

Updating Beliefs when Evidence is Open to Interpretation: Implications for Bias and Polarization

Online Appendix (Not for Publication)

Roland G. Fryer, Jr., Philipp Harms, and Matthew O. Jackson

October 2017

Appendix A: Proofs of Propositions

For all of the results below, we consider the case in which $q = 1$ as the other case is exactly the same simply ignoring dates in which no signal is observed, as the agent takes no action and does not update on those dates, and there are infinitely many dates in which signals are observed, almost surely.

Proof of Proposition 1.

Let us first show for any λ_0 and $\varepsilon > 0$, there exist $\bar{\delta}$ such that if $\delta \leq \bar{\delta}$, then $U(\sigma^1, \delta, \lambda_0) \geq U(\sigma, \delta, \lambda_0) - \varepsilon$ for all strategies σ . Recall that

$$U(\sigma, \delta, \lambda_0) = E \left(\sum_{t=1}^{\infty} \delta^t u_t(\sigma(h_{t-1}, \lambda_0)) \middle| \lambda_0 \right).$$

Write

$$U(\sigma, \delta, \lambda_0) = E(u_1(\sigma(\emptyset, \lambda_0)) | \lambda_0) + E \left(\sum_{t=2}^{\infty} \delta^t u_t(\sigma(h_{t-1}, \lambda_0)) \middle| \lambda_0 \right).$$

The basic idea is that as $\delta \rightarrow 0$, the future does not matter and the decision maker only needs to maximize the current period's payoff which amounts to choosing the most likely interpretation. Note that $E(\sum_{t=2}^{\infty} \delta^t u_t(\sigma(h_{t-1}, \lambda_0)) | \lambda_0)$ lies in the interval $[-\frac{\delta}{1-\delta}, \frac{\delta}{1-\delta}]$ lies within $[-\varepsilon, \varepsilon]$ if $\delta \leq \bar{\delta} = \frac{\varepsilon}{1+\varepsilon}$. Thus, if $\delta \leq \bar{\delta}$, then

$$U(\sigma^1, \delta, \lambda_0) - U(\sigma, \delta, \lambda_0) \geq E(u_1(\sigma^1(\emptyset, \lambda_0)) | \lambda_0) - E(u_1(\sigma(\emptyset, \lambda_0)) | \lambda_0) - \varepsilon. \quad (1)$$

Next, note that

$$E(u_1(\sigma(\emptyset, \lambda_0))|\lambda_0) = E[p \Pr[i_1 = \omega] + (1 - p) \Pr[i_1 \neq \omega]|\lambda_0].$$

Since $p > 1/2$, the maximizing solution is to set i_1 to match the most likely state ω given λ_0 , and so σ^1 is optimal for the first period optimization. This implies that

$$E(u_1(\sigma^1(\emptyset, \lambda_0))|\lambda_0) \geq E(u_1(\sigma(\emptyset, \lambda_0))|\lambda_0).$$

Thus, from (3) it follows that if $\delta \leq \bar{\delta}$, than

$$U(\sigma^1, \delta, \lambda_0) \geq U(\sigma, \delta, \lambda_0) - \varepsilon. \quad (2)$$

Next, let us now show that it is possible to choose $\bar{\delta}$ such that it approaches 1 as λ_0 approaches 0 or 1. For any δ , there exists $T(\delta)$ such that the expected sum of discounted utilities past time $T(\delta)$ amounts to less than $\varepsilon/2$ and so the utility is captured in the first $T(\delta)$ periods:¹

$$U(\sigma, \delta, \lambda_0) \geq E\left(\sum_{t=1}^{T(\delta)} \delta^t u_t(\sigma(h_{t-1}, \lambda_0)) \middle| \lambda_0\right) - \varepsilon/2.$$

Next, note that the expression

$$E\left(\sum_{t=1}^{T(\delta)} \delta^t u_t(\sigma^1(h_{t-1}, \lambda_0))\right)$$

is continuous in λ_0 including the extreme points of $\lambda_0 \in \{0, 1\}$ for any given δ . Note also that σ^1 is the approximately optimal strategy if $\lambda_0 = 1$, since then the expected payoff in any given period (independent of the history) is simply the probability that the interpretation is A times p plus the probability that the interpretation is the interpretation is B times $1 - p$. This is maximized by setting the interpretation to A . Similarly if $\lambda_0 = 0$ the optimal strategy is to interpret things as B in any given period. Thus, maximum likelihood storage rule, σ^1 , is optimal for $\lambda_0 \in \{0, 1\}$. Given the continuity, it is within $\varepsilon/2$ approximately optimal for any λ_0 close enough to 1 or 0. So, for any δ we can find λ_0 close enough to 1 or 0 for which

$$U(\sigma^1, \delta, \lambda_0) \geq U(\sigma, \delta, \lambda_0) - \varepsilon. \quad (3)$$

which completes the proof. ■

¹An upper bound is to set $\frac{\delta^T}{1-\delta} = \varepsilon/2$.

Proof of Proposition 2. We first state an auxiliary result, from Hoeffding (1963), that is useful in proving Proposition 2.

LEMMA 1 (Hoeffding's inequality) *If X_1, \dots, X_t are independent and $a_i \leq X_i \leq b_i$ for $i = 1, 2, \dots, t$, then for $\delta > 0$,*

$$P\left(\sum_{i=1}^t (X_i - E(X_i)) \geq t\epsilon\right) \leq e^{-2t^2\epsilon^2/\sum_{i=1}^t (b_i - a_i)}.$$

Let $n(\lambda)$ be the number of b interpreted signals minus the number of a interpreted signals needed to reach the frontier where $\lambda_t = 1/2$ starting from $\lambda_0 = \lambda$, i.e.,

$$n(\lambda) = \left\lfloor \frac{\log\left(\frac{\lambda}{1-\lambda}\right)}{\log\left(\frac{p}{1-p}\right)} \right\rfloor.$$

The $\lfloor \cdot \rfloor$ reflects starting from a prior below $1/2$, and otherwise it would be rounded up.

The process $n_t = n(\lambda_t)$ is a random walk in the integers such that n_t is increased by 1 every time there is an interpreted a signal and decreased by 1 every time there is an interpreted b signal. The conditional laws given the states A and B are denoted by P_A and P_B , respectively, and E_A and E_B are the corresponding expectations.

- (a) First, note that if $(1 - \pi)p + \pi(1 - \gamma) > 1/2$ (which is rewritten as $\gamma < \frac{1/2 - (1-p)(1-\pi)}{\pi}$), then even if all of the unclear signals are incorrectly interpreted, the majority of signals will still match the true state. Therefore, if the true state is A, then the increments $\Delta n_t = n_{t+1} - n_t$ are positive in expectation, i.e., $E_A(\Delta n_t) > 0$. Moreover, they have bounded first and second moments. It follows from the strong law of large numbers that $(n_t - E_A(n_t))/t$ converges to zero P_A -a.s., which implies that $n_t \rightarrow \infty$ P_A -a.s. and $\lambda_t \rightarrow 1$ P_A -a.s. The P_B -a.s. convergence of λ_t to zero is proven in a similar same way.

Note that this is automatically satisfied if $\pi < (p - 1/2)/p$ for any γ , which establishes the first sentence of the proposition.

- (b) Now suppose that $(1 - \pi)p + \pi(1 - \gamma) < 1/2$ and assume that the true state is B. We claim that P_B assigns positive probability to the event $\lambda_t \rightarrow 1$, which coincides with the event $n_t \rightarrow \infty$. First, we note that n_t reaches any preset level with positive probability if t is large enough. Therefore, it is sufficient to prove the proposition for n_0 large. Whenever n_t is positive, it is more likely to increase than to decrease, i.e., $P_B(\Delta n_t = 1) \equiv z > 1/2$. As long as this is the case, $E_B(\Delta n_t) = 2z - 1 > 0$ and Hoeffding's inequality states that for any $\epsilon > 0$,

$$P_B(n_t - n_0 \leq (2z - 1 - \epsilon)t) \leq e^{-t\epsilon^2/2}.$$

Setting $\epsilon = (2z - 1)/2$ leads to the bound

$$P_B(n_t \leq (z - 1/2)t) \leq P_B(n_t - n_0 \leq (z - 1/2)t) \leq e^{-t(z-1/2)^2/2}.$$

When n_0 is large, n_t cannot immediately fall below $(z - 1/2)t$. More specifically, this is impossible for $t \leq \lfloor n_0/(z + 1/2) \rfloor$. It follows that

$$P_B(\forall t : n_t > (z - 1/2)t) = 1 - P_B(\exists t : n_t \leq (z - 1/2)t) \geq 1 - \sum_{t > \lfloor n_0/(z+1/2) \rfloor} e^{-t(z-1/2)^2/2}.$$

The last expression is positive if n_0 is large enough. This proves that

$$P_B(\lim_{t \rightarrow \infty} \lambda_t = 1) = P_B(\lim_{t \rightarrow \infty} n_t = \infty) > 0.$$

It can be shown in a similar way that P_A assigns positive probability to the event $\lambda_t \rightarrow 0$.

■

Proof of Proposition 4. In the limit, the belief process places a.s. weight 1 on the true state of Nature when the $\gamma = 1/2$ rule is used, as shown in proposition 2 (and could also be deduced from Levy's 0-1 law and Martingale convergence of beliefs). Therefore, the belief process remains eventually on the correct side of the frontier. Formally, under the $\gamma = 1/2$ rule, the random time

$$S = \inf\{t \geq 0 : \forall s \geq t : 1_{\lambda_t > 1/2} = \omega\}$$

is P_ω -a.s. finite for any $\omega \in \{A, B\}$. Therefore, for any T ,

$$\begin{aligned} U(\sigma^T, \delta, \lambda) &\geq E \left(1_{S \leq T} \sum_{t=T}^{\infty} \delta^t u_t(\sigma^T(h_{t-1}, \lambda_{t-1})) \right) = E \left(1_{S \leq T} \sum_{t=T}^{\infty} \delta^t u_t(\sigma^{FI}(\omega)) \right) \\ &> E \left(\sum_{t=T}^{\infty} \delta^t u_t(\sigma^{FI}(\omega)) \right) - \epsilon/2 > E \left(\sum_{t=0}^{\infty} \delta^t u_t(\sigma^{FI}(\omega)) \right) - \epsilon = U(\sigma^{FI}, \delta) - \epsilon. \end{aligned}$$

In the above equation, the first relation holds because some non-negative terms are dropped, the second relation holds because the agent calls out the right state past time S , the third relation holds for large enough T because $P(S \leq T) \rightarrow 1$ as $T \rightarrow \infty$, and the fourth relation holds for δ close enough to 1 because then the first T stages do not matter relative to the rest. ■

Proof of Proposition 5. Solving the expressions in section 4 iteratively, it follows that

$$\hat{\mu}_t = \mu_0 (\times_{\tau=1}^t W_\tau) + \sum_{\tau=1}^t s_\tau \frac{x_\tau^2}{(1+x_\tau)^2} (\times_{\tau'=\tau+1}^t W_{\tau'}) \quad (4)$$

where

$$W_\tau = \left[\frac{(1+x_\tau)^2 - x_\tau^2}{(1+x_\tau)^2} \right].$$

Note that since $x_t = \frac{\sigma_{t-1}^2}{\sigma_s^2}$ and $\sigma_{t-1}^2 = \frac{\sigma_0^2 \sigma_s^2}{\sigma_s^2 + (t-1)\sigma_0^2}$, it follows that $x_t = \frac{\sigma_0^2}{\sigma_s^2 + (t-1)\sigma_0^2}$, and so

$$W_\tau = \frac{(\sigma_s^2 + \tau\sigma_0^2)^2 - (\sigma_0^2)^2}{(\sigma_s^2 + \tau\sigma_0^2)^2} = \frac{(\sigma_s^2 + (\tau-1)\sigma_0^2)(\sigma_s^2 + (\tau+1)\sigma_0^2)}{(\sigma_s^2 + \tau\sigma_0^2)^2}.$$

Thus,

$$(\times_{\tau=1}^t W_\tau) = \frac{(\sigma_s^2 + 0\sigma_0^2)(\sigma_s^2 + (t+1)\sigma_0^2)}{(\sigma_s^2 + 1\sigma_0^2)(\sigma_s^2 + t\sigma_0^2)}$$

or

$$(\times_{\tau=1}^t W_\tau) = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_0^2} \left[1 + \frac{\sigma_0^2}{\sigma_s^2 + t\sigma_0^2} \right].$$

Similarly,

$$(\times_{\tau'=\tau+1}^t W_{\tau'}) = \frac{\sigma_s^2 + \tau\sigma_0^2}{\sigma_s^2 + (\tau+1)\sigma_0^2} \left[1 + \frac{\sigma_0^2}{\sigma_s^2 + t\sigma_0^2} \right].$$

Substituting for these expressions and x_τ into (4), we obtain:

$$\hat{\mu}_t = \mu_0 \frac{\sigma_s^2}{\sigma_s^2 + \sigma_0^2} \left[1 + \frac{\sigma_0^2}{\sigma_s^2 + t\sigma_0^2} \right] + \sum_{\tau=1}^t s_\tau \left(\frac{(\sigma_0^2 \sigma_s^2)^2}{(\sigma_s^2 + (\tau-1)\sigma_0^2 + \sigma_0^2 \sigma_s^2)^2} \right) \left(\frac{\sigma_s^2 + \tau\sigma_0^2}{\sigma_s^2 + (\tau+1)\sigma_0^2} \right) \left[1 + \frac{\sigma_0^2}{\sigma_s^2 + t\sigma_0^2} \right]. \quad (5)$$

This gives the expression claimed in the proposition. The expression for the Bayesian updater is standard:

$$\mu_t = \mu_0 \frac{\sigma_s^2}{\sigma_s^2 + t\sigma_0^2} + \sum_{\tau=1}^t s_\tau \frac{\sigma_0^2}{\sigma_s^2 + t\sigma_0^2},$$

and completes the proof. ■

Proof of Proposition 3. The argument in the proof of proposition 2b shows that when $\pi < 1$, then the values 0 and 1 occur with positive probability as the limit of the belief process λ_t as $t \rightarrow \infty$. It remains to show that $\lambda_t \rightarrow 1$ with probability tending to one as $\pi \rightarrow 1$ if $\lambda_0 > 1/2$. So let us assume $\lambda_0 > 1/2$. Obviously, in the extreme case that $\pi = 1$, all signals are interpreted as a and therefore, the probability that $\lambda_t \rightarrow 1$ equals 1. This probability depends continuously on the parameter π by Lemma 2. ■

LEMMA 2 *Under any strategy, the belief process Λ is continuous in the total variation norm with respect to the parameters p and π .*

Proof. It is equivalent to show the proposition for the point process n_t defined in the proof of Proposition 2 instead of the process λ_t . Let $p^k \rightarrow p$, $\pi^k \rightarrow \pi$, and let P^k and P be the corresponding laws of the process n_t . Furthermore, let \mathcal{F}_t be the sigma algebra generated by n_0, \dots, n_t and P_t^k the restriction of P^k to \mathcal{F}_t . It follows directly from the Chapman-Kolmogorov equations or from Jacod and Shiryaev (2003, corollary V.4.39a) applied to the point process $(n_t - n_0 + t)/2$ that the total variation of the signed measure $P_t^k - P_t$ tends to zero, i.e.,

$$\|P_t^k - P_t\| = \sup\{|P_t^k(\phi) - P_t(\phi)| : \phi \text{ } \mathcal{F}_t\text{-measurable function on } \Omega \text{ with } |\phi| \leq 1\} \rightarrow 0.$$

The restriction that ϕ is \mathcal{F}_t -measurable can be removed by an approximation argument: for any \mathcal{F}_t -measurable function ϕ_t , one has

$$\begin{aligned} |P^k(\phi) - P(\phi)| &\leq |P^k(\phi) - P^k(\phi_t)| + |P^k(\phi_t) - P(\phi_t)| + |P(\phi_t) - P(\phi)| \\ &\leq |P^k(\phi) - P^k(\phi_t)| + \|P^k - P\| + |P(\phi_t) - P(\phi)|. \end{aligned} \tag{6}$$

Setting k large enough, $\|P_t^k - P_t\|$ can be made smaller than $\epsilon/3$. Then ϕ_t can be set equal to the \mathcal{F}_t -conditional expectation of ϕ under the measure $(P^k + P)/2$. It follows that $\phi_t \rightarrow \phi$ a.s. under P^k and P . By the dominated convergence theorem, the first and third term in the right-hand side of equation (6) are smaller than $\epsilon/3$ when t is large enough. It follows that (6) is arbitrarily small for large enough values of k . Thus P^k converges to P in the total variation norm. ■

Appendix B: Data Appendix

Interpretation of Summaries

These variables are on a scale of -8 to 8.

- -8 implies “I am certain that the death penalty does NOT deter people from committing murder”
- 0 implies “I am not certain whether the death penalty deters people from committing murder”
- 8 implies “I am certain thatt he death penalty DOES deter people from committing murder.”

We used the raw numbers that participants entered. Climate change questions were worded identically.

Prior Beliefs

These variables are on a scale of -8 to 8. To the scale is identical to that described above. We used the raw numbers that participants entered.

Gender

This question was a free response. We coded the variable as a 1 if participants entered “F”, “Female”, “Woman” etc. We coded the variable as 0 if the participant entered “Man”, “M”, “Male” etc. We coded missing gender as 1 if participants left the answer blank, or entered something that was not decipherable, and 0 otherwise.

Additional Demographic Indicators: For each of the following variable categories, participants were required to select an answer to continue. Each of the variables corresponds to a single answer choice within the category that was offered to participants. We coded these variables as 0 if the participant did not select the corresponding box, and 1 otherwise.

Race/Ethnicity

Answer choices include: Black, Chinese, Indian, Other asian, Hispanic, Native American, White, Other race, or Prefer not to answer. Individuals could mark multiple choices.

Religion

Individuals could choose: Buddhist, Hindu, Christian, Jewish, Muslim, Not religious, or Prefer not to answer

Educational Attainment

We asked what was the highest level of education that the participant achieved. Answer choices and corresponding indicator variables were Some High School, High School Graduate, College with no degree, Bachelor’s Degree, Graduate degrees (Master, PhD, etc.), Other and Prefer not to answer.

College - We coded the variable College as equal to 1 if participants selected Bachelor’s degree or Graduate degrees. We coded the variable as 0 if participants did not select Bachelor’s degree or Graduate degree and did not select Prefer not to answer.

Employment Status

Employment status choices include: Employed, A Student, Unemployed and seeking work, Not formally employed and not seeking formal employment, Retired, Other, or Prefer not the answer.

Political affiliation

Indicators were constructed for Democrat, Republican, Independent, and Prefer not to answer

Wages and Annual Income

Hourly wage categories include: \$0.00-\$2.00, \$2.01-\$4.00, \$4.01-\$7.00, \$7.01-\$10.00, \$10.01-\$15.00, \$15.01-\$20.00, \$20.01-\$30.00, \$30.01-\$50.00, \$50.01 or more and Prefer not to answer.

For the purposes of calculating the mean in Table ??, we used the midpoint of each category (e.g. \$1.00 for the first category, \$3.00 for the next category, etc.). We coded the \$50.00 or more category as equal to \$60.00.

Approximate annual income categories include: \$0.00-\$5,000, \$5,001-\$10,000, \$10,001-\$20,000, \$20,001-\$30,000, \$30,001-\$40,000, \$40,001-\$60,000, \$60,001-\$80,000, \$80,001 or more and Prefer not to answer.

For the purposes of calculating the mean in Table ??, we used the midpoint of each category (e.g. \$2,500 for the first category, \$7,500 for the next category etc.). We coded the \$80,000 or more category as equal to \$100,000.

Location

We first defined the boundaries of the continental United States using the following boundary lines: Western boundary = -124.8 degrees; Eastern boundary = -66.9 degrees; Northern boundary = 49.4 degrees; Southern boundary = 24.4 degrees. Then we split it into quadrants along the East-West line = -95.85 degrees and the North-South line = 36.9 degrees.

We used the location coordinates that were recorded based on the participant's IT address to assign each participant to a quadrant: Northwest, Northeast, Southwest and Southeast.

We also collected location type indicator variables: Urban, Suburban, Rural, and Prefer not to answer

Appendix C: Further Results and Evidence Consistent with the Model

Appendix Table 1: Summary of Belief Divergence Results

Lord, Ross, and Lepper (1979)	Experimental subjects were provided with evidence for and against the deterrent effect of the death penalty. Subjects of all beliefs report that the article matching their baseline is more convincing, and students became more confident in their original position.
Darley and Gross (1983)	Subjects were asked to rate a student's academic ability and performance after seeing different videos of the student's playground either a poor-looking inner-city neighborhood or a wealthier-looking suburban neighborhood. Subjects gave lower grades in the inner-city treatment. Subjects who also viewed a video of the child answering a variety of quiz questions (some correctly, some incorrectly, sometimes paying attention, sometimes not) before rating the child displayed even greater divergence.
Plous (1991)	Subjects with varying opinions on nuclear energy and deterrence were provided with articles on the Three Mile Island disaster and a narrowly-averted accidental missile launch. Subjects of all viewpoints expressed increased confidence in their original viewpoints after reading the articles.
Munro and Ditto (1997)	Subjects with high and low levels of prejudice towards homosexuals were presented with two fictional studies on the empirical prevalence of a homosexual stereotype. Follow-up interviews revealed evidence of both biased assimilation and attitude polarization.
Russo, Meloy, and Medvec (1998)	Experimenters sequentially provided subjects with information on two fictional brands. In later stages, once participants have formed preferences, neutral information causes subjects to identify more strongly with their preferred brand.
McHoskey (2002)	Students were randomly selected to review either information supporting the claim that Lee Harvey Oswald acted alone in assassinating John F. Kennedy and or information pointing to a larger conspiracy. Students with extreme opinions intensify their positions when presenting with information supporting their beliefs and relax their beliefs to a lesser degree when confronted with contradictory evidence.
Kahan et al (2007)	Subjects were surveyed on their beliefs about the safety of nanotechnology after half were randomly provided with factual information about risks and benefits. Those who were exposed to information displayed greater polarization than those who were not.
Nyhan and Reifler (2010), Nyhan, Reifler, and Ubel (2013)	These studies detail a series of five experiments in which participants are asked to assess the validity of a false or misleading statement by a politician. In each case, the additional information leads the most-committed members of the targeted subgroup to intensify their misperceptions, rather than weakening them.

Appendix Table 2: Regressions of Interpretations on Priors - Question by Question

	Summary 1 (1)	Summary 2 (2)	Summary 3 (3)	Summary 4 (4)	Summary 5 (5)	Summary 6 (6)
<i>Climate Change</i>						
Prior Belief	0.147*** (0.025)	0.117*** (0.024)	0.084*** (0.028)	0.100*** (0.027)	0.032 (0.026)	0.079*** (0.024)
Constant	5.210*** (0.131)	5.587*** (0.124)	-5.663*** (0.148)	-5.311*** (0.138)	-0.850*** (0.135)	0.200 (0.127)
Observations	608	608	608	608	608	608
<i>Death Penalty</i>						
Prior Belief	0.070*** (0.025)	0.084*** (0.028)	0.127*** (0.030)	0.102*** (0.027)	0.051*** (0.016)	0.036 (0.030)
Constant	5.569*** (0.123)	3.535*** (0.136)	-3.600*** (0.148)	-5.132*** (0.131)	-0.118 (0.079)	-1.823*** (0.147)
Observations	608	608	608	608	608	608

Notes: This table presents estimates of the influence of prior beliefs on interpretation of each individual summary. Column (1) refers to summary 1, column(2) refers to summary 2 etc. *, **, and *** denote significance at the 90%, 95%, and 99% confidence levels, respectively.

Appendix Table 3: Main Results with Controls

	Climate Change		Death Penalty	
	(1)	(2)	(3)	(4)
<i>Pro Abstracts</i>				
Prior Belief	0.069** (0.028)	-0.066 (0.061)	0.080*** (0.025)	0.108 (0.073)
Constant	6.885*** (1.495)	-2.093 (4.279)	9.294*** (2.605)	14.480*** (4.555)
Observations	982	208	982	208
<i>Con Abstracts</i>				
Prior Belief	0.060*** (0.022)	0.149*** (0.050)	0.054** (0.024)	0.064 (0.067)
Constant	-5.423** (2.124)	3.739 (3.594)	-2.024 (2.060)	0.598 (6.650)
Observations	1,473	312	1,473	312
<i>Unclear Abstracts</i>				
Prior Belief	0.030 (0.033)	0.151 (0.097)	0.044* (0.024)	0.042 (0.078)
Constant	11.468*** (2.676)	1.789 (6.193)	-1.254 (2.344)	-10.993* (6.455)
Observations	491	104	491	104

Notes: This table presents estimates of the influence of prior beliefs on interpretation of the summaries by category of summary, controlling for a full set of demographic variables. The Data Appendix contains a description of all demographics and their definitions. Columns (1) and (3) contain those participants who took the survey and were not screened with reading comprehension questions. Columns (2) and (4) contain those participants who were presented with surveys that tested reading comprehension and successfully answered all test questions. Pro summaries include summaries 1 and 2 for both climate change and death penalty. Con summaries include summaries 3,4,5 for climate change and summaries 3,4 and 6 for death penalty. Unclear summaries include summary 6 for climate change and summary 5 for death penalty. *, **, and *** denote significance at the 90%, 95%, and 99% confidence levels, respectively. All standard errors are clustered at the individual level.

References

- [1] Darley, John M. and Paget H. Gross (1983). "A Hypothesis-Confirming Bias in Labeling Effects." *Journal of Personality and Social Psychology*, 44(1): 20-33.
- [2] Hoeffding, Wassily (1963). "Probability Inequalities for Sums of Bounded Random Variables." *Journal of the American Statistical Association*, 58(301): 13-30.
- [3] Jacod, Jean, and Albert Shiryaev (2003). *Limit Theorems for Stochastic Processes*. Berlin: Springer-Verlag.

- [4] Lord, Charles, Lee Ross and Mark Lepper (1979). "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology*, 37(11): 2098-2109.
- [5] Kahan, Dan M., Paul Slovic, Donald Braman, John Gastil, and Geoffrey L. Cohen (2007). "Affect, Values, and Nanotechnology Risk Perceptions: An Experimental Investigation." GWU Legal Studies Research Paper No. 261; Yale Law School, Public Law Working Paper No. 155; GWU Law School Public Law Research Paper No. 261; 2nd Annual Conference on Empirical Legal Studies Paper. Available at SSRN: <http://ssrn.com/abstract=968652>
- [6] McHoskey, John W. (2002). "Case Closed? On the John F. Kennedy Assassination: Biased Assimilation of Evidence and Attitude Polarization." *Basic and Applied Social Psychology*, 17(3): 395-409.
- [7] Munro, Geoffrey D. and Peter H. Ditto (1997). "Biased Assimilation, Attitude Polarization, and Affect in Reactions to Stereotype-Relevant Scientific Information." *Personality and Social Psychology Bulletin*, 23(6): 636-653.
- [8] Nyhan, Brendan and Jason Reifler (2010). "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior*, 32:303-330.
- [9] Nyhan, Brendan, Jason Reifler, and Peter Ubel (2013). "The Hazards of Correcting Myths about Health Care Reform." *Medical Care* 51(2): 127-132.
- [10] Plous, Scott (1991). "Biases in the Assimilation of Technological Breakdowns: Do Accidents Make Us Safer?" *Journal of Applied Social Psychology*, 21(13): 1058-1082.
- [11] Russo, J. Edward, Margaret G. Meloy, and Victoria Husted Medvec (1998). "Predecisional Distortion of Product Information." *Journal of Marketing Research*, 35(4): 438-452.