendencies: b_1, $b_2 > 1$ (item sampling bias), $\beta > 1$ (the rich getting richer), and $b_1 > b_2\beta$ (ceiling effects). They do not and need not say anything about the extent to which b_1, $b_2 > 1$, $\beta > 1$, or $b_1 > b_2\beta$ (the strength of those tendencies). Similarly, the reasoning behind b_1, $b_2 > 1$ requires only that $U \supset Q$ and $n(U)$ thus $> n(Q)$, not that $n(U) > n(Q)$ to any particular degree. This is enough to establish that $\rho_{\Delta\Delta^*}$ can be < 0, that $\rho_{x_2\Delta^*}$ is always > 0, and that $\rho_{\Delta\Delta^*}$ can be $< \rho_{x_2\Delta^*}$ (and the conditions under which the first and third of those inequalities obtain). Our simulations, suggesting how often these inequalities hold and how large $\rho_{\Delta\Delta^*}$, $\rho_{x_2\Delta^*}$, and $\rho_{x_2\Delta^*} - \rho_{\Delta\Delta^*}$ tend to be, do stipulate plausible ranges of b_1, b_2, and β. But these are wide ranges, and Appendix C shows that moving them up or down a fair bit leaves the thrust of the results unaltered—that only settings describing an uninformative process or a minute topic can make Δ_i look good.

So much for vindication: Δ_i is not a good measure. Can it be *rehabilitated*—made into one? No, not really. It can probably be improved, but not nearly enough. One largely unavailing idea is simply to ask harder items, reducing the item-sampling bias. The trouble is, the item-sampling bias can only be reduced so much, for reasons previously given. The question-writers do not know many of the hardest items, and legitimately shy away from asking the hardest they do know, lest $x_{it} = 0$ (lucky guessing aside) for all or nearly all i, obscuring absolutely small but relatively large differences in X_{it}. Bankrupting x_{it} is not a good way of bailing out Δ_i.

Another largely unavailing idea is to ask more items. *Ceteris paribus,* increasing $n(Q)$ should increase x_{it}'s reliability. Expanding a 6-item index whose reliability is .6 to 12 items can be expected to increase its reliability to .75.[26] But would that help Δ_i? Conditions 1a and 3a can be reexpressed as $\rho_{x_1x_1'} > b_1^2 b_2 \sigma_v^2 / [(\beta - 1)(b_1 - b_2\beta)\sigma_{x_1}^2]$ and $\rho_{x_1x_1'} < b_1/2b_2\beta$, but both sides of both inequalities depend on b_1, making the righthand-sides moving targets and the practical implications hard to read.[27] Our simulations, however, allow us to chart the mean $\rho_{\Delta\Delta^*}$ and mean $\rho_{x_2\Delta^*}$ as functions of $\rho_{x_1x_1'}$ (within its plausible range of $.4 \le \rho_{x_1x_1'} \le .75$). Figure 1 shows that the mean $\rho_{\Delta\Delta^*}$ increases with increasing $\rho_{x_1x_1'}$, but only slightly. It remains low, never getting much above .25.

(Figure 1 about here)

The mean $\rho_{x_2\Delta^*}$, meanwhile, holds roughly steady, well above .6. In fine, increasing $\rho_{x_1x_1'}$ does help—but not enough to help. Δ_i remains a weak measure, both absolutely and in relation to x_{i2}.

But perhaps Δ_i could be rehabilitated by "observing" differently—by defining x_{it} as somethimg other than the unweighted proportion of a fixed set of items answered correctly. One might, for example, estimate a knowledge score on the basis of an item response theory (IRT) model (as in Levendusky and Jackman 2000); weight by difficulty, giving greater credit for answering harder items correctly (as in Sood 2011); or branch the items, asking harder/easier items of those answering the previous ones correctly/incorrectly (as in Montgomery and Cutler 2013).

The first two approaches seem unlikely to make much difference. Luskin and Bullock (2006) compare the scores from a two-parameter (difficulty and discrimination) IRT model, as in Jackman (2000), with conventionally scored indices varying in such details as the treatment of DK responses. The correlations are all above .96. Weighting, for its part, does not reduce the 1's at all, may not decrease the near-1's very much, and scrunches the lower scores together—slightly increasing discrimination at the high end but slightly reducing it at the low end).[28]

Branching (as in computer assisted testing, or CAT) has more promise. By increasing the number of items (not all asked of the same respondents), it can increase the variance and discrimination of the scores. Much depends on further details—the number, focus, and difficulty of the items in the pool, the exact branching criteria, and the like (see Montgomery and Cutler 2013)—but it is important to note that limitations on question writers' expertise and respondents' patience are apt to cap item pools at the high dozens. This is no objection, only a caution about how much improvement can realistically be expected, even in x_{it}.[29]

The impact on Δ_i is likerly still smaller. Branching's (or weighting's or IRT modeling's) ability to compensate or adjust for item-sampling bias is limited: you can't include, up-weight, or branch to items you don't know enough to ask. And, as we have seen, increasing x_{it}'s reliability, a common aspiration (and criterion) of CAT, only slightly boosts Δ_i's correlation with Δ_i^*.

Conclusion

Together, the oversampling of easy items, easily neared or hit ceilings on x_{it}, and the tendency of the most knowledgeable to learn the most make Δ_i a weak measure of Δ_i^*. A simple fix, not requiring any new or additional data, is to switch to x_{i2}. More elaborate fixes, either modeling Δ_i's relation to Δ_i^* or applying other, more refined measures of Δ_i^*, may also avail. For now, we have sketched and explained the problem and indicated a simple way of sidestepping it.

Figure 1

$E(x_{i1}|X_{i1})$, $E(x_{i2}|X_{i1})$, and $E(\Delta_i\,|X_{i1})$

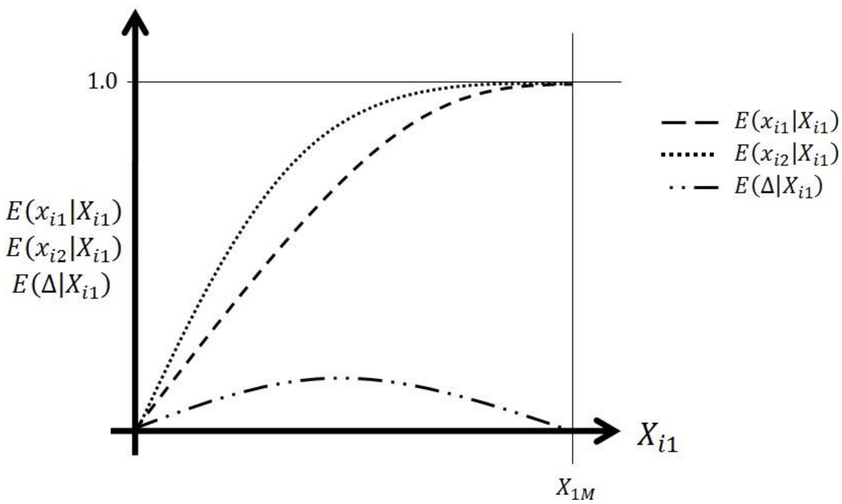


Table 1
Learning and Attitude Change in Two DPs

	Const. Ref. (Australia)†		Crime (U.K.)									
	Intended Vote (+)		Punishment (–)		Police (–)		Procedural Rights (–)		Social Root Causes (+)		Self-Protection (–)	
	Δ_i	x_{i2}	Δ_i	x_{i2}	Δ_i	x_{i2}	Δ_i	x_{i2}	Δ_i	x_{i2}	Δ_i	x_{i2}
γ_1	1.6*	1.98*	.01	-.07*	-.01	-.10*	-.02	-.09*	.04	-.01	.01	-.06*
γ_2	-3.94*	-4.20*	-.18*	-.21*	-.29*	-.32*	-.39*	-.37*	-.44*	-.45*	-.53*	-.54*
R^2	.47	.47	.05	.07	.15	.15	.19	.21	.33	.33	.24	.25
n	298	298	296	296	296	296	295	295	296	296	296	296

Note: Estimated intercepts not reported. The parenthetical signs in the column headings are those of $\bar{P}_2 - \bar{P}_1$ and thus the expected signs of γ_1.
*p < .05, one-tailed.
†Maximum likelihood estimates of ordered logit model and pseudo-R^2s.

Table 2
Parameter Ranges and Other Settings

A. Ranges

	As Applied		More Generally	
	m	*M*	*m*	*M*
X_{M1}	.0001	.000001	X_{M1m}	X_{M1M}
ϕ_1	0.8	0.95	ϕ_{1m}	ϕ_{1M}
ϕ_2	0.5	MIN(ϕ_1,0.9)	ϕ_{2m}	MIN(ϕ_1, $\phi_{2M}{}^{\dagger}$)
σ_v^2	$\frac{1}{3}\beta^2\sigma_{X_{i1}}^2$	$\frac{3}{2}\beta^2\sigma_{X_{i1}}^2$	$\frac{1}{3}\beta^2\sigma_{X_{i1}}^2$	$\frac{3}{2}\beta^2\sigma_{X_{i1}}^2$
$\sigma_{u_1}^2$	$\frac{1}{3}b^2Var(X_{i1}\|A,B)$	$\frac{3}{2}b^2Var(X_{i1}\|A,B)$	$\frac{1}{3}b^2Var(X_{i1}\|A,B)$	$\frac{3}{2}b^2Var(X_{i1}\|A,B)$
$\sigma_{u_2}^2$	$\frac{1}{3}b^2\beta^2Var(X_{i1}\|A) + b^2\sigma_v^2$	$\frac{3}{2}b^2\beta^2Var(X_{i1}\|A) + b^2\sigma_v^2$	$\frac{1}{3}b^2\beta^2Var(X_{i1}\|A) + b^2\sigma_v^2$	$\frac{3}{2}b^2\beta^2Var(X_{i1}\|A) + b^2\sigma_v^2$

B. X_{i1} Descriptive Statistics Settings for Chi-Squared Distribution

Parameter	Value
k	5
d	0.99

NOTE: As in the text, *m* and *M* distinguish plausible minima and maxima.

Table 3

Simulation Results

A. Piecewise

Relative Frequencies

Range	$\rho_{\Delta\Delta^*}$	$\rho_{x_2\Delta^*}$	$\rho_{x_2\Delta^*} - \rho_{\Delta\Delta^*}$
-.1 to 0	.000	-	.174
0 to .1	.003	.000	.495
.1 to .2	.021	.000	.253
.2 to .3	.147	.006	.066
.3 to .4	.527	.221	.011
.4 to .5	.297	.711	.001
.5 to .6	.006	.062	.000
.6 to .7	.000	.000	.000
.7 to .8	.000	.000	.000
.8 to .9	.000	.000	.000
.9 to 1.0	.000	.000	.000

Summary Results

Mean $\rho_{\Delta\Delta^*}$	.365
Mean $\rho_{x_2\Delta^*}$	.463
Proportion ($\rho_{x_2\Delta^*} > \rho_{\Delta\Delta^*}$)	.874

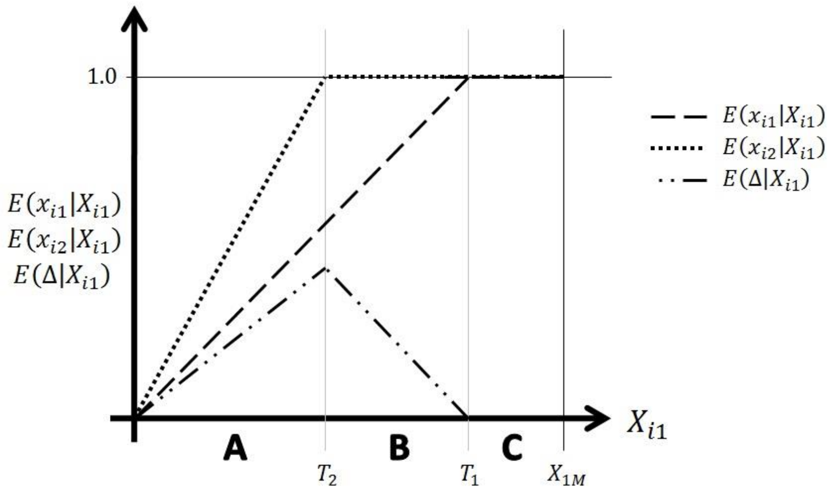
B. Quadratic

Note: Ranges are closed at the bottom and open at the top (e.g. from .1 to .199 …).

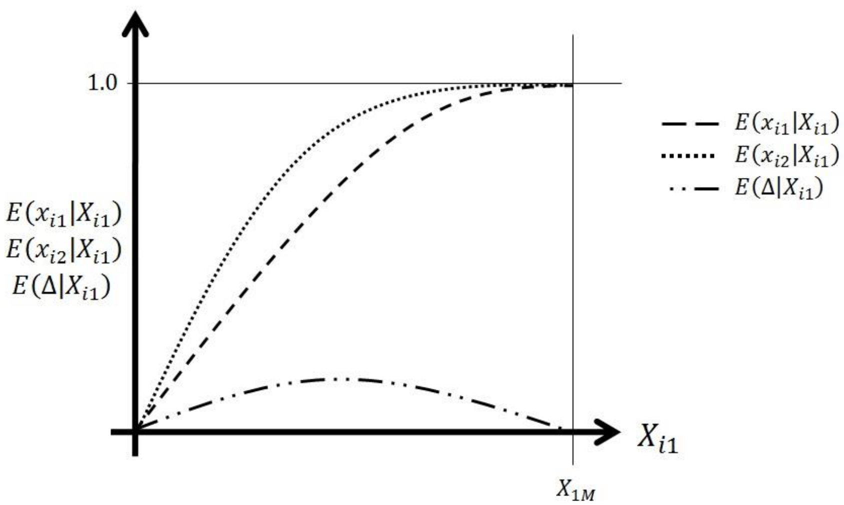
Figure 1

Graphs of the two models

A. Piecewise linear



B. Quadratic



References

Althaus, Scott. 2003. *Collective Preferences in Democratic Politics: Opinion Surveys and the Will of the People*. New York: Cambridge University Press.

Bartels, Larry M. 1996. “Uninformed Votes: Information Effects in Presidential Elections.” *American Journal of Political Science* 40: 194-230.

Cooke, Nancy J., Robert S. Atlas, David M. Lane, and Robert C. Berger. 1993. “Role of high-level knowledge in memory for chess positions.” *American Journal of Psychology,* 106, 321–351.

Converse, Philip E. 2000. “Assessing the Capacity of Mass Electorates.” *Annual Review of Political Science,* 3: 331-353

Converse, Philip E. 1990. “Popular Representation and the Distribution of Information.” In J. A. Ferejohn and J. H. Kuklinski, eds., *Information and Democratic Processes*. Urbana: University of Illinois Press.

Delli Carpini, Michael X. and Keeter, Scott W. 1996. *What Americans Know About Politics and Why It Matters*. New Haven: Yale University Press.

Eckhardt, Beverly B., Mary R. Wood, and Robin S. Jacobvitz. 1991. “Verbal ability and prior knowledge: Contributions to adults’ comprehension of television.” *Communication Re search,* 18, 636–649.

Fuchs, Marek. 2008. Total Survey Error (TSE). In Paul J. Lavrakas (ed.), Encyclopedia of Survey Research Methods. Thousand Oaks, CA: Sage Publicatios, Inc.

Gibson, James L., and Gregory A. Caldeira. 2009. “Knowing the Supreme Court? A Reconsideration of Public Ignorance of the High Court.” *Journal of Politics* 71: 429–41

Gilens, Martin. 2001. Political Ignorance and Collective Policy Preferences. *The American Polit-*

ical Science Review, *95*(2), 379–396.

Gulliksen, Harold. 1987. *Theory of Mental Tests*. L. Erlbaum Associates. Hillsdale, N.J.

Hambrick, Donald Z. 2003. "Why are some people more knowledgeable than others? A longitudinal study of knowledge acquisition." *Memory & Cognition*, 31: 902-917.

Jackman, Simon. 2000. ''Estimation and Inference are Missing Data Problems: Unifying Social Science Statistics via Bayesian Simulation.'' *Political Analysis* 8:307–332.

Krosnick, Jon A., Arthur Lupia, Matthew DeBell, and Darrell Donakowski. 2008. Problems with ANES Questions Measuring Political Knowledge. ANES Report, March 2008

Lupia, Arthur. 2006. How Elitism Undermines the Study of Voter Competence. *Critical Review*, 18 (1-3): 217-232.

Luskin, Robert C., James S. Fishkin, and Roger Jowell. 2002. "Considered opinions: Deliberative Polling in Britain." *British Journal of Political Science*, 32: 455-487.

Luskin, Robert C., Kyu S. Hahn, and James S. Fishkin. 2013. "Deliberation and Net Attitude Change." Presented at the annual General Conference of the European Consortium for Political Research, Bordeaux, France, September 4-7, 2013.

Luskin, Robert C., James S. Fishkin, Ian McAllister, John Higley, and Pamela Ryan. 2000. "Information effects in referendum voting: evidence from Australian Deliberative Poll." Paper presented at the Annual meeting of the American Political Science Association, Washington, D.C.

Luskin, Robert C. and John G. Bullock. 2006. Measuring Political Knowledge. Manuscript, Department of Government, University of Texas, Austin TX 78712.

Luskin, Robert C. and John G. Bullock. 2011. "Don't Know" Means "Don't Know": DK Responses and the Public's Level of Political Knowledge. *Journal of Politics*, 73 (April):

547-57.

Luskin, Robert C., Ariel Helfer, and Gaurav Sood. 2014. Guessing and the Measurement of Knowledge Gains. Manuscript, Department of Government, University of Texas, Austin, TX 78712.

Luskin, Robert C. and Gaurav Sood. 2011. Hidden Knowledge (and Veiled Ignorance): Is the Public Less (or Even More) Ignorant than We Have Thought? xxxx.

Mondak, Jeffery J. 1999. “Reconsidering the Measurement of Political Knowledge.” *Political Analysis* 8: 57–82.

Mondak, Jeffery J., and Belinda Creel Davis. 2001. “Asked and Answered: Knowledge Levels When We Will Not Take ‘Don’t Know’ for an Answer.” *Political Behavior* 23: 199–224.

Montgomery, Jacob M. and Joshua Cutler. 2013. “Computerized Adaptive Testing for Public Opinion Surveys.” *Political Analysis* 21 (2): 141-171 .

Prior, Markus, and Arthur Lupia. 2008. “Money, Time, and Political Knowledge: Distinguishing Quick Recall and Political Learning Skills.” *American Journal of Political Science* 52 (1): 169–83.

Recht, Donna R., and Lauren Leslie. 1988. “Effect of prior knowledge on good and poor readers’ memory of text.” *Journal of Educational Psychology*, 80, 16 –20.

Sood, Gaurav. 2011. A More Discriminating Way to Score: Guessing and Hardness Adjusted Scoring. Manuscript. Department of Communication, Stanford University.

NOTES

*This paper has benefitted from discussions with Larry Bartels, Henry Brady, Martin Elff, James Fishkin, Marta Fraile, James Gibson, Kimmo Grönlund, John Londregan, Markus Prior, David Sanders, Pascal Sciarini, Nicholas Valentino, and Jonathan Wand, among others. Earlier versions were presented at the annual meetings of the American Political Science Association, Washington, DC, September 2-5, 2010; the Midwest Political Science Association, Chicago, IL, March 31-April 3, 2011; the Society for Political Methodology, July 28-30, 2011; and the annual Joint Sessions of the European Consortium for Political Research, Salamanca, Spain, April 10-15, 2014.

[1]We take the bulk of the information to be at least debatably correct. *Dis*informational processes, replacing knowledge with misinformation, may lead to knowledge losses.

[2]For economy, we use "items" to refer both to questionnaire items and the knowable bits they are asking about, a relatively innocuous simplification. Some questionnaire items may refer to more than one knowable bit, making $n(Q)$ a slightly larger fraction of $n(U)$ (with U comparably redefined to have larger, hence fewer elements). But the fraction is always tiny.

[3]Thus Δ_i is generally much larger than Δ_i^*, except when $x_{i1} = 1$, making Δ_i necessarily 0.

[4]This resembles but differs from familiar partitionings of "total survey" error (nicely reviewed in Fuchs 2008). Those refer to the whole sample, and the "sampling error" thus to unit sampling. Here we focus on just the *i*th observation and on item sampling. As remarked below, unit sampling affects the sample mean, over i, of the x_{it}, but it is item sampling, affecting any given x_{it}, that is the key to understanding x_{it} in relation to X_{it} and Δ_i in relation to Δ_i^*.

[5]If Q equaled U, e_{it}^s would = 0, since K_{it}' would then = K_{it}, and if Q were randomly sampled from U, $E(n(K_{it}'))$ would = $[n(\mathrm{Q})/n(U)]n(K_{it})$, and $E(e_{it}^s)$ would thus = 0.

[6]For closed-ended items, constructed so that one and only one response option is correct, taking x_{ijt} as binary is not an approximation at all. Even for open-ended ones, which elicit some (Gibson and Caldeira 2009) but apparently not often verymany (Luskin and Bullock 2011) partially correct responses, it is not too far off. For X_{imt}, it just means taking the elements of U to be small enough to be indiscerptible. One either knows or does not know that Mitch McConnell is a U.S. Senator, despite, in many cases, partially knowing who Mitch McConnell is (knowing some but not all of his numerous attributes) (Luskin and Bullock 2011).

[7]X_{it}, given $n(U)$, *is* virtually continuous; x_{it}, given $n(S)$, much less so. But treating x_{it} as continuous is still a useful pretense in laying out our hypotheses and models.

[8]The rich, i.e., get richer. The poor also get richer, but less so. This recalls the sociological and economic tendencies known as "accumulated advantage," or the "Matthew Effect," after Matthew (25:29), although Matthew also has the poor getting poorer, which is not true of learning.

[9]The poor may not get poorer, but those who have nothing get nothing.

[10]Note that $E(x_{i1}|X_{i1} = X_{M1})$ is not x_{M1}, the maximum attained x_{it} in the population, which only < 1 in the vanishingly unlikely event that nobody (including any of the question writers themselves!) knows all the answers to the questions the question writers have posed or the marginally likelier event that an item amounts to a trick question so tricky that not one of the people knowing all the other items in Q gets it right. Nor is it the maximum x_{i1} in the (unit) sample, which depends on the luck of the draw and may be < 1 and either $<$ or $> E(x_{i1}|X_{i1} = X_{M1})$. Nor is it the observed value of x_{i1} corresponding to X_{M1}, which may $<$ or $> E(x_{i1}|X_{i1} = X_{M1})$, on account of u_{i1}.

[11]On average, $\delta E(x_{it}|X_{it})/\delta X_{it}$ is not merely > 0 but > 1, indeed much > 1. Its mean is the slope of the secant line from (0, 0) to $(X_{Mt}, 1)$, which $= 1/X_{Mt}$, and X_{Mt} much < 1.

[12]Since X_{i1} and X_{i2} are conceptual (undistorted by measurement error), and the knowledge-rich very relaibly tend to get knowledge-richer, the omitted disturbance may not account for very much of the variation in X_{i2} anyway.

[13]Given $E(\Delta_i|X_{i1} = X_{M1}) = 0$, (10′) yields $0 = (b\beta - b)X_{M1}, + (c\beta^2 - c)X_{M1}^2$, boiling down to $X_{M1} = -b/c(\beta + 1) = 2V_\Delta$.

[14]Statistical simulations draw instead from distributions of the disturbance(s) and exogenous regressors, which, for given parameter values, generate multiple datasets, illuminating the repeated-sampling behavior of estimators. Here we are not considering estimation.

[15]The m's and M's subscripting q, ψ, and φ_t have a slightly different meaning than those subscripting the parameters. They are discrete values—ingedients in the ends of the parameter ranges but not the ends of ranges themselves. We draw no q's between q_m and q_M. They just help define $\sigma_{x_1m}^2$ and $\sigma_{x_1M}^2$, between which we draw values of $\sigma_{x_1}^2$. See below.

[16]The corresponding ordinates are $-b^2/4c$ for $E(x_{i1}|X_{i1})$ and $E(x_{i2}|X_{i1})$ and $-(\beta - 1)b^2/4(\beta + 1)c$ for $E(\Delta_i|X_{i1})$. Note that $E(\Delta_i|X_{i1})$'s vertex is lower than $E(x_{i2}|X_{i1})$'s = $E(x_{i1}|X_{i1})$'s.

[17]Define $D = 0$ for $X_{i1} = 0$ and $= 1$ for $X_{i1} \sim U(0, X_{M1})$. Since, generically, $W \sim U(a, b)$ has a mean of $E(W) = (b + a)/2$ and a variance of $V(W) = (b - a)^2/12$, the conditional means and variances of X_{i1} are $E(X_{i1}|D{=}0) = V(X_{i1}|D{=}0) = 0$ and $E(X_{i1}|D{=}1) = X_{M1}/2$ and $V(X_{i1}|D{=}1) = X_{M1}^2/12$. Thus $E(X_{i1}^2|D{=}1) = V(X_{i1}|D{=}1) + (E(X_{i1}|D{=}1))^2 = X_{M1}^2/12 + X_{M1}^2/4 = X_{M1}^2/3$, while of course $E(X_{i1}^2|D{=}0) = 0$. Overall, therefore, $E(X_{i1}) = (1 - q)(X_{M1}/2) + q(0) = (1 - q)X_{M1}/2$, $E(X_{i1}^2) = (1 - q)(X_{M1}^2/3) + q(0) = (1 - q)(X_{M1}^2/3)$, and $V(X_{i1})$ (denoted in the text as $\sigma_{X_1}^2$) $= E(X_{i1}^2) - (\mathrm{E}(X_{i1}))^2 = (1 - q)(X_{M1}^2/3) - (1 - q)^2 X_{M1}^2/4 = [X_{M1}^2(1 - q)(1 + 3q)]/12$.

[18]The stylized distribution has more 0's, fewer near-0's, and more near-X_{M1}'s than the actual one.

[19]Since $\sigma^2_{x_1M}$ is largest for $\sigma^2_{X_t} = \sigma^2_{X_tM}$, one last constraint is $(1+\varphi_{1M})b_1^2\ [(1-q_m)(1+3q_m)]/12\, b_1^2 = (1+\varphi_{1M})[(1-q_m)(1+3q_m)]/12 \leq .25$. The assigned values of φ_{1M} and q_m satisfy this (and $\sigma^2_{x_1M}$ is no worry, since $\sigma^2_{x_2} < \sigma^2_{x_1}$). See Appendix C for further discussion.

[20]The over- and under-shooting stem from $|N(0,1)|$'s being unbounded above. The translations are defined so as to ensure that only 100α% of the draws will correspond to $b_1 < b_{1m}$, $b_2 < b_{1m}$, $\beta > \beta_M$, $\sigma^2_{X_1} > \sigma^2_{X_1M}$, $\sigma^2_{X_2} > \sigma^2_{X_2M}$, $\sigma^2_{x_1} > \sigma^2_{x_1M}$, or $\sigma^2_{x_2} > \sigma^2_{x_2M}$. Those rare draws are simply discarded.

[21]It would be more conventional to fix $\sigma^2_{u_1}$, $\sigma^2_{u_2}$, and σ^2_v (along with $\sigma^2_{X_1}$, b_1, b_2, and β) and then derive $\sigma^2_{X_2}$, $\sigma^2_{x_1}$, and $\sigma^2_{x_2}$, but in absolute terms we have a much better sense of the variances of the x's and X's than of the u's and v (although we do have and invoke some sense of $\sigma^2_{u_1}$'s, $\sigma^2_{u_2}$'s, and σ^2_v's *relative* ranges in setting the ranges of $\sigma^2_{X_2}$, $\sigma^2_{x_1}$, and $\sigma^2_{x_2}$).

[22]Under standard (random-error) assumptions, $\rho_{xx'} \leq \rho_{xX}$, where x and x' are generic indicators of the generic concept X. Thus Δ_i's correlating with Δ_i^* at .247 is even less pleasing than its having a reliability of .247 would be.

[23]The standard deviation may be more appropriate than the variance. But it, too, is not distinctly greater as a proportion of its maximum (of .5) than the mean as a proportion of its maximum.

[24]Learning one obscure fact may seldom affect preferences, but the test of utility is not whether it could do so in a vacuum but whether it could do so if one knew enough else to contextualize it.

[25]$X_{M2} > X_{M1}$, despite $E(x_{i2}|X_{i2} = X_{M2})$ proportionally $> E(x_{i1}|X_{i1} = X_{M1})$.

[26]Lengthening the index from J to K ($> J$) items makes the expected reliability of the M-item

index $s\rho_{x_1x_1'} \big/ (1+(s-1)\rho_{x_1x_1'})$, where $\rho_{x_1x_1'}$ is the reliability of the *J*-item index, and s = *K*/*J*. This is the well-known Spearman-Brown formula (see, e.g. Gulliksen 1987).

[27]*Ceteris paribus*, a higher b_1 means a higher $\rho_{x_1x_1'}$ ($= b_1^2\sigma^2_{X_1} / \sigma^2_{x_1}$) but also a lower $\sigma^2_{x_1}$, a higher $b_1^2\, b_2\, \sigma^2_v/[(\beta - 1)(b_1 - b_2\beta)\sigma^2_{x_1}]$, and a higher $b_1/2b_2\beta$.

[28]Take an index of *J* items, ordered by difficulty, where the *j*th (j = 1, 2, …, J) is ($d(j$ – 1) +1) times harder than the first ($d \geq 0$), making the *J*th and hardest ($d(J$ – 1) + 1) times harder than the first and easiest. If J = 10, and d = 1/3, e.g., the second, third, and tenth items are 4/3, 5/3, and 12/3 = 4 times harder than the first. If the items are strictly Guttman-scalable, meaning that any-one answering the *j*th ($j \geq 2$) item correctly answers all the j – 1 easier ones correctly but may or may not answer any of the 10 – j harder ones correctly, the weights, constrained to sum to 1, are .040, .053, .067, .080, .093, .107, .120, .133, .147, and .160, and the scores for answering 0-10 of the items correctly become 0, .04, .093, .160, .240, .333, .440, .560, .693, .840, and 1. Those an-swering all 10 items correctly still score 1; those answering all but one correctly now score .84 instead of .9; and those answering 0, 1, or 2 correctly now score 0, .04, .093 instead of 0, .1, .2.

[29]It may help to overlay branching with weighting. The *M* items answered correctly by respond-ent A may on average be easier or harder than the *M* answered correctly by respondent B, if the sequences of items they answer correctly and incorrectly are not the same. Weighting can take account of that. Unweighted scores of *M*/*J* become instead a scatter around *M*/*J*.