Guessing and Forgetting: A Latent Class Model for Measuring Learning*

Ken Cor† and Gaurav Sood‡

*We thank Ed Haertel for his advice and encouragement, Robert Luskin for improving some of the language, and Pablo Barberá, Ying Cui, Kabir Khanna, Marc Meredith, and Pete Mohanty for useful comments. For data and scripts for replicating the simulations and the empirical analyses, see Cor and Sood (2016).

†Mathew Kenneth (Ken) Cor is the Director of Assessment in the Faculty of Pharmacy and Pharmaceutical Sciences at the University of Alberta, Edmonton. He can be reached at mcor@ualberta.ca.

‡Gaurav Sood is an independent researcher. He can be reached at gsood07@gmail.com.

ABSTRACT

Guessing on closed-ended knowledge items is common. Under likely to hold assumptions, in presence of guessing, the most common estimator of learning, difference between pre- and post-process scores, is negatively biased. To account for guessing related error, we develop a latent class model of how people respond to knowledge questions and identify the model with the mild assumption that people do not lose knowledge over short periods of time. A Monte Carlo simulation over a broad range of informative processes and knowledge items shows that the simple difference score is negatively biased and the method we develop here, unbiased. To demonstrate its use, we apply our model to data from Deliberative Polls. We find that estimates of learning once adjusted for guessing are about 13% higher. Adjusting for guessing also eliminates the gender gap in learning, and halves the pre-deliberation gender gap on political knowledge.

1 Introduction

A nontrivial number of participants appear to *lose* knowledge over informative processes.[1] For instance, nearly 9% of the participants of Deliberative Polls,[2] which provide vetted briefing materials, and include moderated discussion, appear to go from knowing a piece of information to not knowing it over the course of the poll. Could such informative processes somehow be backfiring, perhaps by replacing information with misinformation? A far likelier explanation, we suspect, lies in the existence of guessing—lucky guessing on the pre-treatment wave, and unlucky guessing on the post-treatment wave.

More generally, guessing inflates the scores—when measured as proportion correct—of the uninformed. The less informed a respondent, the greater the positive bias in their score on average.[3,4,5] Over an informative process, where people typically know more at the end of the process than they do before it, guessing induced bias in the pre-process scores is typically greater than the bias in the post-process scores. Ergo, raw estimates of knowledge gain (post minus pre) are downwardly biased. For a few respondents guessing induced inflation in the pre-process scores (average over multiple items) can even exceed the joint total of learning, and lucky guessing on the post-process instrument.

So what can we do to correct for guessing? One reasonable place to start is 'the

[1]Our evidence for the claim comes from a broad set of cases: 1) Over an experiment that gave students opportunity to learn basic algebra, roughly 10% of the students appear to go from knowing the answer to not knowing it (Cor 2012), 2) Between 6-10% of respondents appear to 'forget' what they know, sometimes across closely spaced waves overlapping periods when correct pertinent information is being widely disseminated (Lenz 2009), 3) Nearly 9% of Deliberative Poll participants suffer the same fate over the course of a poll.

[2]Deliberative Polls bring together a random sample to deliberate about an issue or set of issues. Participants are provided with balanced briefing materials in advance of deliberations and an opportunity to quiz experts (for greater detail, see Fishkin 2009; Luskin et al. 2002).

[3]Guessing by the ignorant contributes solely to error, while guessing by those with related knowledge, under nominal assumptions, increases proportion correct proportional to the amount of related knowledge. In our discussion here, when we refer to guessing, we mean guessing by those who don't know.

[4]Note that when there is an opportunity to narrow options without use of substantive knowledge, guessing by the ignorant can net correct answers at better than chance levels. For instance, asked to identify the person in a photo, respondents can take gender of the person in the photo into account when selecting between options.

[5]The same point is stated as an assumption on pg. 105 in Sniderman et al. (1993).

standard correction,' which assumes that the probability of getting an item correct when guessing is squarely a matter of chance, estimates the true number of guesses based on the number of incorrect responses, and docks appropriately from the observed score. The technique, however, fails to account for differential opportunities to narrow alternatives across items ($\sim$ eliminate non-diagnostic red herrings) before guessing (Nunnally 1967; Lord 1975). Many variations of the standard correction have been proposed to account for non-uniform chance of lucky guessing across items (see, for instance, Hansen et al. 1975). But all of the methods rely on strong assumptions about the data generating process.

A more recent strategy has been to model responses using a three-parameter Item Response Theory (IRT) model (see Lord et al. 1968). The model, rather than assume that the ignorant guess at random, estimates an item-level conditional probability of getting an item correct given the respondent lacks "ability." In doing so, the model accounts for item idiosyncrasies, such as presence of red herrings in response options, that can lead the ignorant to guess correctly at rates very different from chance. IRT models, however, assume that 'ability' is unidimensional, contrary to what some evidence in political science suggests (see, for instance, Iyengar 1990; Stolle and Gidengil 2010). And while multi-dimensional IRT models can be estimated, such models generally need lots of items and respondents for stable estimates; most surveys on politics carry but a few political knowledge questions.

More importantly, perhaps, all these measurement models typically estimate (and correct for) guessing using data from a single wave (which covers a vast majority of the cases). However, in places where we have data from two or more waves, better methods for accounting for guessing may exist. In this paper, we develop a new method for accounting for guessing for such data; we estimate item level guessing parameters by exploiting informative within-item transitions across waves. We conduct a Monte Carlo simulation over a broad range of informative processes to assess how our method compares to the conventional estimator. We find that the conventional estimator is negatively biased and estimates have a large mean squared error. And that our method yields unbiased estimates with a low mean squared error. We next apply our model to data from the Deliberative Polls. We find that estimates of the proportion of people who learn various specific pieces of information, once adjusted for guessing, are considerably higher. Relatedly, we find that accounting for guessing removes the the gender gap in learning and halves the pre-treatment gender gap on political knowledge.

2 Traditional Pre-Post Estimates of Learning

For reasons to do with their popularity, we focus on closed-ended questions. To better reflect the nature of the data we analyze later, and given items with a 'don't know' option can easily be translated to 'forced choice' data— one need only treat 'don't know' as 'don't know' (see Luskin and Bullock 2011; Luskin and Sood 2012; Sturgis et al. 2008)— we focus on closed-ended items that carry a 'don't know' option. (For completeness, we also describe a model of responses on items that don't offer a 'don't know' option in Appendix A.) We further assume that there is no missing data (though no changes to the logic need to be made if the data are missing completely at random).

Let '0' indicate an incorrect answer, '1' a correct answer, and 'c' a 'don't know' response. Additionally, let i denote a response to a knowledge item (0, 1, or c), and let x_{ii} denote the proportion of people who display the subscripted response pattern on a particular item across the two administrations. Responses to an item over two waves can be described as follows:

Table 1: Relationship Between Manifest Responses Across Two Waves

		Post		
		0	1	c
Pre	0	x_{00}	x_{01}	x_{0c}
	1	x_{10}	x_{11}	x_{1c}
	c	x_{c0}	x_{c1}	x_{cc}

The conventional estimator of knowledge gain on any single item is the difference between the proportion of people who got the item right on the posttest $(x_{01} + x_{11} + x_{c1})$ and the proportion of people who got the item right on the pretest $(x_{10} + x_{11} + x_{1c})$. Simplifying, the conventional estimate of learning for a single item is the difference between the proportion of people moving from an incorrect or 'don't know' response to a correct response, and the proportion of people moving from a correct response to either an incorrect or a 'don't know' response or $(x_{01} + x_{c1} - x_{10} - x_{1c})$.

The 'raw' (conventional) estimator, however, may be a biased estimator of the quantity of interest—proportion of people who actually learned the item between the two waves. Inference about the quantity of interest can be compromised in various ways. For instance, if the observed 0 to 1 transition reflects a transition from an unlucky guess to a lucky one, counting such transitions will lead us to overestimate the amount of learning. More generally, if a greater share of correct answers come from lucky guessing by the ignorant on the pretest than on the posttest, the raw estimate will be negatively biased. In processes designed to impart information, where respondents know as much or less before the intervention than after, the proportion of correct answers attributable to lucky guessing in the pretest is liable to be either equal to or larger than in the posttest. Thus, the conventional estimator is a biased estimator of learning in informative processes where people actually learn.

Our goal is to develop an unbiased estimator of learning, the number of true transitions from ignorance to knowledge. To do that, we begin by describing how the latent states of knowledge and ignorance are connected to manifest responses.

3 Manifest Responses and Latent States

A person either knows a piece of information or doesn't. This isn't to say that someone who doesn't know a specific piece of information doesn't have any related cognitions about it. For instance, an individual may not know that Barack Obama is the president of the US but may know that he isn't the president of France. But knowing that Mr. Obama is not the president of France isn't the same as knowing who the president of the US is. We take the knowledge of a specific piece of information as the quantity of interest. (We come back to measuring related cognitions later on in the section.) Assume further that people do not know the wrong thing. This leaves knowledge and ignorance as the only possible latent states of cognition about a particular piece of information.

Let u indicate the manifest response. We expect someone who knows the specific piece of information to be able to recognize it among the few response options and mark the right answer. Hence, we assume that the conditional probability of a respondent marking the right answer given the respondent knows the answer is unity, or $p(u = 1|K) = 1$, where K denotes that the respondent actually knows the item (the latent state is knowledge). This also means that the respondent's probability of getting the question wrong, or marking 'don't know' given she knows the answer is zero: $p(u = 0|K) = p(u = c|K) = 0$.[6] Given we have only two latent states, previous results also mean that only the uninformed get an item incorrect or offer a 'don't know' response. The uninformed, however, can also guess luckily. So we have a world where some of the uninformed confess to their ignorance, marking 'don't know,' and some guess. It is easier to think of the uninformed who guess and who confess as two kinds of people, say 'Guessers,' denoted by G, $p(u = c|G) = 0$, and 'Confessors,' denoted by C, $p(u = c|C) = 1$. When the Guessers do guess, γ of their guesses are lucky, $p(u = 1|G) = \gamma$. We assume γ to be a fixed item-level parameter, constant across time.[7] Table 2 summarizes the information more compactly.

[6]It is reasonable to worry that when a 'Don't Know' option is offered, some of the respondents who know the answer will still mark 'Don't Know.' However, research suggests that there is very little 'hidden knowledge' (about 2-3%) behind 'don't know' responses on closed-ended items (Luskin and Bullock 2011; Luskin and Sood 2012; Sanchez and Morchio 1992; Sturgis et al. 2008). Claims to the contrary (Prior and Lupia 2008) have been shown to be founded on a misreading of the data —interpreting results in terms of percentage increases rather than increases in percentages (Luskin and Bullock 2011). In Appendix B, we describe a latent class model that assumes the probability of a person marking 'don't know' when holding the relevant piece of information as .03.

[7]In order for γ to vary over an informative process, the process must affect participants' ability to use non-diagnostic cues to answer closed-ended questions correctly. Leaving aside 'informative' processes focused on teaching test-wiseness, we do not expect ability to answer correctly without aid of substantive knowledge to change over the course of an informative process.

Table 2: Relationship Between Manifest Responses and Latent Classes

		Latent Classes		
		K	G	C
	$u = 0$	0	$(1 - \gamma)$	0
Observed	$u = 1$	1	γ	0
	$u = c$	0	0	1

The relationships between manifest responses and the set of latent classes described in Table 2 describe one question at a single point in time. The probability of different response patterns across two waves of measurement, given a set of underlying latent class transitions, can be gotten by taking the Kronecker Product of the pretest and posttest classification matrices:

$$\begin{array}{cc} & \begin{matrix} G & K & C \end{matrix} \\ \begin{matrix} u=0 \\ u=1 \\ u=c \end{matrix} & \begin{pmatrix} 1-\gamma & 0 & 0 \\ \gamma & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array} \otimes \begin{array}{cc} & \begin{matrix} G & K & C \end{matrix} \\ \begin{matrix} u=0 \\ u=1 \\ u=c \end{matrix} & \begin{pmatrix} 1-\gamma & 0 & 0 \\ \gamma & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array} = \tag{1}$$

$$\begin{array}{cc} & \begin{matrix} GG & GK & GC & KG & KK & KC & CG & CK & CC \end{matrix} \\ \begin{matrix} u=00 \\ u=01 \\ u=0c \\ u=10 \\ u=11 \\ u=1c \\ u=c0 \\ u=c1 \\ u=cc \end{matrix} & \begin{pmatrix} (1-\gamma)^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ (1-\gamma)\gamma & (1-\gamma) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & (1-\gamma) & 0 & 0 & 0 & 0 & 0 & 0 \\ (1-\gamma)\gamma & 0 & 0 & (1-\gamma) & 0 & 0 & 0 & 0 & 0 \\ \gamma^2 & \gamma & 0 & \gamma & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & (1-\gamma) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \gamma & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

The resulting 9×9 latent class transition matrix describes the conditional probabilities of specific response patterns given different latent class transitions. For example, given a latent transition from guess to guess (GG), the response patterns 0c, 1c, c0, c1, and cc are impossible. The probabilities of 00, 01, 10 and 11 response patterns given an underlying guess-guess transition are represented by the remaining entries in the first column of the 9×9 matrix.

Next, the latent class transition matrix is multiplied by a vector of unknown latent class transition parameters to define the system of equations that describe the multinomial distribution of observed response patterns. There are nine latent class transitions that could hypothetically occur: GG, GK, GC, KG, KK, KC, CG, CK, and CC.

To identify the item-level conditional probability of getting an item correct if the respondent chooses to guess, we make the same assumptions as above —that the respondent who knows the right answer doesn't forget it over the course of the informative process, and that they don't become misinformed. Under these assumptions, the proportion of people moving from knowing to guessing (KG) and confessing to ignorance (KC) is zero.

Multiplying the 9×9 matrix from equation 1 by the latent class transition vector, λ, and setting λ_{KG} and λ_{KC} to zero, we get a system of equations, relating latent parameters to expected response pattern proportions, π:

$$
\begin{array}{r}
\\
u=00 \\ u=01 \\ u=0c \\ u=10 \\ u=11 \\ u=1c \\ u=c0 \\ u=c1 \\ u=cc
\end{array}
\begin{array}{c}
\begin{array}{ccccccccc} GG & GK & GC & KG & KK & KC & CG & CK & CC \end{array} \\
\begin{pmatrix}
(1-\gamma)^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
(1-\gamma)\gamma & (1-\gamma) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & (1-\gamma) & 0 & 0 & 0 & 0 & 0 & 0 \\
(1-\gamma)\gamma & 0 & 0 & (1-\gamma) & 0 & 0 & 0 & 0 & 0 \\
\gamma^2 & \gamma & 0 & \gamma & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & \gamma & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & (1-\gamma) & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \gamma & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}
\begin{pmatrix}
\lambda_{GG} \\ \lambda_{GK} \\ \lambda_{GC} \\ \lambda_{KG}=0 \\ \lambda_{KK} \\ \lambda_{KC}=0 \\ \lambda_{CG} \\ \lambda_{CK} \\ \lambda_{CC}
\end{pmatrix} \tag{2}
$$

$$
= \begin{pmatrix}
(1-\gamma)^2\lambda_{GG} \\
(1-\gamma)\gamma\lambda_{GG} + (1-\gamma)\lambda_{GK} \\
(1-\gamma)\lambda_{GC} \\
\gamma(1-\gamma)\lambda_{GG} \\
\gamma^2\lambda_{GG} + \gamma\lambda_{GK} + \lambda_{KK} \\
\gamma\lambda_{GC} \\
(1-\gamma)\lambda_{CG} \\
\gamma\lambda_{CG} + \lambda_{CK} \\
\lambda_{CC}
\end{pmatrix}
= \begin{pmatrix}
\pi_{00} \\ \pi_{01} \\ \pi_{0c} \\ \pi_{10} \\ \pi_{11} \\ \pi_{1c} \\ \pi_{c0} \\ \pi_{c1} \\ \pi_{cc}
\end{pmatrix}
$$

Based on the model, the proportion of respondents who learned an item is simply the sum of the estimates of the two transition parameters that track movement from ignorance to knowledge, λ_{GK} and λ_{CK}.[8]

3.1 *Related Knowledge*

As we note above, a person either knows a particular piece of information, or he doesn't. A person, however, may have any number of relevant cognitions related to the piece of information. For instance, a person may not know that Pakistan has nuclear weapons, but may know that Poland doesn't. And over the course of an informative process, the repository of related cognitions may grow even while the respondent never learns the particular piece of information. For instance, a respondent may learn that Japan does not have nuclear weapons, but still not learn that Pakistan does. Growth in such related cognitions may be important to the researchers. And given how we motivate the model, it is not immediately clear whether the model we propose captures such learning.

Closed-ended items are, naturally, limited instruments for capturing learning of related cognitions. The only kind of related knowledge that they can capture is growth in cognitions in one of the options in the response set. Staying with the example we use above, if Japan is within the response set of a closed-ended item (with a single correct answer) asking the respondent which of the countries has nuclear weapons, the respondent who has learned that Japan doesn't have nuclear weapons (but is still ignorant of the fact that Pakistan has nuclear weapons) is likelier to pick Pakistan, than another similar respondent without that piece of related knowledge. Such learning is liable to lead to an increase in proportion correct.

[8]Rather than estimate γ using the above model, one can estimate γ under other assumptions, or using alternate methods. The value of γ can be plugged in to the equation and then used to estimate the other unknown parameters. For instance, one can use estimate of γ from the standard guessing correction, inverse of the number of options, or the 3-PL IRT model. Note, however, that data with 'don't know' don't lend themselves to correct estimation of γ in the conventional 3-PL model. Without reconstituting the conventional 3-PL model, there are two ways to handle 'don't know' responses within it—treat them as missing, or convert them to 0s. Doing either requires us to make untenable assumptions. In the first case, we must assume that 'don't know' responses are missing at random when, in fact, we know that they are near perfect indicators of ignorance (Luskin and Bullock 2011). Doing the latter means the 3-PL IRT model sees them as unlucky guesses, which they are not.

Within the multinomial distribution, this means greater number of people going from 0 (or c) to 1, and lower probabilities of going from 1 to 0 (or c). Our model doesn't disambiguate between a correct response as a result of related knowledge, and as a result of knowledge of the relevant piece of information. And as proportions of transitions from c to 1 and 0 to 1 increase, the probability of CK and GK also increase. Thus, while we motivate our model without acknowledging related knowledge, the latent class model that we have developed implicitly accounts for gains in related knowledge.

Estimation

We estimate the parameters of the system of equations that define the multinomial distribution for a single item (equation 2) via maximum-likelihood. Letting n denote the sample size, m the number of multinomial cells, x_i the observed counts in each cell i, and $p_i(\theta)$ the individual cell probabilities, the log-likelihood of a multinomial distribution is:

$$l(\theta) = log(n!) - \sum_{i=1}^{m} log x_i! + \sum_{i=1}^{m} x_i log(p_i(\theta)) \tag{3}$$

To apply this general form to our model, we need to swap x_i and $p_i(\theta)$ with u and π respectively. Given that n and $\sum_{i=1}^{m} log u_i!$ for any one item are constant, parameters are estimated by maximizing the last term in the likelihood function, with the constraint that the transition parameters (λ) (see equation 2) sum to 1:

$$\begin{aligned} \text{max.} \sum_{i=1}^{m} u_i log(\pi_i) = u_{00}((1-\gamma)^2\lambda_{GG}) + u_{01}(1-\gamma)\gamma\lambda_{GG} + (1-\gamma)\lambda_{GK}) + \\ u_{00}((1-\gamma)\lambda_{GC}) + u_{10}(\gamma(1-\gamma)\lambda_{GG}) + u_{11}(\gamma^2\lambda_{GG} + \gamma\lambda_{GK} + \\ \lambda_{KK}) + u_{1c}(\gamma\lambda_{GC}) + u_{c0}((1-\gamma)\lambda_{CG}) + u_{c1}(\gamma\lambda_{CG} + \lambda_{CK}) + u_{cc}(\lambda_{CC}) \\ \text{s.t.} \sum_{i=1}^{m} \lambda_i = 1 \end{aligned} \tag{4}$$

We perform the optimization using a general non-linear optimizer, Rsolnp (Ghalanos and Theussl 2010), that uses the Augmented Lagrange Multiplier Method. (An additional non-linear optimizer, Alabama (Varadhan and Grothendieck 2011), using Adaptive Barrier Minimization, produced virtually identical results.) The sum of λ_{GK} and λ_{CK} gives us the guessing adjusted estimate of proportion of people who learned a piece of information. We calculated the errors of the parameters (and composites of parameters) using non-parametric bootstrap. In particular, we took 100 random samples with replacement of observed response patterns and re-estimated the results. The standard deviation of these estimates was used as the estimate of the standard error.

4 Monte Carlo Study

To assess how well our model fares in recovering actual knowledge gains over an informative process, we simulated a broad range of informative processes and knowledge items used to measure learning. We simulated informative processes that varied in their efficacy (average learning) and in how the information gains were distributed. And to measure learning over these processes, we simulated knowledge items that varied widely in their difficulty and in the ease with which people who didn't know the right answer could correctly guess it.

We started with a model of the underlying data generating process. In particular, we chose a 1-PL IRT model—as our quantity of interest is learning over an item, discrimination parameters are unnecessary— which assumes whether or not a person knows a particular piece of information, $p(x = 1)$, is a stochastic function of their ability (θ), and difficulty of the piece of information (or, more precisely, the difficulty of the item used to measure it) (δ).

$$p(x = 1) = \frac{1}{1 + e^{-(\theta - \delta)}} \tag{5}$$

The stochastic component of the model is best interpreted as follows: a person of a given ability will know some pieces of information of a given difficulty and not know others. We further assume that the items used to test knowledge of the pieces of information are closed-ended, and that those who know the answer always mark the right answer.

We further assumed that the abilities had a standard normal distribution, $N(0, 1)$. From the standard normal, we randomly drew n respondents. Given what is easy for the more able is difficult for the less able, ability and difficulty are on the same scale. And since we had no prior knowledge about the typical difficulty of the items used to measure learning, we took item difficulty to be uniformly distributed. From the uniform distribution across the entire observed range of θ, we randomly drew k item difficulty parameters. Following which, we simulated responses on the pre-process wave using the 1-PL IRT model we specify above.

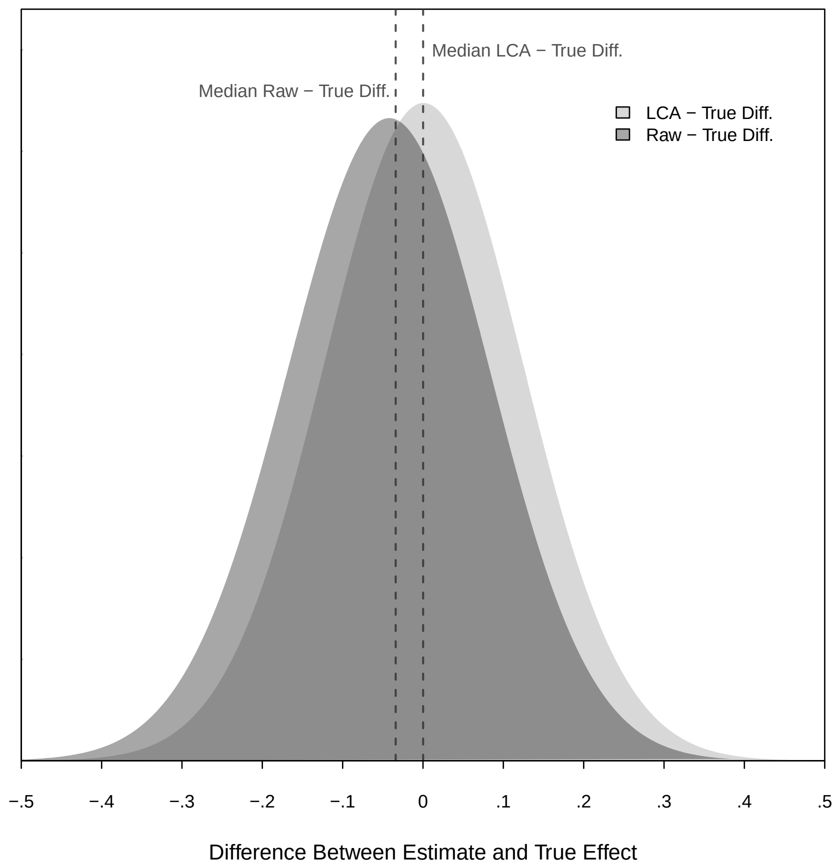
Next, we simulated a broad range of learning processes, that varied in their average efficacy, which we take to be the proportion of respondents learning the item, and in how the gains in learning were distributed. Given we were modeling an informative process, we took average gains (α) to be non-zero, ranging from 5% to a hefty 50%. And since we did not know whether less efficacious informative processes are more or less probable than more informative processes, we took α to be uniformly distributed between .05 and .50. From this uniform distribution, we randomly drew k values of α.

At one end are egalitarian informative processes in which each respondent who doesn't know the item has about an equal chance of learning it. Of course, some respondents will learn the item and some won't. Most naturalistic informative processes, however, produce learning that is correlated with pre-treatment ability (Dooling and Lachman 1971; Bransford and Johnson 1972; Eckhardt et al. 1991; Cooke et al. 1993; Hambrick 2003; Recht and Leslie 1988). We model this kind of process by systematically giving those with greater pre-treatment ability, a greater chance of learning an item. Thus we specified another parameter (β) that was multiplied to pre-process ability, going from 0 (no correlation with pre-process) to .6. Among people who did not know the item in the pre-process instrument (with $x = 0$), we drew a random binomial with probability equal to $\alpha + \beta * \theta$. Thus, learning added to pre-process responses gave us (true) post-process indicators of whether the respondents knew the item or not.

Next, we simulated confessing and lucky guessing on top of true indicators of knowledge. On items that the respondent did not know, we first simulated the choice between guessing and confessing. We assumed that the decision to guess or to confess was random, with probability of confession varying between .05 and .25 (in line with observed rates of confession to ignorance in data we use later). We assumed the chance of guessing an item correctly ranged from .1 and .6. For coverage, we again drew from a random uniform. Simulating these two processes on both waves gave us the observed responses.

As expected, the raw estimator was negatively biased (Bias = -.044). To put the number in perspective, the bias was, on average, 31.7% of the true estimate. However, using the Latent Class Analysis (LCA) model that we develop here yields nearly unbiased estimates (Bias = .001). Or, an average of less than 2% of the true estimate. Not only was our model vindicated on bias, it was also vindicated on Mean Squared Error (MSE). The MSE of the raw estimate was four times LCA's ($\text{MSE}_{\text{raw}} = .0035$, $\text{MSE}_{\text{LCA}} = .0008$). Lastly, as expected, differences between the raw estimates and actual learning were negatively correlated with item difficulty: harder the item, greater the underestimate ($r = -.67$). However, estimates from our model were uncorrelated with item difficulty ($r = -.004$).

Figure 1: Distribution of Difference Between LCA and Raw Estimates and True Effect



5 Application

Next, we apply the model to data from the Deliberative Polls. We begin by describing the data.

5.1 *Data and Measures*

Data are from 19 face-to-face and 4 online Deliberative Polls (see Fishkin 2016). (See Appendix C for details about each of the polls included in the study.) In a face-to-face Deliberative Poll, a random sample is interviewed. Some time later the entire interviewed sample or a random subset is invited to deliberate. Invitees are typically offered a small honorarium. However, not everyone who is invited comes. People who take part in deliberations tend to be more knowledgeable, and better educated than those interviewed initially (see O'Flynn and Sood 2014). But aside from that, participants are alike those who don't come on a variety of demographic characteristics.

If the site of the deliberation is non-local, travel to the site is reimbursed, and a free hotel stay is arranged. Invitees are mailed balanced briefing materials prior to deliberation. Of those who accept the invitation, typically a reasonably large proportion turns up to deliberate. At the site, respondents are randomly assigned to small groups. Deliberation in small groups is facilitated by trained moderators. Participants also get a chance to quiz experts in plenary sessions. At the end of deliberation, participants fill out a survey featuring many of the same questions they were posed initially, along with some new ones.

Online Deliberative Polls work much the same way— like face-to-face Deliberative Polls, a random sample is interviewed, and a random subset of those interviewed is invited to deliberate, and briefing materials is provided in advance of deliberations. But a few things are different in online polls. Participants cannot see each other, and rely on voice alone to communicate. And small groups are not allocated randomly but decided upon opportunistically, based on times that groups of participants find convenient.

In each of these polls, respondents' knowledge of policy relevant facts was measured using, mostly, closed-ended items. (We found only two open-ended items in the 23 polls we analyze here.) Almost always, all the knowledge items that are fielded in the pre-deliberation survey are repeated in the post-deliberation survey. In analyses below, we subset on closed-ended questions that are asked both pre- and post-deliberation. In all, we have data on 177 items. All the closed-ended items offered a 'don't know' option. For placement items, correct absolute placement is scored as correct (see Luskin and Bullock 2004). (See Appendix D for question wording, and response options.)

5.2 *Results*

5.2.1 *Learning Specific Pieces of Information*

Based on differences in proportion correct, a significant proportion of the respondents learned the pieces of information that they were surveyed on.[9] Across all 177 items in 23 polls, on average, approximately 16% of the respondents learned the piece of information they were surveyed about based on the conventional estimator of knowledge gain (Mean = .158, s.e. = .01). The estimated proportion of respondents who learned the piece of information varied widely across items ($s = .158$). Regardless, even based on the 'raw' estimator, which we expect to be negatively biased, participants learned a fair bit on items they were surveyed on.

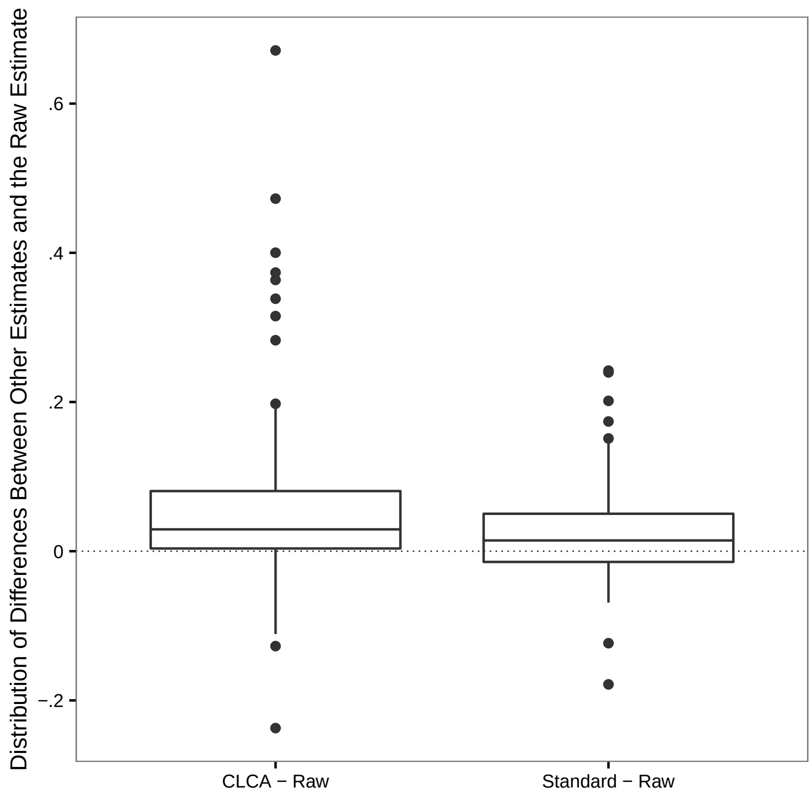
As we reason above, accounting for guessing is liable to lead us to (correctly) infer that a larger proportion of respondents learned items they were surveyed on. But before we present guessing adjusted estimates of learning, we present some evidence consistent with our heuristic account of why adjusted estimates are liable to be higher than raw estimates. Our claim rests upon the insight that when people know less, they have a greater opportunity to guess. And that people know as much, or less, before an informative process, than after. Hence, there is likely more guessing (and very likely more successful guessing) in the survey conducted before the process than after it.

Across all items, 29% of the respondents guessed incorrectly on the pre-process wave. Observable lucky guessing in the pre-process wave —percent '10' and '1c' response patterns —was about 8.7%. In all, at least 38% of the participants guessed in the pre-treatment wave. As expected, somewhat fewer respondents guessed incorrectly on the post-treatment wave (25%).

[9] Note that we lack data on how much people who didn't deliberate learned over the same time. While we think it is unlikely that people would have learned the pieces of information they were surveyed on without the Deliberative Polling treatment, it is possible that the causal effects of Deliberative Poll on knowledge are lower.

Another indicator of greater guessing related error on the pre-process wave, vis-à-vis the post-process wave, would be its lower reliability. We find as much in the data. Across the 23 polls, reliability of pre-process instrument was significantly lower than the reliability of post-treatment instrument (Mean $\alpha_{T_1} = .495$, Mean $\alpha_{T_2} = .561$, Diff. $= .065$, $p < .05$). (The p-value is based on the sign test. In 18 of 23 polls, the pre-process instrument had lower reliability than the post-process instrument.) In all, the evidence is consistent with the idea that more people guessed (luckily) on the pre-process instrument than on the post-process instrument.

Figure 2: Distribution of Differences Between Guessing Adjusted Estimates and Raw Estimates



Next, we present estimates of learning adjusted for guessing using two different methods: the standard guessing correction,[10] and the latent class model we have developed

[10]The standard guessing correction assumes that lucky guessing is squarely a matter of chance. While expositions on standard guessing correction haven't typically dealt with 'Don't Know' responses, they are

in the paper. We estimated learning using both the models for each of the 177 closed-ended knowledge items. As we had expected, once we account for guessing, the estimates of learning are larger ($\text{Mean}_{\text{LCA}} = .210$, $\text{s.e.}_{\text{LCA}} = .016$; $\text{Mean}_{\text{stnd}} = .182$, $\text{s.e.}_{\text{stnd}} = .015$). All together, estimates of learning based on the LCA were greater than the raw estimate for 78.5% of the items. And standard guessing correction based estimates of learning were larger than the raw estimates for 65% of items. The average difference between LCA and raw estimates was about 5% ($p < .01$) (See also Figure 2). (We got virtually identical results when we fixed the probability of knowing given a Don't Know response as 3% as per the model in Appendix B.) To put this in perspective, accounting for guessing using the LCA model we develop here, we find that nearly 30% more respondents learned a piece of information than what the raw estimates suggest. The average difference between estimates based on standard guessing correction and raw estimates was smaller, about 2% ($p < .01$), but substantively meaningful— in terms of percentage change, the difference was approximately 16%.[11]

5.2.2 *Heterogeneous Effects*

Next, we analyze how accounting for guessing changes inferences about how much different groups of people learn over the Deliberative Poll. In particular, given longstanding concerns about gender and guessing on knowledge questions, we analyze how inferences

easily handled. Calculations of how many items an individual guessed on only need omit 'Don't Know' responses.

Where n denotes the number of respondents, i tracks respondents, k is the number of alternatives, u_1 and u_2 are responses on the pretest and the posttest, respectively, and 1 and 0 indicate a correct and incorrect response respectively, the standard guessing correction based estimate of learning runs as follows:

$$\frac{1}{n}\left(\sum_{i=1}^{n}(u_{2_i}=1) - \frac{1}{k-1}\sum_{i=1}^{n}(u_{2_i}=0) - \sum_{i=1}^{n}(u_{1_i}=1) + \frac{1}{k-1}\sum_{i=1}^{n}(u_{1_i}=0)\right) \tag{6}$$

[11] As we note above, the latent class model implicitly accounts for gains in related knowledge. Under some implausible circumstances, however, the latent class model may miss some gains in related knowledge (these latter gains will also be missed by the conventional estimator). We discuss these implausible circumstances and empirically assess whether such circumstances are cause for concern in Appendix E. We find little evidence of these implausible circumstances compromising inference.

about how much men learn vis-à-vis women change when we adjust for guessing.

When a 'Don't Know' option is offered, rates of guessing are perforce equal or lower than when such an option is not available. However, different problems may ensue—opportunity to confess one's ignorance systematically privileges the 'kinds of people' (men, according to Mondak 1999, 2001) who choose to guess when they don't know, or choose not to be reticent when in fact they do know. (Reticence by the knowledgeable on closed-ended items has been shown to be extremely infrequent (see Sturgis et al. 2008; Luskin and Bullock 2011).) Over an informative process, differences in guessing rates can cloud estimates of subgroup differences in learning.

To account for differences in guessing by gender, we calculated the probability that an observed correct answer reflects knowledge, $P(K|u=1)$, at T_1 and at T_2 for men and women separately. To calculate this, we first calculate the item level probability of a lucky guess, $P(u=1|G)$, which is constant, followed by propensity to guess within each subgroup at each time. Next, we replaced all the 1s at $T1$ and $T2$ with the probability of knowing given an observed 1 and recalculated simple pre-post differences. This results in guessing adjusted estimates of learning that also account for sub-group differences in guessing rates.

The $P(K|u=1)$ for men and women separately at each time is calculated using a simple counting principles approach. Let n_l denote total number of lucky guesses, n_u total number of unlucky guesses, and n_g total number of items a person guessed on, the sum of n_l and n_u. The LCA model defines $P(u=1|G)$ as γ while the standard guessing correction model assumes $P(u=1|G)=1/k$, where k is the number of response options on an item. The conditional probabilities of someone picking the right answer when they don't know are then:

$$P(u=1|G)=\gamma=\frac{P(G|u=1)P(u=1)}{P(G)} \tag{7}$$

and

$$P(u=1|G) = \frac{1}{k} = \frac{P(G|u=1)P(u=1)}{P(G)} \tag{8}$$

Using these equations, and that n_g is sum of n_l and n_u, we can solve for the total number of lucky guesses in terms of γ and n_u and k and n_u. Doing so yields the following equations:

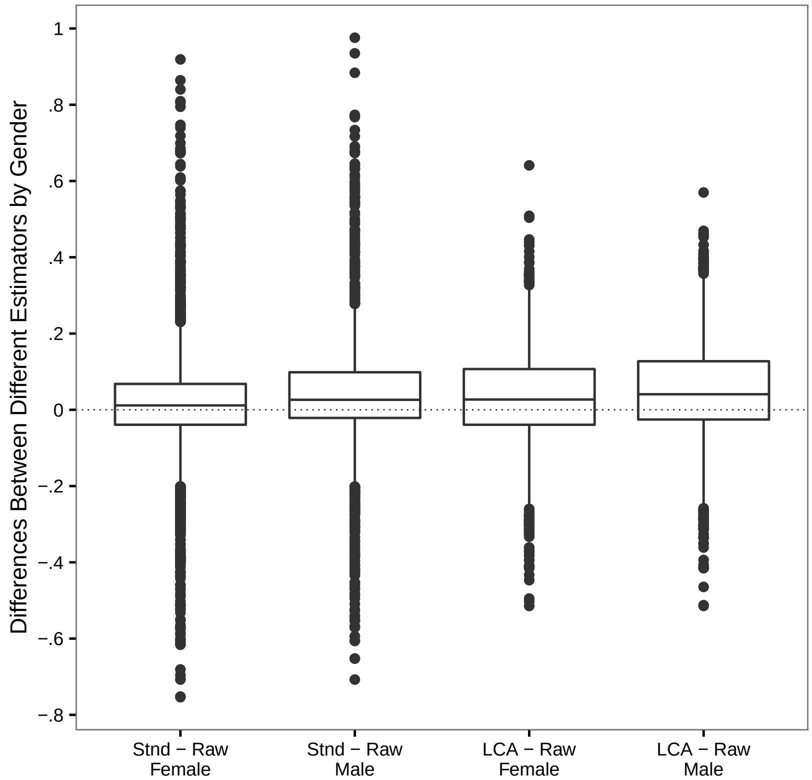
$$n_l = \frac{n_u \gamma}{1-\gamma} \tag{9}$$

$$n_l = \frac{n_u}{k-1} \tag{10}$$

Dividing the above equations by the total number of observed correct answers gives the probabilities that an observed correct answer is a lucky guess. Subtracting them from 1 produces the required probabilities $P(K|u=1)$ for the two different types of corrections.

Across all the 23 polls, based on raw estimates, women learned more than men (Mean Diff. $= -.022$, s.e.$= .006$. As Figure 3) shows, accounting for guessing leads to larger positive corrections for men than for women. Hence, once we account for guessing, the 'learning gap' nearly disappears (Mean Diff.$_{\text{LCA}} = -.007$, s.e. $= .004$; Mean Diff.$_{\text{stnd}} = -.003$, s.e. $= .005$).

Figure 3: Distribution of Differences Between Guessing Adjusted and Raw Estimates By Gender



One plausible explanation for why the learning gap between men and women disappears after we adjust for guessing is that men and women vary in their propensity to guess when they don't know. In a world where men are more likely to guess (as opposed to confess) when they don't know than women, guessing related attenuation would be greater in men's scores than women's in the conventional estimates.

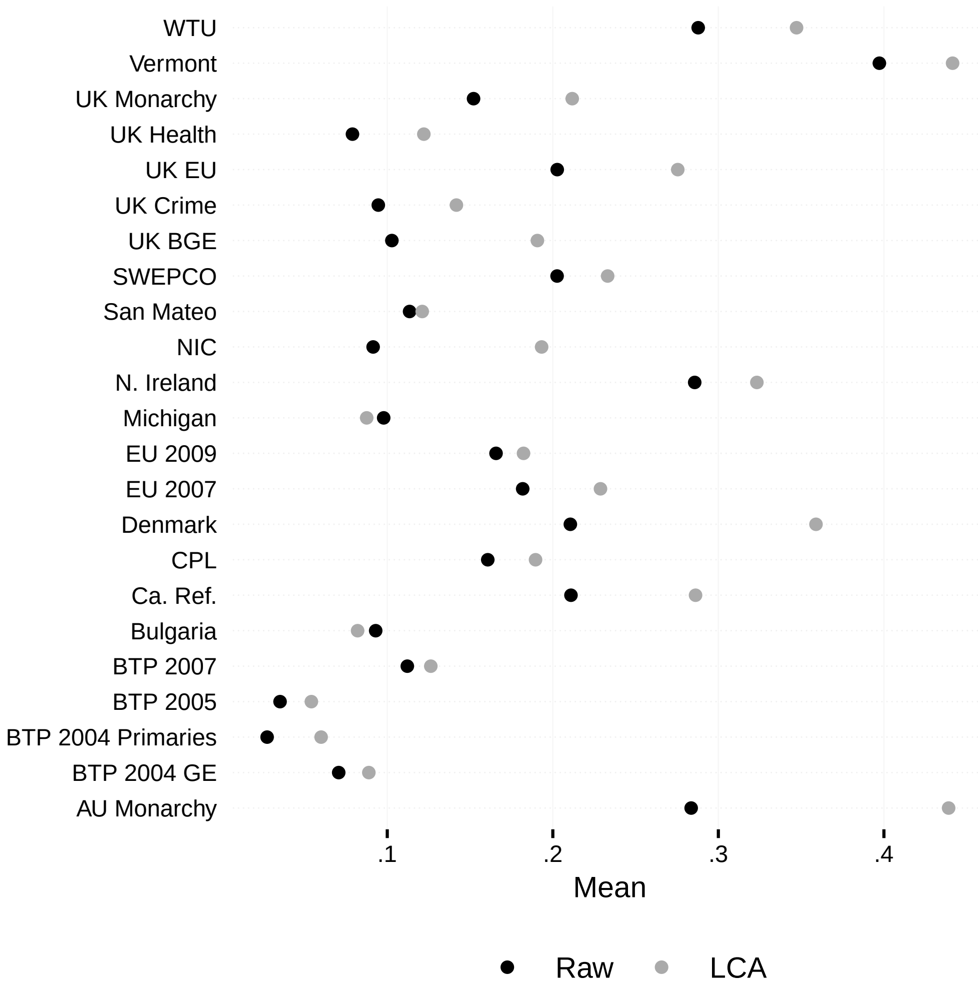
The differences that cloud subgroup differences in learning also cloud subgroup differences in knowledge at any particular wave. For instance, differences in guessing may partly explain 'knowledge gap' across genders (Verba et al. 1997; Mondak and Anderson 2004; Delli Carpini and Keeter 1996; Frazer and Macdonald 2003). To assess this possibility, we compared gender gap in conventional estimates to gap in guessing adjusted estimates on the pre-deliberation wave. On average, based on the raw scores, women knew far less than men (Mean Diff. = .067, s.e. = .005). Once you adjust for guessing, however, the knowledge gap nearly halves (Mean $\text{Diff.}_{\text{stnd}}$ = .037, s.e. = .004; Mean $\text{Diff.}_{\text{LCA}}$ = .039, s.e. = .003). (The decline in estimated gender gap when you account for guessing is itself highly statistically significant, $p < .001$.) Significantly, our results match those obtained by Mondak and Anderson 2004, who also find that "approximately 50% of the gender gap is illusory."

5.2.3 *Learning Over Polls*

One can use estimates of observed learning over specific pieces of information to infer learning over other unsampled pieces of information. Such inference rests upon one assumption—dimensionality of learning—and its generalizability is limited by biases in sampling. For dimensionality, we simply assume that learning is unidimensional. As for biases in sampling, like many other surveys, the pieces of information people are surveyed on in Deliberative Polls are considerably 'easier' (Luskin et al. 2014). This limits inference to learning on similarly easy bits of information, though see Luskin et al. (2014) for some other issues that may sabotage inference.

We use simple average of learning across items as an estimate of average learning of similar such items in a poll. On average, Deliberative Poll participants got a significantly greater proportion of items correct in surveys conducted right after the polls than in surveys conducted weeks before. While the increase over the polls —mean of item-level increases —varied considerably (Range: .027, .397), the average increase (inverse variance weighted) was a hefty .14 (s.e. = .003) (see F1 in Appendix F and Figure 4). Expectedly, adjusting for guessing (using either standard guessing correction or LCA), boosts the estimates of learning—average learning based on the LCA estimator is .171 (s.e. = .004; $p_{\text{Diff.}} < .01$), and estimate based on the standard guessing correction is .168 (s.e. = .003; $p_{\text{Diff.}} < .01$).

Figure 4: LCA and Raw Estimates of Learning by Poll



5.2.4 *Goodness of Fit*

Lastly, we check how good our model is at reproducing the observed distribution of response patterns. We find that the latent class model captures the data reasonably well. Based on the χ^2 test, the model reproduces response patterns that are consistent with the data on 83.1% of the items (see Table 3).

In all, the analyses suggest three things. First, conventional estimates of learning are negatively biased. Second, using any of the major models that adjust for guessing reduces bias. And third, adjusting for guessing using latent class models yields particularly good

Table 3: Proportion of Items in Each Poll With Predicted Patterns Different From Observed Data

Poll	LCA
AU Monarchy	.3
BTP 2004 Primaries	0
BTP 2004 GE	.333
BTP 2005	0
BTP 2007	.125
Bulgaria	.143
Ca. Ref.	0
CPL	.286
Denmark	.222
EU 2007	0
EU 2009	.167
N. Ireland	.143
Michigan	.222
NIC	.125
San Mateo	.25
SWEPCO	.2
UK BGE	.067
UK Crime	0
UK EU	0
UK Health	.167
UK Monarchy	.125
Vermont	.667
WTU	.2
Weighted Average	.831

estimates of the true effect.

6 Discussion

When analyzing the impact of informative processes one basic question that we often want an answer to is "What proportion of the respondents learned a particular item over the course of an informative process?" To answer this, one needs to know how many knew the information before the process, and how many knew it right after the process. Ask once, ask twice. Rinse and repeat. However, as we show above, this may not be enough. Our ability to detect whether a respondent knows something at each time point is hindered by lucky guessing by the uninformed.

Naive estimates of learning over an intervention are biased due to guessing. And while guessing by the uninformed may be random, its effects are distinctly non-random. Guessing related error is negatively correlated with true knowledge. And not taking guessing into account leads us to underestimate learning over an informative process. Guessing may also bias inferences about differences in how much various subgroups learn over an informative process. This may happen simply because some groups of people start out knowing less than other groups of people. Or, because a group has a greater propensity to guess than another. Adjusting for guessing using any of the popular methods, such as the standard guessing correction, and IRT, reduces bias. However, the latent class model developed here is unbiased and has the lowest mean squared error.

Accounting for guessing post hoc is, however, necessarily imperfect— for any particular response, there is no way to know the 'latent state' without error. Thus, rather than account for guessing post hoc, it may be better to reduce guessing in the measurement instrument. As is readily intuited, guessing is largely a concern on closed-ended items. Thus, one cure may be to replace closed-ended items with open-ended items. However, that may be no panacea. Concerns about motivation to search memory for the answer become important on open-ended items. In particular, estimates of learning may be biased by differential motivation to recall the answer, pre- and post-process (see also, Tallmadge 1982). For instance, in the specific informative process we discussed— Deliberative Polling— it may well be the case that respondents are more motivated to recall a piece of information after having intensely discussed an issue than before deliberation. These kinds of concerns aren't material when the task is merely to recognize the correct answer from a small set of options. Separately, coding responses to open-ended questions presents its own challenges (Gibson and Caldeira 2009). But perhaps more to the point, open-ended questions do not preclude guessing. The respondent who does not know the correct answer may nonetheless guess from among a set of possibilities. For example, a respondent may not know that the president of the United States is Mr. Obama, but may be pretty sure that it is either Mr. Obama or Mr. Clinton or Mr. Bush, and guess randomly from among those possibilities.

Another, perhaps more promising, way to preempt the need for building post hoc models to account for guessing may be to include additional questions that assess respondents' certainty about their responses (see, for instance, Alvarez and Franklin 1994; Pasek et al. 2015). Measuring certainty has the virtue of expanding the scope of what can be measured in three important ways. Assume, for instance, an absolute scale on which politician's positions could be placed. Assume also that, for strategic reasons, the politician is deliberately unclear about his or her position, indicating just a range. And that the respondent, a particularly avid consumer of politics, knows the range exactly. In conventional closed-ended questions, such a respondent may be scored as uninformed. But certainty and range assessments can be used to capture such knowledge (see, for instance, Franklin 1991; Alvarez and Franklin 1994; Alvarez and Glasgow 1996). Not only that, certainty measures allow tracking of changes in certainty, which may capture changes in how easy it is for the respondent to recall the information, over an informative process (see again, Franklin 1991; Alvarez and Franklin 1994; Alvarez and Glasgow 1996).

Certainty assessments can also be used to measure misinformation. On conventional instruments, there is no way to distinguish misinformation from ignorance without making further assumptions. And, in presence of misinformation (and possibility of knowledge being replaced by misinformation), our identifying assumption would no longer hold. Supplementary questions that assess respondent's certainty about the responses, however, may provide one way to resolve the issue. One may take positions on which a respondent is certain and incorrect as indications of misinformation— belief in incorrect information (see, for instance, Pasek et al. 2015).

Including certainty measures has yet another benefit. A great deal of research suggests that certainty assessments are substantively important, capturing substantively important things like candidate strategies (Franklin 1991) and affecting respondent decision making (Alvarez and Franklin 1994; Alvarez and Glasgow 1996; Bartels 1986).

Hitherto, we have focused on concerns related to measurement of learning of specific pieces of information. However, we often want to infer how much people learned in general over an informative process. Estimates of gross learning over a rich naturalistic informative process like Deliberative Polling are subject to multiple other concerns. For one, as Luskin et al. (2014) note, there is a great deal of item sampling bias in typical knowledge batteries. Add to it the tendency of the rich to get richer over these naturalistic informative processes and the fact that ceiling effects cap observable growth in learning. A combination of these issues can contrive to create a scenario where observed gains in knowledge are *negatively* correlated with true knowledge gain (Luskin et al. 2014).

Besides affecting estimates of learning, guessing likely also affects cross-sectional measures of political knowledge. Guessing likely abrades correlations with criteria. Guessing, as we note above, causes only positive error when the items are conventionally scored as proportion correct. And this positive error is negatively correlated with knowledge—the more uninformed one is, the greater the positive error in their score on average. When error is negatively correlated with the true measure, it weakens correlation with variables the latent concept ought to be correlated with. Guessing related error in cross-sectional measures may be addressed by another measurement wave, followed by a latent class model, much like ours, except identified by the assumption that respondents cannot go from marking an item correctly to marking it incorrectly, and vice versa.

In all, our aim in this paper was to draw attention to an important problem that affects measurement of learning, and to propose a way to solve the problem. The problem we highlight affects areas other than politics—most prominently, education. In a non-trivial number of cases, educational researchers still use naive estimates of learning when measuring gains from implementing certain programs. And accounting for guessing using a latent class estimate may provide a much better account of the efficacy of the programs.

However, many opportunities for improving the model still remain. A natural extension of the latent class model developed in the paper would be a two-wave latent trait model that allows for constraints on cross-wave transitions. Such a model of learning, which accounts for guessing by exploiting cross-wave transitions, is liable to provide better estimates of learning than simple averages, and perhaps even IRT. Secondly, extensions to the model that exploit data from more than two waves are liable to allow us to better account for guessing related error by leveraging more data to estimate γ. (We present a model that exploits data from three waves in Appendix Appendix G.) And thirdly, ancillary data, such as certainty of responses (for instance, Pasek et al. 2015), and data from experiments in which correct answers are incentivized (see, for instance, Prior and Lupia 2008; Prior et al. 2015), may prove useful in building larger (and better) models for measuring learning.

REFERENCES

Alvarez, R. M. and C. H. Franklin (1994). Uncertainty and political perceptions. *The Journal of Politics 56*(03), 671–688.

Alvarez, R. M. and G. Glasgow (1996). Do voters learn from presidential election campaigns? Technical report.

Bartels, L. M. (1986). Issue voting under uncertainty: An empirical test. *American Journal of Political Science*, 709–728.

Bransford, J. D. and M. K. Johnson (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior 11*(6), 717–726.

Cooke, N. J., R. S. Atlas, D. M. Lane, and R. C. Berger (1993). Role of high-level knowledge in memory for chess positions. *The American journal of psychology*.

Cor, K. and G. Sood (2016). Replication data for: Learning without forgetting: A latent class model for measuring learning. http://dx.doi.org/10.7910/DVN/HZHVCU. Accessed: 2016-02-03.

Cor, M. K. (2012). *An Experimental Test of the Effect of Norm-Referenced and Criterion Referenced Feedback on Just World Beliefs, Motivation, and Performance: Does Social Disadvantage Matter?* Ph. D. thesis, Stanford University, Stanford, CA.

Delli Carpini, M. X. and S. Keeter (1996). *What Americans know about politics and why it matters*. Yale University Press.

Dooling, D. J. and R. Lachman (1971). Effects of comprehension on retention of prose. *Journal of experimental psychology 88*(2), 216.

Eckhardt, B. B., M. R. Wood, and R. S. Jacobvitz (1991). Verbal ability and prior knowledge contributions to adults' comprehension of television. *Communication Research 18*(5), 636–649.

Fishkin, J. (2009). *When the people speak: Deliberative democracy and public consultation*. Oxford University Press.

Fishkin, J. (2016). Center for deliberative polling, stanford university. http://cdd.stanford.edu. Accessed: 2016-02-04.

Fishkin, J., T. Kousser, R. C. Luskin, and A. Siu (2015, 12). Deliberative agenda setting: Piloting reform of direct democracy in california. *Perspectives on Politics 13*, 1030–1042.

Fishkin, J. S., R. C. Luskin, and A. Siu (2014). Europolis and the european public sphere: Empirical explorations of a counterfactual ideal. *European Union Politics*, 1465116514531507.

Franklin, C. H. (1991). Eschewing obfuscation? campaigns and the perception of us senate incumbents. *American Political Science Review 85*(04), 1193–1214.

Frazer, E. and K. Macdonald (2003). Sex differences in political knowledge in britain. *Political Studies 51*(1), 67–83.

Ghalanos, A. and S. Theussl (2010). Rsolnp: General non-linear optimization. *R package version*, 1–0.

Gibson, J. L. and G. A. Caldeira (2009). Knowing the supreme court? a reconsideration of public ignorance of the high court. *The Journal of Politics 71*(02), 429–441.

Hambrick, D. Z. (2003). Why are some people more knowledgeable than others? a longitudinal study of knowledge acquisition. *Memory & cognition 31*(6), 902–917.

Hansen, E. A., F. L. Schmidt, and J. C. Hansen (1975). A model for the correction for guessing on multiple-choice tests. *ACM SIGSOC Bulletin 7*(1-4), 24–28.

Hansen, K. M. (2004). *Deliberative democracy and opinion formation*. University Press of Southern Denmark Odense.

Iyengar, S. (1990). Shortcuts to political knowledge: The role of selective attention and accessibility. *Information and democratic processes*, 160–85.

Iyengar, S., R. C. Luskin, and A. Fishkin (2004). Deliberative public opinion in presidential primaries. In *Voice and citizenship conference*. Citeseer.

Lenz, G. S. (2009). Learning and opinion change, not priming: Reconsidering the priming hypothesis. *American Journal of Political Science 53*(4), 821–837.

Lord, F. M. (1975). Formula scoring and number-right scoring. *Journal of Educational Measurement 12*(1), 7–11.

Lord, F. M., M. R. Novick, and A. Birnbaum (1968). *Statistical theories of mental test scores.* Addison-Wesley.

Luskin, R. C. and J. G. Bullock (2004). Measuring political knowledge. In *Annual Meeting of the Midwest Political Science Association*.

Luskin, R. C. and J. G. Bullock (2011). 'don't know' means 'don't know': Dk responses and the public's level of political knowledge. *The Journal of Politics 73*(2), 547–557.

Luskin, R. C., D. B. Crow, J. S. Fishkin, W. Guild, and D. Thomas (2008). Report on the deliberative poll® on "vermont's energy future". *Center for Deliberative Opinion Research, University of Texas at Austin. A copy of this report is available at http://www.vermontsenergyfuture.info/Results% 20and% 20reports.html*.

Luskin, R. C. and J. S. Fishkin (1998). Deliberative polling, public opinion, and democracy: The case of the national issues convention. In *Slightly revised from a paper presented at the annual meeting of the American Political Science Association, Boston.*

Luskin, R. C., J. S. Fishkin, S. Boucher, and H. Monceau (2008). Considered opinions on further eu enlargement: Evidence from an eu-wide deliberative poll. *presentation at the International Society of Political Psychology, Paris, France*.

Luskin, R. C., J. S. Fishkin, and R. Jowell (2002). Considered opinions: Deliberative polling in britain. *British Journal of Political Science 32*(3), 455–487.

Luskin, R. C., J. S. Fishkin, R. Jowell, and A. Park (1999). Learning and voting in britain: insights from the deliberative poll. In *annual meeting of the American Political Science Association, Atlanta*.

Luskin, R. C., J. S. Fishkin, I. McAllister, J. Higley, and P. Ryan (2000). Information effects in referendum voting: evidence from the australian deliberative poll. In *annual meeting of the European Consortium for Political Research, Copenhagen*, Volume 4, pp. 2000.

Luskin, R. C., J. S. Fishkin, and D. L. Plane (1999). Deliberative polling and policy outcomes: Electric utility issues in texas. In *annual meeting of the Midwest Political Science Association, Chicago*, Volume 4, pp. 1999.

Luskin, R. C., K. Hahn, J. S. Fishkin, and S. Iyengar (2006). The deliberative voter.

Luskin, R. C., I. O'Flynn, J. S. Fishkin, and D. Russell (2014). Deliberating across deep divides. *Political Studies 62*(1), 116–135.

Luskin, R. C. and G. Sood (2012). Hidden knowledge (and veiled ignorance): Is the public distinctly less (or even more) ignorant than we have thought? In *Annual Meeting of the Midwest Political Science Association*.

Luskin, R. C., G. Sood, and A. Helfer (2014). Measuring learning in informative processes. In *European Consortium of Political Science Research Workshop, Salamanca, Spain*.

Mondak, J. J. (1999). Reconsidering the measurement of political knowledge. *Political Analysis 8*(1), 57–82.

Mondak, J. J. (2001). Developing valid knowledge scales. *American Journal of Political Science*, 224–238.

Mondak, J. J. and M. R. Anderson (2004). The knowledge gap: A reexamination of gender-based differences in political knowledge. *Journal of Politics 66*(2), 492–512.

Nunnally, J. C. (1967). *Psychometric Theory*. Tata McGraw-Hill Education.

O'Flynn, I. and G. Sood (2014). What would dahl say?: An appraisal of the democratic credentials of deliberative polls and other mini-publics. In *Deliberative Mini-Publics*, pp. 41–58. ECPR Press.

Park, A., R. Jowell, and S. McPherson (1998). The future of the national health service: results from a deliberative poll. *Paper available at http://cdd. stanford. edu/research/papers/2000/report_kings_fund. pdf*.

Pasek, J., G. Sood, and J. Krosnick (2015). Misinformed about the affordable care act? leveraging certainty to assess the prevalence of misinformation. *Journal of Communication 65*(4), 660–673.

Prior, M. and A. Lupia (2008). Money, time, and political knowledge: Distinguishing quick recall and political learning skills. *American Journal of Political Science 52*(1), 169–183.

Prior, M., G. Sood, and K. Khanna (2015). You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions. *Quarterly Journal of Political Science 10*(4), 489–518.

Recht, D. R. and L. Leslie (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology 80*(1), 16.

Sanchez, M. E. and G. Morchio (1992). Probing "dont know" answers: Effects on survey estimates and variable relationships. *Public Opinion Quarterly 56*(4), 454–474.

Siu, A., J. Fishkin, R. C. Luskin, N. Kim, and S. Westwood (2011). Hard times, hard choices: Michigan citizens deliberate. In *APSA 2011 Annual Meeting Paper*.

Sniderman, P. M., R. A. Brody, and P. E. Tetlock (1993). *Reasoning and choice: Explorations in political psychology*. Cambridge University Press.

Stolle, D. and E. Gidengil (2010). What do women really know? a gendered analysis of varieties of political knowledge. *Perspectives on Politics 8*(01), 93–109.

Sturgis, P., N. Allum, and P. Smith (2008). An experiment on the measurement of political knowledge in surveys. *Public Opinion Quarterly 72*(1), 90–102.

Sturgis, P., C. Roberts, and N. Allum (2005). A different take on the deliberative poll information, deliberation, and attitude constraint. *Public Opinion Quarterly 69*(1), 30–65.

Tallmadge, G. K. (1982). The correction for guessing a case in which its use made treatment effects appear larger than they really were. *Evaluation Review 6*(6), 837–841.

Varadhan, R. and G. Grothendieck (2011). alabama: Constrained nonlinear optimization. *R package version 1*.

Verba, S., N. Burns, and K. L. Schlozman (1997). Knowing and caring about politics: Gender and political engagement. *The Journal of Politics 59*(04), 1051–1072.

Weiksner, G. M. (2008). Measurement error as a threat to causal inference: Acquiescence bias and deliberative polling. In *Draft Paper prepared for the Political Methodology Conference in Ann Arbor*.

Appendix A A Latent Class Model For Forced Choice Items

Many closed-ended knowledge items do not include an explicit 'Don't know' option. Since rates of guessing are perforce equal or higher when a 'Don't know' option is not offered than when it is, the threat to inference when not accounting for guessing is yet greater on items without a 'Don't know' option. Here we develop a latent class model for such items.

Using the notation in the paper, responses to such closed-ended multiple choice items across two measurement waves can be represented by the following two-way table:

Table A1: Two-way Table of Manifest Pre-Post Response Patterns for a Single Closed-Ended Question

	Post	
Pre	0	1
0	x_{00}	x_{01}
1	x_{10}	x_{11}

Thus, the traditional estimate of the effect for any single item is: $(x_{01} + x_{11}) - (x_{10} + x_{11})$, or simply, $(x_{01} - x_{10})$. Next, we specify the latent class model for a single wave. We once again assume that there is no partial knowledge, and that either an individual knows the answer or doesn't— so knowledge and ignorance are the only two latent states. We again assume that the conditional probability of answering correctly given a respondent knows the answer is 1 or $P(u = 1|\text{Knowing}) = 1$ and $P(u = 0|\text{Knowing}) = 0$. Only the ignorant guess and when they guess they are lucky γ number of times or $P(u = 1|\text{guess}) = \gamma$. The following table summarizes the:

Table A2: Latent Class Model for A Closed-ended Question

	Latent Class	
Observed Response (u)	Guesser	Knower
u=0	$(1 - \gamma)$	0
u=1	γ	1

The probability of different response patterns across two waves can be described by multiplying the latent class models for each wave -

$$\otimes \begin{array}{c} \\ u=0 \\ u=1 \end{array} \begin{array}{c} \begin{array}{cc} G & K \end{array} \\ \begin{pmatrix} (1-\gamma) & 0 \\ \gamma & 1 \end{pmatrix} \end{array} \begin{array}{c} \\ u=0 \\ u=1 \end{array} \begin{array}{c} \begin{array}{cc} G & K \end{array} \\ \begin{pmatrix} (1-\gamma) & 0 \\ \gamma & 1 \end{pmatrix} \end{array} =$$

$$\begin{array}{c} \\ u=00 \\ u=01 \\ u=10 \\ u=11 \end{array} \begin{array}{c} \begin{array}{cccc} GG & GK & KG & KK \end{array} \\ \begin{pmatrix} (1-\gamma)^2 & 0 & 0 & 0 \\ (1-\gamma)\gamma & (1-\gamma) & 0 & 0 \\ \gamma(1-\gamma) & 0 & (1-\gamma) & 0 \\ \gamma^2 & \gamma & (1-\gamma) & 1 \end{pmatrix} \end{array}$$

Next, in order to define the system of equations that describe the underlying multinomial distribution of observed response patterns, we multiply the latent class transition matrix by a vector of latent class transition parameters. If we let, G stand for Guessing, and K stand for Knowledgeable, there are four possible latent class transitions that could hypothetically occur: GG, GK, KG, and KK.

Assuming the respondent who knows the right answer doesn't forget it over the course of the experiment, and that she doesn't become misinformed, the proportion of people moving from knowing to guessing (KG) must be zero.

Multiplying the 4×4 matrix from equation 1 by the latent class transition vector, λ, and setting λ_{KG} to zero, we get the following system of equations that defines the vector of expected proportions, π:

$$\begin{array}{c} \\ u=00 \\ u=01 \\ u=10 \\ u=11 \end{array} \begin{array}{c} \begin{array}{cccc} GG & GK & KG & KK \end{array} \\ \begin{pmatrix} (1-\gamma)^2 & 0 & 0 & 0 \\ (1-\gamma)\gamma & (1-\gamma) & 0 & 0 \\ \gamma(1-\gamma) & 0 & (1-\gamma) & 0 \\ \gamma^2 & \gamma & (1-\gamma) & 1 \end{pmatrix} \end{array} \begin{pmatrix} \lambda_{GG} \\ \lambda_{GK} \\ \lambda_{KG} = 0 \\ \lambda_{KK} \end{pmatrix}$$

$$= \begin{pmatrix} (1-\gamma)(1-\gamma)\lambda_{GG} \\ (1-\gamma)\gamma\lambda_{GG} + (1-\gamma)\lambda_{GK} \\ \gamma(1-\gamma)\lambda_{GG} \\ \gamma^2\lambda_{GG} + \gamma\lambda_{GK} + \lambda_{KK} \end{pmatrix} = \begin{pmatrix} \pi_{00} \\ \pi_{01} \\ \pi_{10} \\ \pi_{11} \end{pmatrix}$$

Appendix B Accounting for Hidden Knowledge in Don't Know Responses

Some research suggests that roughly 3% of the people who know a piece of information still mark 'don't know' on a closed-ended item (Luskin and Bullock 2011; Luskin and Sood 2012; Sanchez and Morchio 1992; Sturgis et al. 2008). In this section, we describe a version of the latent class model developed in the paper that accounts for this: the model takes the probability of a respondent who knows the answer offering a Don't Know response to be .03.

In the notation developed in the main body of the paper, responses to such closed-ended multiple choice items that offer a 'Don't Know' across two measurement waves can be represented by the following two-way table:

Table B1: Relationship Between Manifest Responses Across Two Waves

		Post		
		0	1	c
Pre	0	x_{00}	x_{01}	x_{0c}
	1	x_{10}	x_{11}	x_{1c}
	c	x_{c0}	x_{c1}	x_{cc}

Thus, the traditional estimate of the effect for any single item is: $(x_{01} + x_{c1} - x_{10} - x_{1c})$

Next, we specify the latent class model for a single wave. We once again assume that there is no 'partial knowledge', that either an individual knows the answer or doesn't. And those who don't know the answer can choose to guess or confess. We now assume that the conditional probability of answering correctly given a respondent knows the answer is .97 or $P(u = 1|K) = .97$, $P(u = 0|K) = 0$ and the $P(u = c|K) = .03$. Only the ignorant guess and when they guess they are lucky γ number of times or $P(u = 1|G) = \gamma$. The following table summarizes the model:

Table B2: Relationship Between Manifest Responses and Latent Classes when DK can Indicate Knowledge

		Latent Classes		
		G	K	C
	$u=0$	$(1-\gamma)$	0	0
Observed	$u=1$	γ	.97	0
	$u=c$	0	.03	1

The probability of different response patterns across two waves can be described by multiplying the latent class models for each wave:

$$\left(\begin{array}{cccc} & G & K & C \\ u=0 & 1-\gamma & 0 & 0 \\ u=1 & \gamma & .97 & 0 \\ u=c & 0 & .03 & 1 \end{array}\right) \otimes \left(\begin{array}{cccc} & G & K & C \\ u=0 & 1-\gamma & 0 & 0 \\ u=1 & \gamma & .97 & 0 \\ u=c & 0 & .03 & 1 \end{array}\right) = \tag{11}$$

$$\left(\begin{array}{cccccccccc} & GG & GK & GC & KG & KK & KC & CG & CK & CC \\ u=00 & (1-\gamma)^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ u=01 & (1-\gamma)\gamma & .97(1-\gamma) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ u=0c & 0 & 0 & .03(1-\gamma) & 0 & 0 & 0 & 0 & 0 & 0 \\ u=10 & (1-\gamma)\gamma & 0 & 0 & .97(1-\gamma) & 0 & 0 & 0 & 0 & 0 \\ u=11 & \gamma^2 & .97\gamma & 0 & .97\gamma & (.97)^2 & 0 & 0 & 0 & 0 \\ u=1c & 0 & .03\gamma & \gamma & 0 & .03(.97) & .97 & 0 & 0 & 0 \\ u=c0 & 0 & 0 & 0 & .03(1-\gamma) & 0 & 0 & (1-\gamma) & 0 & 0 \\ u=c1 & 0 & 0 & 0 & .03\gamma & .03(.97) & 0 & \gamma & .97 & 0 \\ u=cc & 0 & 0 & 0 & 0 & (.03)^2 & .03 & 0 & .03 & 1 \end{array}\right)$$

Next, the latent class transition matrix is multiplied by a vector of unknown latent class transition parameters to define the system of equations that describe the multinomial distribution of observed response patterns. There are nine latent class transitions that could hypothetically occur: GG, GK, GC, KG, KK, KC, CG, CK, and CC.

To identify the item level conditional probability of getting an item correct if the respondent chooses to guess, we assume that the respondent who knows the right answer doesn't forget it over the course of the experiment, and that she doesn't become misinformed. Under these assumptions, the proportion of people moving from knowing to guessing (KG) or confessing to ignorance (KC) must be zero.

Multiplying the 9×9 matrix from equation 11 by the latent class transition vector, λ, and setting λ_{KG} and λ_{KC} to zero, we get the following system of equations that defines the vector of expected response pattern proportions, π:

$$\begin{pmatrix} & GG & GK & GC & KG & KK & KC & CG & CK & CC \\ u=00 & (1-\gamma)^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ u=01 & (1-\gamma)\gamma & .97(1-\gamma) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ u=0c & 0 & 0 & .03(1-\gamma) & 0 & 0 & 0 & 0 & 0 & 0 \\ u=10 & (1-\gamma)\gamma & 0 & 0 & .97(1-\gamma) & 0 & 0 & 0 & 0 & 0 \\ u=11 & \gamma^2 & .97\gamma & 0 & .97\gamma & (.97)^2 & 0 & 0 & 0 & 0 \\ u=1c & 0 & .03\gamma & \gamma & 0 & .03(.97) & .97 & 0 & 0 & 0 \\ u=c0 & 0 & 0 & 0 & .03(1-\gamma) & 0 & 0 & (1-\gamma) & 0 & 0 \\ u=c1 & 0 & 0 & 0 & .03\gamma & .03(.97) & 0 & \gamma & .97 & 0 \\ u=cc & 0 & 0 & 0 & 0 & (.03)^2 & .03 & 0 & .03 & 1 \end{pmatrix} \begin{pmatrix} \lambda_{GG} \\ \lambda_{GK} \\ \lambda_{GC} \\ \lambda_{KG}=0 \\ \lambda_{KK} \\ \lambda_{KC}=0 \\ \lambda_{CG} \\ \lambda_{CK} \\ \lambda_{CC} \end{pmatrix} \tag{12}$$

$$= \begin{pmatrix} (1-\gamma)^2\lambda_{GG} \\ (1-\gamma)\gamma\lambda_{GG} + .97(1-\gamma)\lambda_{GK} \\ .03(1-\gamma)\lambda_{GG} + (1-\gamma)\lambda_{GC} \\ \gamma(1-\gamma)\lambda_{GG} \\ \gamma^2\lambda_{GG} + .97\gamma\lambda_{GK} + (.97)^2\lambda_{KK} \\ .03\gamma\lambda_{GK} + \gamma\lambda_{GC} + .03(.97)\lambda_{KK} \\ (1-\gamma)\lambda_{CG} \\ .03(.97)\lambda_{KK} + \gamma\lambda_{CG} + .97\lambda_{CK} \\ (.03)^2\lambda_{KK} + .03\lambda_{CK} + \lambda_{CC} \end{pmatrix} = \begin{pmatrix} \pi_{00} \\ \pi_{01} \\ \pi_{0c} \\ \pi_{10} \\ \pi_{11} \\ \pi_{1c} \\ \pi_{c0} \\ \pi_{c1} \\ \pi_{cc} \end{pmatrix}$$

Appendix C Details of Deliberative Polls

The polls were conducted by James Fishkin in collaboration with other scholars. More information about Deliberative Polling is available at http://cdd.stanford.edu. The references in the reference column in the table below are to other papers, research notes, and reports based on the data from the relevant polls. (Note: While we limit ourselves to citing a single paper, research note or report, often enough there are multiple papers and research notes that utilize data from the poll.)

Table C1: Details of Deliberative Polls

Poll Name	Year	Region	Country	Mode	Issue(s)	Reference
UK EU	1995		U.K.	Face-to-Face	Future of Britain in Europe	Sturgis et al. (2005)
UK Health	1998		U.K.	Face-to-Face	National Health Service	Park et al. (1998)
UK Monarchy	1996		U.K.	Face-to-Face	Future of Monarchy	Sturgis et al. (2005)
UK BGE	1997		U.K.	Face-to-Face	Electoral Choice	Luskin et al. (1999)
AU Monarchy	1999		Australia	Face-to-Face	Referendum on becoming a Republic	Luskin et al. (2000)
UK Crime	1994		U.K.	Face-to-Face	Crime	Luskin et al. (2002)
EU 2007	2007		EU-Wide	Face-to-Face	Pension systems, Globalization, and Enlargement	Luskin et al. (2008)
CPL	1996	Texas	U.S.	Face-to-Face	Energy Choices	Luskin et al. (1999)
WTU	1996	Texas	U.S.	Face-to-Face	Energy Choices	Luskin et al. (1999)
SWEPCO	1996	Texas	U.S.	Face-to-Face	Energy Choices	Luskin et al. (1999)
Bulgaria	2007		Bulgaria	Face-to-Face	Roma: Housing, Crime and Education	CDD[a]
NIC	1996		U.S.	Face-to-Face	Foreign and Economic Policy	Luskin and Fishkin (1998)
Denmark	2000		Denmark	Face-to-Face	Adoption of Euro	Hansen (2004)
N. Ireland	2007	Omagh	N. Ireland	Face-to-Face	Local Education	Luskin et al. (2014)
Vermont	2007	Vermont	U.S.	Face-to-Face	Energy Choices	Luskin et al. (2008)
Ca. Ref.	2011	California	U.S.	Face-to-Face	Government Reform	Fishkin et al. (2015)
San Mateo	2008	California	U.S.	Face-to-Face	Housing	Weiksner (2008)
EU 2009	2009		EU-Wide	Face-to-Face	Environment and EU-Reform	Fishkin et al. (2014)
BTP 2004 Primaries	2004		U.S.	Online	Online Primaries	Iyengar et al. (2004)
BTP 2004 GE	2004		U.S.	Online	General Elections	Luskin et al. (2006)
BTP 2005	2005		U.S.	Online	Healthcare and Education	PBS[b]
BTP 2007	2007		U.S.	Online	Electoral Reform	
Michigan	2009	Michigan	U.S.	Face-to-Face	Economy and Budget Choices	Siu et al. (2011)

[a] http://cdd.stanford.edu/polls/bulgaria/2007/bulgaria-results.pdf

[b] http://www.pbs.org/newshour/spc/btp/articles/events_deliberation.html

Appendix D Question Text and Response Options

Correct answers are in bold text.

AU Monarchy

- Issue Related Items: Which one of these best describes the current role of the Queen in relation to the Australian Governor General? - She chooses the Governor General, **She appoints the Governor General only on the advice of the Prime Minister**, She appoints the Governor General only on the recommendation of parliament, She has no role in choosing or appointing the Governor General
- Which one of these is true of the current role of the Governor-General in Australia - **He performs only ceremonial duties**, He can decide whether or not to dismiss the government, He acts only on the Queen's instructions, He controls the government
- If the referendum on the republic is passed, which one of these would the role of the President be more like? - **The current Governor General**, The American President, The British Prime Minister, the current Australian Prime Minister
- If the referendum on the republic is passed, which one of these best describes how the Australian Prime Minister in the Republic can remove the President? - At any time without reporting to parliament, **At any time but must later obtain approval from the House of Representatives**, Only after a fair trial in parliament, The Prime Minister could not remove president.
- Under the proposed change to a Republic, which of these, as far as you know, would definitely change.
  - The Australian Flag
  - The National Anthem

- **The word 'Royal' in the names of the Royal Australian Navy and the Royal Australian Air Force**
- Australia's participation in the Commonwealth Games

- On the whole would you say the Liberal Party, or the Labor Party, is more in favour of Australia becoming a Republic?

- Which one of these best describes Aden Ridgeway - Leader of the federal opposition, Leader of the Australian Democrats, **An aboriginal senator in parliament**, A justice of the high court, Or is he none of these

- Which one of these best describes Jennie George - A Labor member of federal parliament, Secretary of the Teacher's Federation, **President of the ACTU**, Leader of the Australian Worker's Union, Or is she none of these

- And now about social and welfare issues. On the whole would you say the Liberal Party, or the Labor Party, is more concerned about social and welfare issues, they are equally concerned, or do you not have much of an im-pression about that? - the Liberal Party, **the Labor Party**, equally close

- And now thinking about business. On the whole would you say the Liberal Party, or the Labor Party, is closer to business, they are equally close or do you not have much of an impression about that? - **the Liberal Party**, the Labor Party, equally close

UK EU

- EU has recently expanded to 15 members - **True** or False.
- Switzerland is to join EU - True or **False**.
- Britain's income tax rates are decided in Brussels - True or **False**.
- Elections to European Parliament are held every 5 years - **True** or False.

- Of the three British parties, Liberal Democrats least in favour of EU - True or **False**.

UK Health

- Standard charge for NHS prescription is 10 pounds - True or **False**.
- Proportion of old people in Britain getting larger - **True** or False.
- Even with inflation taken into account, spending on NHS doubled over last 20 years - **True** or False.
- Most people use private medical treatment instead of NHS - True or **False**.
- All British women get free breast cancer screening on NHS - True or **False**.
- Cosmetic surgery is never available on NHS - True or **False**.

UK Crime

- British courts are allowed to sentence a murderer to death - True or **False**.
- Britain has largest prison population in Western Europe - **True** or False.
- Britain has more people serving life sentences than rest of European Community put together - True or **False**.
- Possible to be tried by jury in local magistrate court - True or **False**. General Politics Items -
- Number of MPs about 100 - True or **False**.
- Longest time allowed between general elections is 4 years - True or **False**.
- Britain's electoral system is based on proportional representation - True or **False**.

UK Monarchy

- Princess Anne next in line to throne - True or **False**.
- PM, not the queen, is Britain's Head of State - True or **False**.

- Queen is head of Commonwealth - True or **False**.
- Queen's duty to decide election date - True or **False**.
- No PM takes office without being asked by the Queen - True or **False**.
- Queen is head of Church of England - **True** or False.
- Queen is head of Church of Scotland - True or **False**.
- Britain has unwritten constitution - **True** or False.

UK BGE

- Inflation has been less than 5
- Interest rates are decided by Bank of England - True or **False**.
- Unemployment in UK is higher than in Germany - True or **False**.
- (Conservative, Labor, Liberal Democrat) Party's View on Government role in making incomes equal: 7 point semantic scale going from "Government should do nothing to make incomes in Britain more equal" to "Government should try much harder to make incomes in Britain more equal."
- (Conservative, Labor, Liberal Democrat) Party's View on Spending: 7 point semantic scale ranging from "Government should spend much less on services like education and health in order to cut taxes a lot" to "Government should spend a lot more on services like education and health, even if it means putting up taxes a lot."
- (Conservative, Labor, Liberal Democrat) Party's View on minimum wage: 7 point semantic scale ranging from "Government should definitely not introduce a minimum wage because too many low paid workers would then lose their jobs." to "Government should definitely introduce a minimum wage so that no employer can pay their workers too little."
- (Conservative, Labor, Liberal Democrat) Party's View on European Union: 7 point semantic scale ranging from "Britain should do much more to keep its distance

from the European Union." to "Britain should do much more to unite fully with the European Union."

CPL

- What energy source do you think produces the greatest amount of CPL's electricity; coal, wind, **natural gas**, fuel oil, nuclear, or solar?
- Overall, what group of CPL customers consume the most kilowatt hours of electricity; residential, commercial, or **industrial**?
- Which customer group pays the highest electric rates; residential, **commercial**, or industrial?
- Which of the following resources used to produce electricity produces the most air emissions; nuclear, **coal**, or natural gas?
- Roughly, how much profit do Texas electric utilities make on fuel; **0%**, 10%, 20%, or 30%?
- Which agency sets electric rates in Texas? - **PUC**, Texas Utility Commission, Government, CPL, WT, SWEPCO, Other.
- How much of your electric bill goes for fuel; less than 20%, **21 to 30%**, 31 to 40%, 41 to 50%, or more than 50%?

SWEPCO

- What energy source do you think produces the greatest amount of SWEPCO's electricity; **coal**, wind, natural gas, fuel oil, nuclear, or solar?
- Overall, what group of SWEPCO customers consume the most kilowatt hours of electricity; residential, commercial, or **industrial**?
- Which customer group pays the highest electric rates; **residential**, commercial, or industrial?

- Which of the following resources used to produce electricity produces the most air emissions; nuclear, **coal**, or natural gas?
- Which agency sets electric rates in Texas? - **PUC**, Texas Utility Commission, Government, CPL, WT, SWEPCO, Other.

WTU

- What energy source do you think produces the greatest amount of WTU's electricity; coal, wind, **natural gas**, fuel oil, nuclear, or solar?
- Overall, what group of WTU customers consume the most kilowatt hours of electricity; **residential**, commercial, or industrial?
- Which customer group pays the highest electric rates; **residential**, commercial, or industrial?
- Which of the following resources used to produce electricity produces the most air emissions; nuclear, **coal**, or natural gas?
- Which agency sets electric rates in Texas? - **PUC**, Texas Utility Commission, Government, CPL, WT, SWEPCO, Other.

NIC

- Has the United States sent ground troops to Iraq in the past ten years? – **Yes**, No
- Has the United States sent ground troops to Somalia in the past ten years? – **Yes**, No
- Has the United States sent ground troops to Rwanda in the past ten years? – Yes, **No**
- Has the United States sent ground troops to Bosnia in the past ten years? – **Yes**, No
- On which of the following does the federal government spend more money? – The space program, Foreign Aid, **National defense**, Agriculture subsidies

- With which of the following countries does the United States conduct the most international trade? **Canada**, Japan, Germany, Mexico
- Where would you put the Republican party on the following scale: Extremely liberal, Liberal, Slightly liberal, Moderate middle of road, **Slightly conservative, Conservative, Extremely conservative**
- Where would you put the Democratic party on the following scale: **Extremely liberal, Liberal, Slightly liberal**, Moderate middle of road, Slightly conservative, Conservative, Extremely conservative

Bulgaria

- If a policeman wants to detain a citizen for questioning at any time of the day or night, the citizen does not have the right to refuse – **True** or False.
- The Court has the right to detain suspects as long as needed to prove if they are guilty or not – True or **False**.
- The prosecution has the right to detain suspects, until the crime is solved – True or **False**.
- The Chief Prosecutor is responsible and accountable to Parliament – **True** or False.
- The police do not have the right to use violence against a detainee even if proven that he/she has committed a crime – True or **False**.
- The police has the right to keep you in detention for three days – True or **False**.
- A confession is not enough to find the defendant guilty – True or **False**.

Northern Ireland

- What percentage of majority-Protestant or majority-Catholic schools in Northern Ireland have at least 10% of the other religion in their enrolment? – 40 - 50%, 20 - 30%, **5 - 10%**, less than 1%

- By approximately what percentage has the number of children entering Omagh schools increased or decreased over the past five years – Increased by 20%, Increased by 10%, Stayed about the same, **Decreased by 10%**, Decreased by 20%
- The new entitlement framework requires that . . . – **Every school provides all 14-year-olds with a choice of at least 24**, subjects, Every child has the right to attend any school his or her parents wish, Every child be provided a free school meal every school day, Every child receives free textbooks
- The new entitlement framework requires that . . . – Every child receives tuition in the language of his or her parents' choice, Every child receives free transportation to and from school, Every denominational group has the right to run its own schools, **One-third of all the subjects offered must be applied**
- Which of the following is true of what pupils in Northern Ireland do after they leave school? - **About one-quarter go directly into employment**, About one-quarter leave school to be unemployed, About three-quarters of grammar school pupils attend university, About three-quarters of secondary school students attend Further Education College
- Which of the following is true of current school funding? – **Schools receive more funding for older pupils**, Schools receive the same funding for all pupils, regardless of age, Schools receive more funding for younger pupils
- Which of the following is true of the employing authority in the schools? – The official employer for all teachers is the Education and Library Board, The official employer for all teachers is the Department of Education, The official employer for all teachers in Catholic schools is the CCMS (Council for Catholic Maintained Schools), **The official employer for all teachers in voluntary grammar schools is the school's Board of Governors**

Denmark

- As a member of the single currency Denmark will, in case of a large public budget deficit, be sanctioned economically by the other European member-states: **Yes**, No, It has not been decided.

- Denmark can have an independent currency policy even if it decides to join the single currency. (Currency policy is the interest rate, currency rate of number of money in circulation) – Yes, **No**, It has not been decided.

- The value of the euro in Danish kroner is – 3,80 kr., **7,50 kr.**, 8,50 kr., 12,00 kr.

- If we vote yes at the referendum on the 28th September, the new currency will be in circulation in Denmark on – January 1, 2001, **January 1 2004**, January 1, 2007, January 1, 2010

- If Denmark joins the single currency, the Danish National Bank will be closed down, continue to operate as now, **become part of the European Central Bank**, or be a museum of Danish currency through times

- Will the single currency have a national part on its coins and bills? – **Yes**, No, It has not been decided

- Is Denmark already involved in a monetary union where the member states help each other in situations of an unstable foreign exchange market? – **Yes**, No

- I will now read some of the movements and parties up names. Please answer whether you think that these parties and movements recommend a yes or no on the euro in the coming referendum.

  - People's Movement against union – Recommend a yes, **Recommend a no**
  - June Movement – Recommend a yes, **Recommend a no**
  - Social Democrats – Recommend a yes, **Recommend a no**
  - Social Liberal – **Recommend a yes**, Recommend a no
  - The Conservative Party – **Recommend a yes**, Recommend a no

– Centre Democrats – **Recommend a yes**, Recommend a no

– Socialist People's Party – Recommend a yes, **Recommend a** no

– Danish People's Party – Recommend a yes, **Recommend a no**

– Christian Party – Recommend a yes, **Recommend a no**

– Left Denmark's Liberal Party – **Recommend a yes**, Recommend a no

– Progress Party – Recommend a yes, **Recommend a no**

– Unity – Recommend a yes, **Recommend a no**

– Freedom 2000 – Recommend a yes, **Recommend a no**

EU 2007

- Which of the following countries is an official candidate to join the EU? Romania, Montenegro, **Croatia**, or Morocco.

- Are the members of the European Parliament ...? Answer options: **directly elected by the citizens of their country**, elected for three year terms, elected by the parliament of their country, or appointed by their national head of government

- Are new EU laws in the field of employment adopted by ...? the European Commission and in some cases with the European Parliament, **the Council of Ministers and in some cases with the European Parliament**, the European Parliament by itself or the European commission by itself

- Is the EU's role regarding unemployment benefits to ...? Finance the member states' unemployment benefit systems, decide the level and length of unemployment benefits in the member states, require that the member states merge their unemployment benefit systems by 2010 or **guarantee that all EU citizens have access to unemployment benefits where they live**

- Roughly a third of the EU's budget is devoted to one of the following. Is it ...? Answer options: **Helping the EU's less prosperous regions**, subsidizing the EU's fishing industry, financing diplomatic missions abroad or maintaining the EU's administration and bureaucracy

- By 2050, is the percentage of the adult EU population that is 65 or older projected to be ... Answer options: About one quarter of what it is now, about half of what it is now, about the same as what it is now, **about twice what it is now**, or about four times what it is now

- Which of the following is true of the amount of foreign aid given by the EU and its member states, combined, versus the amount given by the US? The EU and its member states give roughly four times as much, **the EU and its member states give roughly twice as much**, the EU and its member states and the US give about the same amount, the US gives roughly twice as much, or the US gives roughly four times as much

- On average, what percentage of the total spending by the governments of the EU member states for foreign aid? Answer options: **About 1%**, About 5%, About 9%, About 13%, or About 17%

- Which of the following countries does NOT possess nuclear weapons? Pakistan, India, North Korea or **Japan**

- In politics, people often talk of "left" and "right." On a scale from 0 to 10, where 0 means as far left as possible, 10 means as far right as possible, and 5 is exactly in the middle, where would you place the views of Nicholas Sarkozy **(6 to 10)**

- In politics, people often talk of "left" and "right." On a scale from 0 to 10, where 0 means as far left as possible, 10 means as far right as possible, and 5 is exactly in the middle, where would you place the views of Gordon Brown **(0 to 5)**

EU 2009

- Is the main decision-making body of the European Union the ...? European

Commission, **Council of Ministers**, European Parliament, European Court of Auditors, Couldn't say about that

- Only one of the following statements about the European Parliament is false. Which one is it? **It passes all EU laws**, It can dismiss the European Commission, It can reject the budget proposed by the Council of Ministers, It is involved in decisions about the admission of new Member States, Couldn't say about that

- Is the European Union represented on the international stage by the ...? **European Commission**, Council of Ministers, European Parliament, European Court of Auditors, Couldn't say about that

- Which of the following is true of Blue card workers? They can work anywhere in the EU, **They must have a university education**, They cannot bring family members to join them any faster than other immigrants, They are subject to the Returns Directive, Couldn't say about that

- Which of the following is true about the ways in which immigration policy is currently made? **The EU sets the basic rules about entry and residency requirements**, The EU decides how many immigrants can be admitted to each country, Work permits for immigrants must be approved by the EU, The EU plays no role in immigration policy, Couldn't say about that

- Which of the following is true of the EU's immigrants ...? **Most illegal immigrants enter the EU legally but outstay their visas**, Roughly 9% of the EU's total population was born outside the EU, Because of immigration, the EU's total population will increase by about 30 million by 2050, There are more illegal than legal immigrants in the EU, Couldn't say about that

- The percentage of the EU's total energy consumption that comes from fossil fuels (coal, gas or oil) is about ...? 50%, 60%, 70%, **80%**, Couldn't say about that, Couldn't say about that

- Which of the following produces the most greenhouse gases? **China**, The European Union, The United States, India, Couldn't say about that

- Which of the following is true about wind power in the European Union? Most of the EU's wind power is currently produced off-shore, **Wind power's share of EU energy consumption is increasing by about roughly 30% a year**, Wind power's share of the EU energy consumption is about three times that of solar power, Wind power's share of the EU energy consumption is currently about 5%, Couldn't say about that

Vermont

- Is the surcharge Vermonters pay on their electric bills for programs to reduce the need for electricity currently . . . ? Zero -there is no surcharge, **About half a cent per kilowatt hour**, About two cents per kilowatt hour, About five cents per kilowatt hour, About seven-and-a-half cents per kilowatt hour, Couldn't say

- What effect has Vermont's energy efficiency program had on the annual increase in the amount of electricity used by Vermonter's? Has it - Had almost no impact on the increase, Reduced it by 20%, **Reduced it by 50%**, Reduced it by 70%, Reduced it by 90%, Couldn't say

- Excluding Hydro Quebec, about what percentage of its electricity does Vermont currently get from renewable resources? 5%, **15%**, **25%**, 40%, Couldn't say

- And about what percentage of its electricity Does Vermont currently get from the Vermont Yankee nuclear plant? 5%, 10%, 20%, **33%**, Couldn't say

- About what percentage of its electricity does Vermont currently get from Hydro Quebec? 15%, **33%**, 45%, 60%, Couldn't say

- Roughly what percentage of Vermont's electricity is currently generated within Vermont? 12%, 33%, **55%**, 72%, Couldn't say

- How do Vermont's electricity rates compare with those of the rest of New England? Are they, on average, . . . Roughly 20% higher, Roughly 10% higher, About the same, **Roughly 10% lower**, Roughly 20% lower, Couldn't say

- What is the average electric bill for the typical Vermonter? Is it . . . Roughly $60, **Roughly $80**, Roughly $120, Roughly $180, Couldn't say

- Does Vermont's contract with the Vermont Yankee nuclear power plant expire in . . . 2010, **2012**, 2018, 2025, Couldn't say

San Mateo

- According to the Association of Bay Area Governments, between 1999 and 2006, what was the gap each year between the number of new homes needed and the number actually produced in San Mateo County? about 200, about 500, **about 1,000**, about 5,000, about 8,000, Or couldn't you say about that?

- As of September 2007, about how much was the median price of a single-family home in San Mateo County? $550,000, $680,000, $750,000, $830,000, **$940,000**, Or couldn't you say about that?

- As of 2006, about what percentage of County households could afford a median-priced house in San Mateo County? **12%**, 23%, 37%, 43%, 55%, Or couldn't you say about that?

- As of September 2007, what was the average rent for a 2-bedroom apartment in San Mateo County? $1,100, $1,300, **$1,700**, $2,200, $3,000, Or couldn't you say about that?

- According to the 2006 San Mateo County Housing Needs Study, by about how many housing units is the demand for housing in San Mateo County projected to grow between 2005 and 2025? 20,000, 40,000, 60,000, **70,000**, 90,000, Or couldn't you say about that?

- According to the same Housing Needs Study, among the new households that are expected to form in the County between 2005 and 2025, about what percentage, combined, would be households with low, very low and extremely low-incomes? less than one-third, less than one-half, **about half**, more than one-half, more than two-thirds, Or couldn't you say about that?

- According to the 2006 San Mateo County Housing Needs Study, roughly how many new homes (including houses, townhomes, condos, and apartments) will be built between 2005 and 2025 at the current rate of housing development? **25,000**, 50,000, 75,000, 100,000, Or couldn't you say about that?

- What percentage of San Mateo County's land is agricultural use, watershed, open space, wetlands, or parks? less than 25%, 33%, 58%, 67%, **more than 75%**, Or couldn't you say about that?

Michigan

- Which political party holds the majority in the Michigan State Senate? Or couldn't you say about that? Democrats, **Republicans**, Other, Or couldn't you say about that?

- How about the Michigan State House of Representatives? Which political party holds the majority there? Democrats, **Republicans**, Other, Or couldn't you say about that?

- Which of the following states has unemployment rates similar to Michigan's? **Oregon**, West Virginia, Ohio, Louisiana, Or, couldn't you say about that?

- People who have reached the 48-month limit can stay eligible for the Family Independence Program if. . . ? **they live in a county with a high unemployment rate**, they were fired from a job while on the Family Self-Sufficiency Plan, they are non-citizens on the Refugee Assistance Program, they have a small child who is active in the ChildFirst program, Or couldn't you say about that?

- About what percentage of African American children in Michigan live in poverty? Under 20%, About 30%, **About 40%**, Over 50%, Or couldn't you say about that?

California Referendum

- Which political party holds the majority in the California State Senate? **The Democratic Party**, Independents, The Republican Party, Couldn't say
- How about the California State Assembly? **The Democratic Party**, Independents, The Republican Party, Couldn't say
- How large a majority of the State Legislature is needed to approve a proposed constitutional amendment? A simple majority of both houses, a simple majority of the Assembly and two-thirds of Senate, **two-thirds of both houses**, three-fourths of both houses, Couldn't say
- How large a majority of the State Legislature is needed to increase taxes? A simple majority of both houses, a simple majority of the Assembly and two-thirds of Senate, **two-thirds of both houses**, three-fourths of both houses, Couldn't say
- Ballot measures can be signed by ...? anyone over 18 years of age, any citizen over 18 years of age, **only registered voters**, only registered voters who voted in last election, Couldn't say

BTP 2004 Primaries

- As far as you know, what did President Bush's tax cut do to the tax rate on income from investments such as dividends and capital gains? Did it ... **Lower it**, Keep it the same, Raise it, Don't know

- As far as you know, what was Wesley Clark's most recent position in government? —Senator from Arkansas, Secretary of the Air Force, **Supreme Allied Commander of NATO**, National Security Advisor to the President, Don't Know
- As far as you know, did President Bush end U.S. tariffs on imported steel because . . . The steel industry and steelworkers unions objected, **He wanted to avoid increased tariffs by the European Union on American exports**, Domestic steel production is insufficient to meet demand, thus raising the price of steel, He has never supported tariffs, Don't Know
- Which one of the Democratic presidential candidates served as Majority Leader in the House of Representatives? John Edwards, Howard Dean, Joe Lieberman, **Richard Gephardt**, Don't know
- Since President Bush took office, would you say that the rate of unemployment in the US has —decreased, stayed about the same, **increased**, don't know
- So far the number of Americans killed in action in Iraq can be numbered as: fewer than a hundred, **several hundred**, several thousand, more than twenty thousand, don't know
- As far as you know, what is the Bush administration's position on creating a new Free Trade Area of the Americas? Does the Bush administration . . . **Support it**, Neither support nor oppose it, Oppose it, Don't know

BTP 2004 GE

- How did John Kerry vote in the Senate on the resolution authorizing President Bush to go to war with Iraq? Did he . . . **Vote for it**, Vote against it, Not vote on it, Don't Know
- Which of the following countries now harbors the most Al Qaeda and Taliban fighters? India, **Pakistan**, Sri Lanka, Indonesia, Don't Know

- Iraq was directly involved in the attacks on the WTC and the Pentagon on 9-11 2001. True, **False**, Don't Know
- Large quantities of weapons of mass destruction have been found in Iraq. True, **False**, Don't Know
- On average, prescription drugs cost more in Canada than in the US. True, **False**, Don't Know
- Which of the following was true of George W. Bush during the Vietnam War? He was drafted but never went to Vietnam, He was a decorated officer serving in Vietnam, He was ineligible to serve in the military because of a medical deferment, **He served in the Texas Air National Guard**, Don't Know
- Which of the following was true of John Kerry during the Vietnam War? He was drafted but never went to Vietnam, **He was a decorated officer serving in Vietnam**, He was ineligible to serve in the military because of a medical deferment, He served in the Massachusetts Air National Guard, Don't Know
- A major destination of white collar jobs that have gone to other countries is? South Africa, Japan, Brazil, **India**, Don't Know
- Which of the following is closest to the number of Americans killed in Iraq since the war began? 100, 500, **1000**, 10,000, Don't Know

BTP 2005

- Among 29 wealthy industrialized nations, does the US rank top 3, top 10, **bottom 10**, or bottom 3 in students' math skills?
- In the last few years, it has become **harder**, easier or about the same to obtain US work permits for foreigners.

- In the last few years, the gap between minority and white students in math and reading tests at the elementary school level has **increased**, decreased, or stayed about the same.

- Number of uninsured Americans. 1 million, 25 million, **45 million**, 65 million or 85 million.

- Over the next fifty years, do you think the number of Americans over age 65 will . . . ? tripled, **doubled**, increased by half, or stayed about the same.

- The percentage of uninsured low income families. a quarter, fifty percent, **two-thirds**, or ninety percent.

BTP 2007

- About how many Americans are prevented from voting because of criminal convictions, or couldn't you say about that? 50,000, 500,000, **5,000,000**, 15,000,000, Couldn't say

- Does "gerrymandering" reshape districts to...? Increase electoral competition, **Ensure a majority for one party over another**, Include independents in the primary process, Make them much smaller, Couldn't say

- Approximately what percentage of eligible American voters actually vote in a given presidential general election, or couldn't you say about that? 10%, 30%, **50%**, 70%, Couldn't say

- Which of the following countries has a system of "compulsory voting," or couldn't you say about that? Ireland, United Kingdom, **Australia**, South Korea, Couldn't say

- At present, who must register with the selective service system in the US, or couldn't you say about that? Nobody, **Men between 18 and 25 years old**, Both men and women between 18 and 25 years old, Men and women who receive federal grants for education, Couldn't say

- A candidate is elected President of the United States if he or she . . . Gets more votes than any other candidate, Gets a majority of all the votes cast, Gets more Electoral College votes than any other candidate, **Gets a majority of Electoral College votes**, Couldn't say

- How often are congressional districts redrawn, or couldn't you say about that? Every 5 years, **Every 10 years**, Every 15 years, Every 20 years, Couldn't say

- Which two states have traditionally had the earliest presidential primary events for determining their delegates to the parties' national conventions, or couldn't you say about that? Nevada and South Carolina, **Iowa and New Hampshire**, Indiana and New Mexico, Wyoming and Delaware, Michigan and Florida, Couldn't say

Appendix E Unaccounted Gains in Related Knowledge?

As we note above, the latent class model implicitly accounts for gains in related knowledge as most such gains are already reflected in increases in proportion correct. However, it is possible that some gains in related knowledge are not reflected in increases in proportion correct. For instance, consider the following tendentious case: an individual learns that a particular response option isn't correct, and also by some accident chooses only between the other remaining incorrect options. Such growth in related knowledge paired with that particular pattern of responding would not increase probability of marking the right answer. To address worries that our model fails to account for gains in related knowledge, we present estimates from closed-ended responses that offer just two response options. On the mild assumption that people choose the option that they are most confident about (for empirical support, see Luskin and Sood 2012), all respondents who are more than 50% sure about the correct answer will choose the correct answer. If we assume confidence in correct answer to be uniformly distributed across the entire number line, and if the process only creates an intercept shift in certainty, increase in proportion correct ought to completely reflect all growth in knowledge.

Thus we compare differences in learning between the 51 items that offer only two options, and 124 items that offer more than two options using various estimators. For identification, we assume that items with two options and items with more than two options are exchangeable. The mean difference in learning between items with more than two options and items with just two options was .060 based on the conventional estimator. If there are greater unaccounted for gains in items with two-options (items in which fewer, if not no, gains in related knowledge go unmeasured) than more than two-options, one ought to see this difference shrink when we account for guessing. But adjusting for guessing using the LCA, rather than decrease this difference, slightly increases the difference to .067 ($p < .01$). In all, worries about unaccounted for gains in related knowledge are not supported by the data.

Appendix F Estimates of Learning over Deliberative Polls using Different Estimators

Table F1: Estimates of Learning over Deliberative Polls

Poll	n	n *(items)*	Raw	LCA	Stnd.
AU Monarchy	347	10	.284	.439	.345
			(.014)	(.018)	(.023)
BTP 2004 Primaries	328	7	.027	.06	.034
			(.016)	(.017)	(.014)
BTP 2004 GE	250	9	.071	.089	.084
			(.01)	(.013)	(.021)
BTP 2005	454	6	.035	.054	.039
			(.011)	(.015)	(.011)
BTP 2007	301	8	.112	.126	.122
			(.01)	(.012)	(.022)
Bulgaria	278	7	.093	.082	.075
			(.017)	(.026)	(.017)
Ca. Ref.	401	5	.211	.286	.231
			(.015)	(.016)	(.024)
CPL	216	7	.161	.19	.155
			(.016)	(.02)	(.016)
Denmark	363	9	.211	.359	.293
			(.011)	(.014)	(.023)
EU 2007	335	11	.182	.229	.206
			(.01)	(.013)	(.015)
EU 2009	348	6	.166	.182	.175
			(.014)	(.018)	(.016)
N. Ireland	124	7	.286	.323	.337
			(.022)	(.028)	(.023)
Michigan	310	9	.098	.088	.082
			(.013)	(.022)	(.013)
NIC	466	8	.091	.193	.116
			(.009)	(.013)	(.015)
San Mateo	239	8	.113	.121	.096
			(.012)	(.014)	(.015)
SWEPCO	232	5	.203	.233	.222
			(.02)	(.026)	(.032)
UK BGE	275	15	.103	.191	.157
			(.012)	(.018)	(.018)
UK Crime	299	7	.095	.142	.102
			(.013)	(.02)	(.016)
UK EU	224	5	.203	.275	.244
			(.019)	(.025)	(.032)
UK Health	230	6	.079	.122	.134
			(.014)	(.02)	(.025)
UK Monarchy	258	8	.152	.212	.209
			(.01)	(.015)	(.022)
Vermont	146	9	.397	.441	.4
			(.018)	(.021)	(.027)
WTU	230	5	.288	.347	.329
			(.017)	(.021)	(.023)
Inverse Variance Weighted			.14	.17	.168
			(.003)	(.004)	(.004)

Appendix G Three-wave Latent Class Model of the Relationship between Observed Response Patterns and Underlying Latent Transition Classes

The system of equations defining the multinomial distribution for response patterns across three waves of testing is gotten by taking the Kronecker product of a wave 1 classification matrix (see Table 2) and the Kronecker product of the classification matrices for two additional waves of testing. This result is then multiplied by a vector of latent class transition parameters, (λ). The sequence of steps and resulting system of 27 equations are shown below.

$$
\left(\begin{array}{lccc}
 & G & K & C \\
u=0 & 1-\gamma & 0 & 0 \\
u=1 & \gamma & 1 & 0 \\
u=c & 0 & 0 & 1
\end{array}\right) \otimes
$$

$$
\left(\begin{array}{lccccccccc}
 & GG & GK & GC & KG & KK & KC & CG & CK & CC \\
u=00 & (1-\gamma)^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
u=01 & (1-\gamma)\gamma & (1-\gamma) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
u=0c & 0 & 0 & (1-\gamma) & 0 & 0 & 0 & 0 & 0 & 0 \\
u=10 & (1-\gamma)\gamma & 0 & 0 & (1-\gamma) & 0 & 0 & 0 & 0 & 0 \\
u=11 & \gamma^2 & \gamma & 0 & \gamma & 1 & 0 & 0 & 0 & 0 \\
u=1c & 0 & 0 & \gamma & 0 & 0 & 1 & 0 & 0 & 0 \\
u=c0 & 0 & 0 & 0 & 0 & 0 & 0 & (1-\gamma) & 0 & 0 \\
u=c1 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma & 1 & 0 \\
u=cc & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right)
$$

This Kronecker product results in a 27×27 latent class transition matrix (not shown here), similar to one produced from equation 1. The matrix describes the conditional probabilities of the possible three-wave response patterns given each possible latent class transition. For example, given the latent transition from guess to guess to guess (GGG), equations defining the probability of the 27 possible response patterns (000, 001, 00c,..., cc0, cc1, ccc) are produced.

To get the final system of equations, the 27×27 latent class transition matrix is multiplied by the vector of possible latent class transition parameters, λ. Since we assume that participants cannot transition from knowing to guessing or confessing, we assign those transition parameters as 0. The resulting vector of transition parameters is as follows:

$$\begin{pmatrix} \lambda_{GGG} \\ \lambda_{GGK} \\ \lambda_{GGC} \\ \lambda_{GKG} = 0 \\ \lambda_{GKK} \\ \lambda_{GKC} = 0 \\ \lambda_{GCG} \\ \lambda_{GCK} \\ \lambda_{GCC} \\ \lambda_{KGG} = 0 \\ \lambda_{KGK} = 0 \\ \lambda_{KGC} = 0 \\ \lambda_{KKG} = 0 \\ \lambda_{KKK} \\ \lambda_{KKC} = 0 \\ \lambda_{KCG} = 0 \\ \lambda_{KCK} = 0 \\ \lambda_{KCC} = 0 \\ \lambda_{CGG} \\ \lambda_{CGK} \\ \lambda_{CGC} \\ \lambda_{CKG} = 0 \\ \lambda_{CKK} \\ \lambda_{CKC} = 0 \\ \lambda_{CCG} \\ \lambda_{CCK} \\ \lambda_{CCC} \end{pmatrix}$$

The final result is the system of 27 equations defining π which identifies the underlying multinomial distribution of response patterns shown below.

$$
\begin{pmatrix}
(1-\gamma)^3\lambda_{GGG} \\
(1-\gamma)^2\gamma\lambda_{GGG} + (1-\gamma)^2\lambda_{GGK} \\
(1-\gamma)^2\lambda_{GGC} \\
(1-\gamma)^2\gamma\lambda_{GGG} \\
(1-\gamma)\gamma^2\lambda_{GGG} + (1-\gamma)\gamma\lambda_{GGK} + (1-\gamma)\lambda_{KKK} \\
(1-\gamma)\gamma\lambda_{GGC} \\
(1-\gamma)^2\lambda_{GCG} \\
(1-\gamma)\gamma\lambda_{GCG} + (1-\gamma)\gamma\lambda_{GCK} \\
(1-\gamma)\lambda_{GCC} \\
(1-\gamma)^2\gamma\lambda_{GGG} \\
(1-\gamma)\gamma^2\lambda_{GGG} + (1-\gamma)\gamma\lambda_{GGK} \\
(1-\gamma)\gamma\lambda_{GGC} \\
(1-\gamma)\gamma^2\lambda_{GGG} \\
\gamma^3\lambda_{GGG} + \gamma^2\lambda_{GGK} + \lambda_{KKK} \\
\gamma^2\lambda_{GGC} \\
(1-\gamma)\gamma\lambda_{GCG} \\
\gamma^2\lambda_{GCG} + \gamma\lambda_{GCK} \\
\gamma\lambda_{GCC} \\
(1-\gamma)^2\gamma\lambda_{CGG} \\
(1-\gamma)\gamma\lambda_{CGG} + (1-\gamma)\lambda_{CGK} \\
(1-\gamma)\lambda_{CGC} \\
(1-\gamma)\gamma\lambda_{CGG} \\
\gamma^2\lambda_{CGG} + \gamma\lambda_{CGK} + \lambda_{CKK} \\
\gamma\lambda_{CGC} \\
(1-\gamma)\lambda_{CCG} \\
\gamma\lambda_{CCG} + \lambda_{CCK} \\
\lambda_{CCC}
\end{pmatrix}
=
\begin{pmatrix}
\pi_{000} \\ \pi_{001} \\ \pi_{00c} \\ \pi_{010} \\ \pi_{011} \\ \pi_{01c} \\ \pi_{0c0} \\ \pi_{0c1} \\ \pi_{0cc} \\ \pi_{100} \\ \pi_{101} \\ \pi_{10c} \\ \pi_{110} \\ \pi_{111} \\ \pi_{11c} \\ \pi_{1c0} \\ \pi_{1c1} \\ \pi_{1cc} \\ \pi_{c00} \\ \pi_{c01} \\ \pi_{c0c} \\ \pi_{c10} \\ \pi_{c11} \\ \pi_{c1c} \\ \pi_{cc0} \\ \pi_{cc1} \\ \pi_{ccc}
\end{pmatrix}
$$

The model that defines the three-wave multinomial distribution of response patterns affords some additional opportunities beyond the ones available in the two-wave model. In particular, the model allows for more accurate estimates of γ given that the number of lucky guesses at wave one and two are identified further with the addition of the third wave. The only ambiguous response patterns that remain are '111' response patterns that represent three lucky guesses which are inherently highly unlikely. The three-wave model can also be used to calculate guessing adjusted estimates of learning between waves 1 and 2, and waves 2 and 3.