Optimal Persuasion under Confirmation Bias

Online Appendix

Table of Contents

A Proof of Proposition 1

Suppose that individuals have a prior belief about the outcome of some political reform, $\hat{\mu}_0 \sim \mathcal{N}(\hat{\mu}, \sigma_0^2)$. The voter is then exposed to a prediction, $x \sim \mathcal{N}(\mu_x, \sigma_x^2)$ where $\sigma_0^2, \sigma_x^2 > 0$. Without loss of generality, let $x \geq \hat{\mu}_0 \geq 0$. If we assume that the variance of the message x is known by the individual, we can express the updated belief $\hat{\mu}_1$ as

$$\hat{\mu}_1 = \hat{\mu}_0 \frac{\sigma_x^2}{\sigma_0^2 + \sigma_x^2} + x \frac{\sigma_0^2}{\sigma_0^2 + \sigma_x^2}. \tag{1}$$

by a well-known result of Bayes' theorem.

I will show how three different forms of discounting affects the persuasiveness of extreme messages. I do this by letting the variance of the prediction, σ_x^2, be a function of the distance between the prior, $\hat{\mu}_0$, and the prediction, x. Thus, let

$$\sigma_x^2 = g(\hat{\mu}_0, x) \tag{2}$$

where $g(\cdot) > 0$. I denote the discounting function $g(\cdot)$.

No discounting Let us first consider the case with no discounting, such that $g(\hat{\mu}_0, x) = s^2$, where $s > 0$ is some constant. This yields

$$\hat{\mu}_1 = \hat{\mu}_0 \frac{s^2}{s^2 + \sigma_0^2} + x \frac{\sigma_0^2}{s^2 + \sigma_0^2}. \tag{3}$$

Taking the derivate with respect to x,

$$\frac{d\hat{\mu}_1}{dx} = \frac{\sigma_0^2}{\sigma_0^2 + s^2} > 0, \tag{4}$$

we see that it is constant and always greater than zero. It follows that the persuasion function is not bounded and tends to ∞ as $x \to \infty$.

Linear discounting Let $g(x, \hat{\mu}_0) = a + b(x - \hat{\mu}_0)$ for any $a, b \in \mathcal{R}^+$. Using that $g'_x = b$ and taking the derivate with respect to x,

$$\frac{d\hat{\mu}_1}{dx} = \frac{d}{dx}\left(\hat{\mu}_0 \frac{g(x, \mu_0)}{\sigma_0^2 + g(x, \mu_0)} + x \frac{\sigma_0^2}{\sigma_0^2 + g(x, \mu_0)}\right) \tag{5}$$

$$= \frac{\sigma_0^2(a + \sigma_0^2)}{(a + b(x - \hat{\mu}_0) + \sigma^2)^2} > 0 \tag{6}$$

we find that marginal persuasion is strictly positive. Letting the distance between the prior and the prediction go to infinity,

$$\lim_{x \to \infty} \frac{\sigma_0^2(a + \sigma_0^2)}{(a + b(x - \hat{\mu}_0) + \sigma^2)^2} = \frac{0}{b^2} = 0 \tag{7}$$

we find that marginal persuasion tends to zero. Lastly, we show that the persuasion function is bounded under linear discounting,

$$\lim_{x \to \infty} \left(\hat{\mu}_0 \frac{a + b(x - \hat{\mu}_0)}{\sigma_0^2 + a + b(x - \hat{\mu}_0)} + x \frac{\sigma_0^2}{\sigma_0^2 + a + b(x - \hat{\mu}_0)}\right) = \hat{\mu}_0 + \frac{\sigma_0^2}{b}. \tag{8}$$

The last expression shows that, under linear discounting, persuasion is bounded by the constant σ_0^2/b. Intuitively, the upper bound of persuasion grows as the strength of the prior decreases (σ_0^2 increases) and as confirmation bias decreases (b decreases). Although there is an upper bound of persuasion, assuming that there is no cost to making extreme arguments, it follows from Equation 6 that political actors are still incentivized to use extreme messages since marginal persuasion is strictly positive. However, Equation 8 implies that there is a first mover advantage, since the prior beliefs of the individual bounds maximal persuasion. To see this, suppose that a political actors manages to maximally shift the beliefs of a voter. The best a competing political actor can do is to undo this shift, since the voter's posterior belief after receiving the first message now bounds persuasion.

Exponential discounting Let $g(x, \hat{\mu}_0) = e^{r(x-\hat{\mu}_0)}$ for any $r \in \mathcal{R}^+$. Taking the derivative with respect to x,

$$\frac{d\hat{\mu}_1}{dx} = \frac{d}{dx}\left(\hat{\mu}_0 \frac{g(x,\hat{\mu}_0)}{\sigma_0^2 + g(x,\hat{\mu}_0)} + x\frac{\sigma_0^2}{\sigma_0^2 + g(x,\hat{\mu}_0)}\right) \tag{9}$$

$$= \frac{\sigma_0^2 e^{r\hat{\mu}_0}(\sigma_0^2 e^{r\hat{\mu}_0} + \hat{\mu}_0 r e^{rx} + e^{rx} - rxe^{rx})}{(\sigma_0^2 e^{r\hat{\mu}_0} + e^{rx})^2} \tag{10}$$

we see that whether the sign of the derivative is positive or negative depends on the difference

$$\sigma_0^2 e^{r\hat{\mu}_0} + r\hat{\mu}_0 e^{rx} + e^{rx} - rxe^{rx}. \tag{11}$$

The derivative must initially be positive. Let $x = \hat{\mu}_0$, i.e., the prior belief. This yields

$$\sigma_0^2 e^{r\hat{\mu}_0} + r\hat{\mu}_0 e^{r\hat{\mu}_0} + e^{r\hat{\mu}_0} - r\hat{\mu}_0 e^{r\hat{\mu}_0} = \sigma_0^2 e^{r\hat{\mu}_0} + e^{r\hat{\mu}_0} > 0 \tag{12}$$

since $\sigma_0^2 > 0$. Yet, it is clear that for sufficiently large x the derivative eventually changes sign and becomes negative. This can be seen by re-writing the difference and finding the limit

$$\lim_{x\to\infty} \left(\sigma_0^2 e^{r(\hat{\mu}_0 - x)} + r(\hat{\mu}_0 - x) + 1\right) = -\infty. \tag{13}$$

Consequently, under exponential discounting the persuasion function is non-monotonic. In fact, we can show that the persuasion function is unimodal. Let $t = \hat{\mu}_0 - x$. Since $x \geq \hat{\mu}_0$ it follows that $t \leq 0$. The derivative of the persuasion function equals 0 if and only if

$$\sigma_0^2 e^{rt} + rt + 1 = 0. \tag{14}$$

When $t \to 0$, the expression above approaches $\sigma_0^2 + 1 > 0$. Since $\sigma_0^2, r > 0$ the expression is monotonically strictly increasing in t. It follows that, as t decreases, the expression strictly decreases monotonically. Hence, it equals 0 for a single value of t and becomes

negative for all values thereafter. Thus, under exponential discounting, the persuasion function is unimodal.

Lastly, the limit of the persuasion function as x tends to infinity is

$$\lim_{x\to\infty} \hat{\mu}_1 = \lim_{x\to\infty} \left(\hat{\mu}_0 \frac{e^{r(x-\hat{\mu}_0)}}{e^{r(x-\hat{\mu}_0)} + \sigma_0^2} + x \frac{\sigma_0^2}{e^{r(x-\hat{\mu}_0)} + \sigma_0^2} \right) \tag{15}$$

$$= \lim_{x\to\infty} \frac{\hat{\mu}_0 + \frac{x\sigma_0^2}{e^{x-\hat{\mu}_0}}}{1 + \frac{\sigma_0^2}{e^{r(x-\hat{\mu}_0)}}} = \frac{\hat{\mu}_0 + 0}{1+0} = \hat{\mu}_0. \tag{16}$$

In other words, when the distance between the prediction and the prior grows extreme, individuals increasingly rely on their prior beliefs. In the limit, predictions are completely self-defeating, and individuals do not update their beliefs at all despite the provision of new information. As for linear discounting, the results imply a first mover advantage, since the prior beliefs of the individual again bounds maximal persuasion. □

B Three Observations on Discontinuous Discounting

Following the same setup as in Section A, let

$$\hat{\mu}_1 = \hat{\mu}_0 \frac{\sigma_x^2}{\sigma_0^2 + \sigma_x^2} + x \frac{\sigma_0^2}{\sigma_0^2 + \sigma_x^2}. \tag{17}$$

Let us now consider the case when uncertainty does not increase continuously. Let

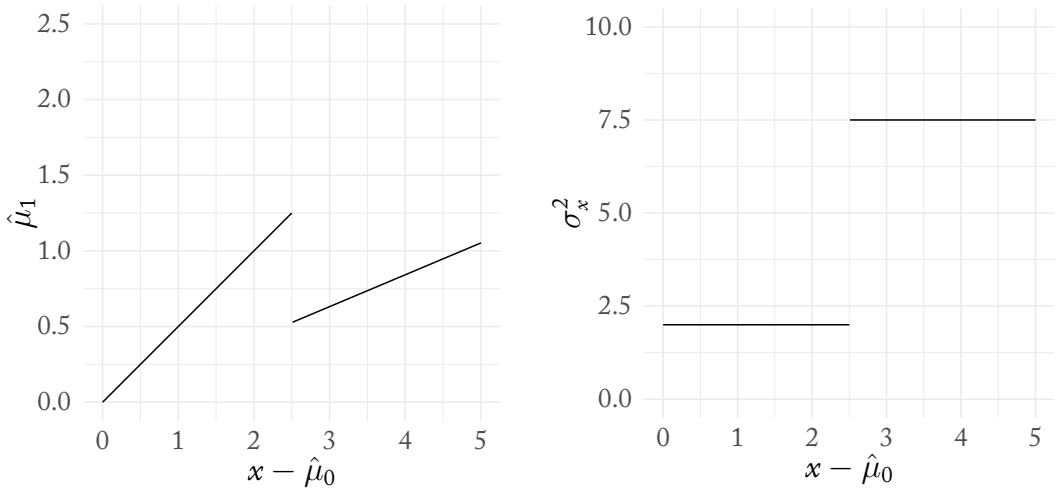
$$\sigma_x^2 = g(\hat{\mu}_0, x) = \begin{cases} \sigma_x^{2\prime} & \text{if } x \leq \bar{x} \\ \sigma_x^{2\prime\prime} & \text{if } x > \bar{x} \end{cases} \tag{18}$$

where $\sigma_x^{2\prime} < \sigma_x^{2\prime\prime}$ and $\bar{x}$ is some constant. This means that uncertainty increases stepwise and not continuously. We can now rewrite our expression of $\hat{\mu}_1$ such that

$$\begin{aligned} \hat{\mu}_1 = &\mathbb{I}_{x \leq \bar{x}} \left[\hat{\mu}_0 \frac{\sigma_x^{2\prime}}{\sigma_0^2 + \sigma_x^{2\prime}} + x \frac{\sigma_0^2}{\sigma_0^2 + \sigma_x^{2\prime}} \right] + \\ &\mathbb{I}_{x > \bar{x}} \left[\hat{\mu}_0 \frac{\sigma_x^{2\prime\prime}}{\sigma_0^2 + \sigma_x^{2\prime\prime}} + x \frac{\sigma_0^2}{\sigma_0^2 + \sigma_x^{2\prime\prime}} \right] \end{aligned} \tag{19}$$

where $\mathbb{I}$ is an indicator function. I plot an example of such discounting in Figure 1.

Figure 1: An Illustration of Discontinuous Discounting



Note: For simplicity, I let $\hat{\mu}_0 = 0$ and $x \geq 0$. This implies that the posterior belief, $\hat{\mu}_1$, is equivalent to the magnitude of updating.

The figure shows how, after the threshold of 2.5, perceived uncertainty jumps, which produces a clear discontinuity in $\hat{\mu}_1$. We also see that the marginal rate of persuasion is lower after the threshold. Below I address three observations on discontinuous discounting in detail.

Observation 1. *Discontinuous shifts in perceived uncertainty imply discontinuities in $\hat{\mu}_1$ at the threshold.*

Consider two cases. For both cases, without loss of generality, let $\hat{\mu}_0 = 0$ and $x \geq 0$. First, let $x = \bar{x}$, which yields

$$\hat{\mu}_1(\bar{x}) = \bar{x}\frac{\sigma_0^2}{\sigma_0^2 + \sigma_x^{2\prime}}. \tag{20}$$

Second, let $x = \bar{x} + \epsilon$, where ϵ is some arbitrarily small but positive number, such that

$$\hat{\mu}_1(\bar{x} + \epsilon) = (\bar{x} + \epsilon)\frac{\sigma_0^2}{\sigma_0^2 + \sigma_x^{2\prime\prime}}. \tag{21}$$

We proceed to determine the difference

$$\hat{\mu}_1(\bar{x}) - \hat{\mu}_1(\bar{x} + \epsilon) = \bar{x}\left[\frac{\sigma_0^2}{\sigma_0^2 + \sigma_x^{2\prime}} - \frac{\sigma_0^2}{\sigma_0^2 + \sigma_x^{2\prime\prime}}\right] + \tag{22}$$

$$\epsilon\left[\frac{\sigma_0^2}{\sigma_0^2 + \sigma_x^{2\prime\prime}}\right]. \tag{23}$$

Since $\sigma_x^{2\prime} < \sigma_x^{2\prime\prime}$, it follows immediately that the first bracket is positive. Letting $\epsilon \to 0$, it is clear that the discontinuity in updating at the uncertainty threshold $\bar{x}$ is equal to the difference $\bar{x}\left[\frac{\sigma_0^2}{\sigma_0^2+\sigma_x^{2\prime}} - \frac{\sigma_0^2}{\sigma_0^2+\sigma_x^{2\prime\prime}}\right]$.

Observation 2. *Discontinuous shifts in perceived uncertainty imply shifts in the marginal rate of persuasion at the threshold.*

Since $\sigma_x^{2\prime} < \sigma_x^{2\prime\prime}$, it follows that marginal persuasion is attenuated after the uncertainty threshold:

$$\frac{d\hat{\mu}_1(\bar{x})}{d\bar{x}} = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_x^{2\prime}} > \frac{\sigma_0^2}{\sigma_0^2 + \sigma_x^{2\prime\prime}} = \frac{d\hat{\mu}_1(\bar{x} + \epsilon)}{d\bar{x}} \tag{24}$$

Observation 3. *A finite number of uncertainty thresholds imply that persuasion is unbounded.*

It suffices to show that persuasion is unbounded after the last uncertainty threshold. Note that, after the last threshold, σ_x^2 is constant. Thus, we can simply apply the same reasoning as for the case of no discounting from the proof of proposition 1 in Section A. It follows that persuasion is unbounded and tends to ∞ as $x \to \infty$.

C Additional Analyses

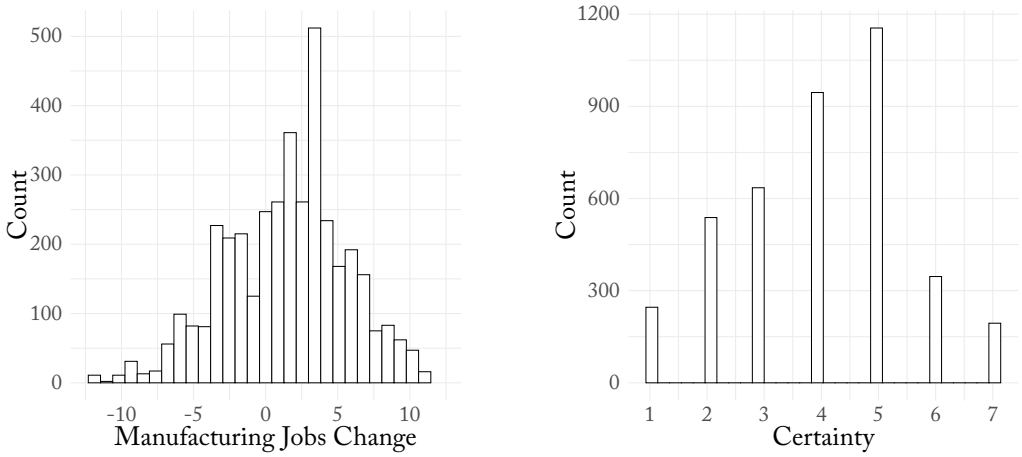
C.1 Consistency with Pre-Analysis Plan

No alterations were made to the experimental design between the registration and the data collection. All heterogeneity analyses and the robustness exercises are available in the appendix. Any *post-hoc* analyses in the main paper are clearly labeled as such. Please note that neither the randomization checks in section C.2 nor the dummy regressions in the heterogeneity analyses were pre-registered.

C.2 Distribution of Prior Beliefs and Randomization Checks

Figure 2 shows the distribution of priors in the sample. A majority of respondents believe that joining the TPP will increase the number of manufacturing jobs in the U.S. The average belief is that it will increase the number of jobs by 1.7 million. The modal response on how certain they feel about this guess is "somewhat certain" and approximately 41% of respondents answered that they were more certain than "neither certain nor uncertain" on the seven-point scale. Please note that this measures the certainty of the initial belief, and does not measure the prior belief about the credibility of the prediction.

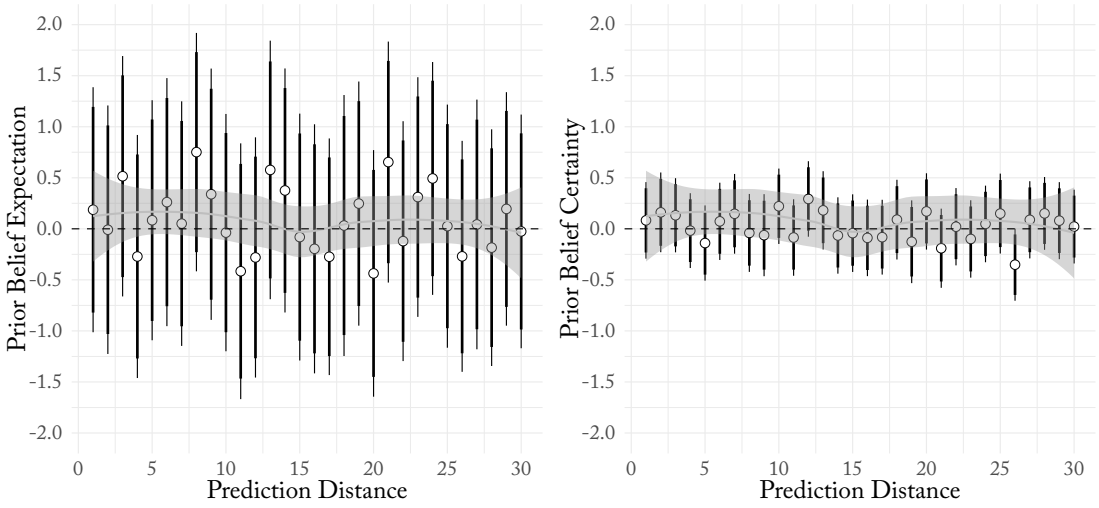
Figure 2: Distribution of Prior Beliefs



Note: The left panel shows that distributions of beliefs about the effect of joining the TPP on manufacturing employment in millions of jobs. The right panel shows the certainty of this prior belief. Higher values of certainty means more certain.

We can also use the priors to do a randomization check. If the randomization was successful, the treatment variable should not be a predictor of the priors. For maximum transparency, I regress every value of the prediction distance treatment, operationalized as dummies, on both the expected change in jobs and the certainty of the expectation. The results are shown in Figure 3.

Figure 3: Effect of Treatment on Prior Belief and Prior Certainty



Note: OLS estimates with robust standard errors. Null prediction distance (i.e. prediction distance = 0) is the reference category. Prior belief expectation refers to beliefs about the effect of joining the TPP on manufacturing employment in millions of jobs. Prior belief certainty is the certainty of this prior belief.

We see that prediction distance does not predict either prior variable, showing that the randomization was successful.

C.3 Bounding the Proportion of Discounters

The theoretical framework of this paper focuses on the individual discounting, while the experiment focuses on outcomes in the aggregate. This makes it difficult to determine heterogeneity of discounting among the respondents. I will, however, be able to conclude some things with certainty in this study.

First, if marginal persuasion is constant in the prediction distance, respondents do not engage in confirmation bias. To see this, consider the case where some respondents engage in exponential discounting and all others do not discount at all. Then marginal persuasion must be decreasing, since the exponential discounters update less at some

point. However, if no decrease in marginal persuasion is observed, this means that none, or very few, discount based on the prediction distance. The results from the experiment showed substantial discounting, and we can thus reject that marginal persuasion is constant among all respondents.

Second, if marginal persuasion becomes negative, at least a substantial proportion of respondents engage in exponential discounting. This follows since the discounting must be sufficiently strong (or the discounters sufficiently many) to counterweigh the positive marginal persuasion of respondents who do not discount. The findings from the experiment, especially for the discontinuous dummy regression, suggests that the marginal effect on beliefs does not become negative, suggesting that respondents do not discount exponentially.

Sorting out the proportions of different discounters is tricky, since it depends both on the treatment effects and the number of respondents who discount in different ways. Still, we can gain some insight into these proportions using back-of-the-envelope calculations.

Assume that there is no discounting when the prediction distance is small, and let this be the estimate of marginal persuasion for individuals who do not discount. I base the estimates for initial and end point marginal persuasion on the square regression controlling for the null prediction distance, which is the model with the best model fit. The initial marginal persuasion is then

$$p'(0.1) = -0.6 + 0.32 \cdot 0.1 = -0.57, \tag{25}$$

that is, the derivative of the fitted second degree polynomial evaluated at $x = 0.1$. We estimate end point marginal persuasion for both discounters and non-discounters as the

derivative of the fitted second degree polynomial evaluated at $x = 3$,

$$p'(3) = -0.6 + 0.32 \cdot 3 = 0.36. \tag{26}$$

To bound the proportions of discounters we must make some further assumptions. First, assume that all respondents who discount do so in the same way. That is, all respondents who discount are either exponential or linear discounters, but our sample does not contain both types at the same time. Second, let the estimate of the end point marginal persuasion be the sum of the marginal persuasion of the share of respondents who do not discount and the marginal persuasion of the share of respondents that do discount. From this information, we can determine, for every share of discounting respondents, what the end point marginal persuasion must be for the discounting respondents. Using the numbers above, and assuming that 25% of the respondents in the sample discount, we get

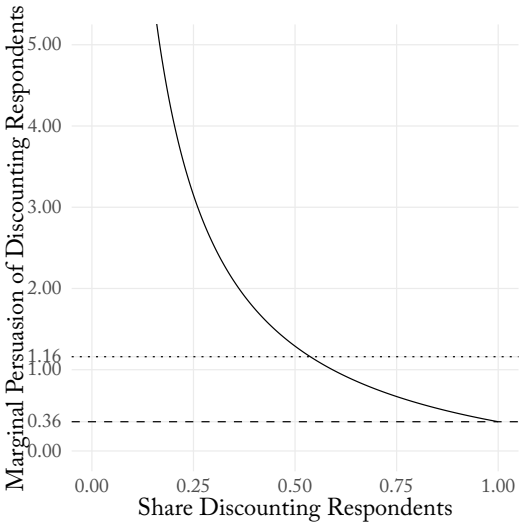
$$0.36 = -0.57 \cdot 0.75 + x \cdot 0.25 \iff x = 3.15 \tag{27}$$

and, analogously, if 75% of the respondents discount

$$0.36 = -0.57 \cdot 0.25 + x \cdot 0.75 \iff x = 0.67 \tag{28}$$

as our estimates of end point marginal persuasion for discounters. In Figure 4, I plot the end point marginal persuasion for discounting respondents as a function of the share of discounting respondents for the values of initial and end point marginal persuasion above.

Figure 4: Estimating the Share of Discounters



Note: Marginal persuasion of discounting respondents refers to the end point marginal persuasion of the respondents who discount. This is determined as the value which makes the weighted sum of the initial and end point marginal persuasion for a given share of discounting respondents equivalent to the initial marginal persuasion.

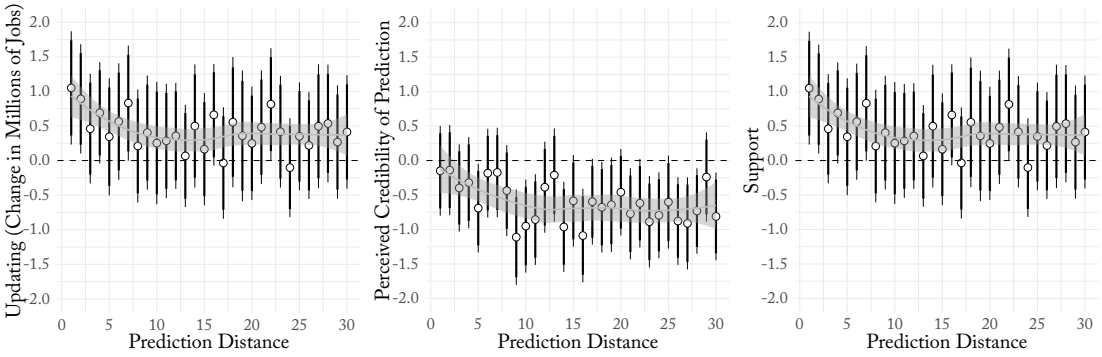
Let us first make two observations on the limits. Since the end slope is positive and the initial slope is negative, at least some respondents must discount. Second, if all respondents discount, end point marginal persuasion is 0.36. One way to bound the estimate is to ask how many must discount for a given value of end point marginal persuasion. If we assume that end point marginal persuasion is equal in magnitude to initial persuasion, i.e., 0.57, then approximately 80% of the respondents must discount. If we go further and assume that end point marginal persuasion among discounters is twice in magnitude, i.e., 1.14, then approximately 54% of the respondents must discount. If only 25% of respondents discount, end point marginal persuasion among discounters must be 3.00, almost six times greater in magnitude compared to initial discounting. Consequently, this analysis suggests that discounting is not a marginal phenomenon

among voters, but applies to a large subset or even a majority of voters.

C.4 Non-Continuous Estimates

In the main paper, I present the results of a dummy regression, where I split the prediction distance variable into six categories. Of course, there is no obvious reason to prefer this partitioning over another, and I therefore present the results from a similar analysis, but where every value of the prediction distance treatment is operationalized as a dummy variable. The results on updating, credibility and support are shown in Figure 5.

Figure 5: Effects of the Factorized Prediction Distance Variable



Note: OLS estimates with robust standard errors. Updating is the difference between posterior and prior beliefs. Null prediction distance (i.e. prediction distance = 0) is the reference category.

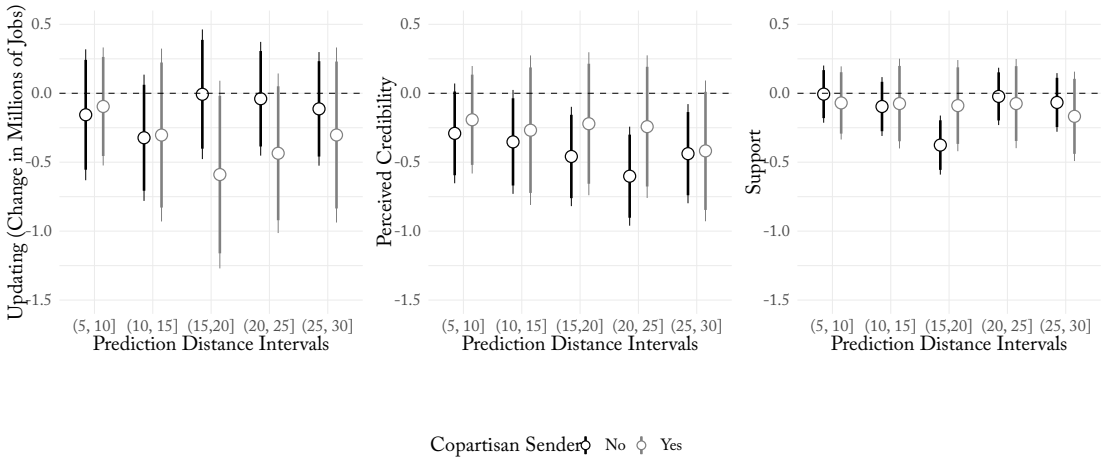
Beginning with the effect of the treatment on updating, we see that almost all point estimates are positive. This is because of the unexpected negative updating for respondents who received the null prediction distance. This is most obvious in comparison with respondents who received a prediction distance of 0.1 million jobs. The difference in updating is, in expectation, approximately 1 million jobs. Except for this curious result, the updating pattern is highly consistent with linear discounting and weak confirmation bias. Respondents initially follow the prediction, but halt at a certain point. We do not see

the same unexpected effect on the perceived credibility of the predictions, shown in the center panel. Perceived credibility decreases as the prediction distance grows. Here, the effect also appears to taper off after a prediction distance of 1-1.5 million jobs. Lastly, the effect on support for the TPP follows a pattern very similar to that of updating.

C.5 Heterogeneity: Partisanship

In the experiment, the senders are Democrats or Republicans in Congress. It is highly plausible that copartisans assess and assimilate predictions differently than, for example, Democratic respondents who receive a prediction from a Republican senders. I explore this heterogeneity by classifying respondents who identify with a certain party as *copartisans* when the senders come from the same party. Independents are always classified non-partisan, even when they lean toward a certain party. I focus on copartisan heterogeneity rather than partisan heterogeneity (i.e. comparing the response of Democratic respondents to predictions from Democratic and Republican senders to increase statistical power). For ease of interpretation, I present the results from a dummy regression, where each dummy has been interacted with an copartisan dummy variable. I show the results in Figure 6.

Figure 6: Copartisan Heterogeneity



Note: OLS estimates with robust standard errors. Updating is the difference between posterior and prior beliefs. Prediction distance values between [0, 5] millions jobs is the reference category.

Beginning with the left panel, we see that copartisan senders appear to have an advantage when influencing their party identifiers. Although copartisans also discount statements from their preferred party, the discounting appears much later compared to when the sender comes from the competing party. This advantage is also reflected in the perceived credibility of the predictions. The effect of prediction distance on copartisan prediction credibility is never statistically significant from zero, in stark contrast to the significant and negative effects of prediction distance when the sender comes from the competing party. Somewhat curiously, however, it is only when the sender is not a copartisan that the sender is able to affect the policy preferences of the voters. However, since respondent partisanship is not randomly assigned, the causal status of this conditional effect is not obvious. For the sake of transparency, I also present the findings from the polynomial regression in Table 1.

Table 1: Heterogeneity: Copartisan Sender

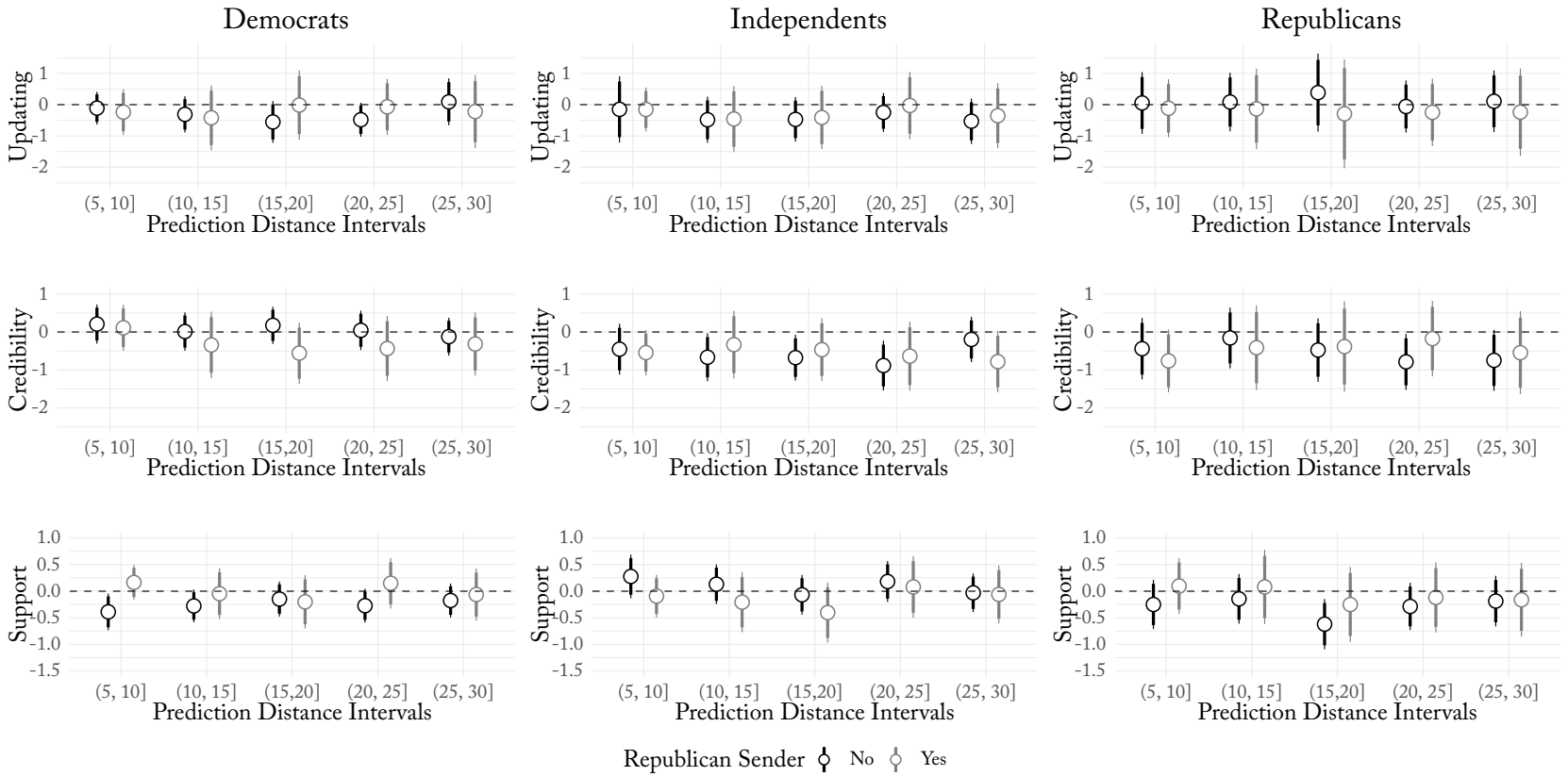
	Updating			Credibility	
Distance	-0.04	-0.53	-1.50	-0.20**	-0.83**
	(0.07)	(0.31)	(0.79)	(0.06)	(0.26)
Distance^2		0.16	0.94		0.21*
		(0.10)	(0.59)		(0.08)
Distance^3			-0.17		
			(0.13)		
Copartisan × Distance	-0.15	-0.18	0.57	0.10	0.63
	(0.11)	(0.43)	(1.05)	(0.09)	(0.35)
Copartisan × Distance^2		0.01	-0.62		-0.18
		(0.14)	(0.82)		(0.11)
Copartisan × Distance^3			0.14		
			(0.18)		
Copartisan	0.26	0.28	0.10	0.65***	0.40
	(0.19)	(0.28)	(0.36)	(0.16)	(0.23)
Prior = Prediction (Dummy)	-0.62	-0.89*	-1.05**	0.27	0.09
	(0.33)	(0.36)	(0.39)	(0.26)	(0.28)
AIC	20515	20514	20516	18819	18816
Observations	4026	4026	4026	4031	4031

Note: OLS estimates with robust standard errors within parentheses. Higher values for updating implies manufacturing job increases and higher values for credibility means more credible predictions.
$^*p < 0.05$,$^{**}p < 0.01$,$^{***}p < 0.001$

In Figure 7, I present the findings from a set of dummy regression, where I have split the sample according respondent partisanship. I also include an interaction for whether the prediction sender is from the Republican or Democratic Party. Similar to the findings for the copartisan heterogeneity, partisan respondents seem to follow the predictions from their party more closely and are more reluctant to negatively assess the

credibility of a prediction when it comes from their favored party. Independents seem to assess both Democratic and Republican senders similarly. A proper interaction model in the polynomial regression framework is available in the replication material.

Figure 7: Partisan Heterogeneity

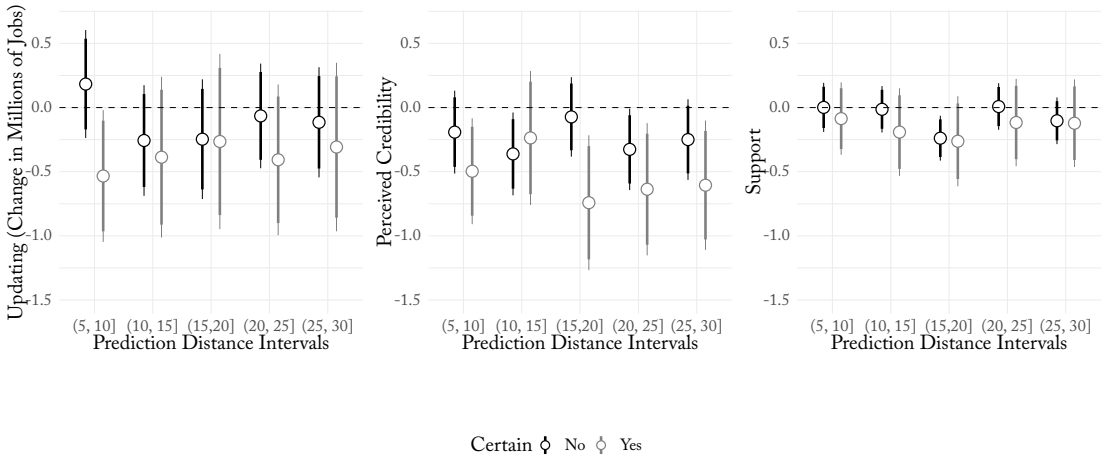


Note: OLS estimates with robust standard errors. Updating is the difference between posterior and prior beliefs. Prediction distance values between [0, 5] millions jobs is the reference category.

C.6 Heterogeneity: Certainty

It is possible that respondents react differently to predictions depending on the strength of their prior beliefs. Specifically, respondents with strong priors may react less to the predictions, and perceive the predictions to be less credible, than respondents with weak priors. I explore this using the same approach as for the copartisan analysis above, but interact the prediction distance dummies with a dummy for certain beliefs, defined as giving an answer of "somewhat certain", "certain" or "highly certain" when asked about the certainty of their prior beliefs. I present the results in Figure 8.

Figure 8: Prior Certainty Heterogeneity



Note: OLS estimates with robust standard errors. Updating is the difference between posterior and prior beliefs. Prediction distance values between [0, 5] millions jobs is the reference category.

Starting with the effects on updating, in the left panel of the figure, we see that respondents who are certain react much more strongly to predictions with small prediction distance compared to respondents who are not certain. However, they also appear to discount more strongly than respondents who are uncertain. The results for credibility show

that certain respondents perceive lower credibility of predictions as prediction distance grows. There are possibly some effects for uncertain respondents as well, but they are much less distinct both in magnitude and precision. Lastly, we see that the conditional effects for the two groups follow each other quite closely on support.

A possible confounder for the effect of certainty is that it may not only capture the certainty of beliefs, but numerical literacy as well. That is, respondents who say that they are certain might actually mean that they are certain giving a numerical estimate of their beliefs and vice versa.

For the sake of transparency, I also present the findings from the polynomial regression in Table 2.

Table 2: Heterogeneity: Prior Certainty

	Updating			Credibility	
Distance	-0.10	-0.53	-0.22	-0.06	-0.38
	(0.07)	(0.32)	(0.81)	(0.05)	(0.22)
Distance^2		0.14	-0.15		0.10
		(0.10)	(0.60)		(0.07)
Distance^3			0.07		
			(0.13)		
Certain × Distance	-0.02	-0.18	-2.47*	-0.17*	-0.31
	(0.11)	(0.42)	(1.03)	(0.09)	(0.34)
Certain × Distance^2		0.05	1.97*		0.04
		(0.14)	(0.81)		(0.11)
Certain × Distance^3			-0.42*		
			(0.18)		
Certain	-0.00	0.07	0.62	1.67***	1.73***
	(0.19)	(0.27)	(0.34)	(0.15)	(0.22)
Prior = Prediction (Dummy)	-0.60	-0.86*	-1.02**	0.44	0.24
	(0.33)	(0.35)	(0.38)	(0.24)	(0.26)
AIC	20499	20498	20495	18580	18580
Observations	4022	4022	4022	4027	4027

Note: OLS estimates with robust standard errors within parentheses. Higher values for updating implies manufacturing job increases and higher values for credibility means more credible predictions.
$^*p < 0.05$,$^{**}p < 0.01$,$^{***}p < 0.001$

C.7 Robustness: Floor Effects

The prediction distance treatment is assigned independent of the prior for most respondents. However, due to the numerical treatment, this is not possible for some respondents. Specifically, respondents who have a prior belief that joining the TPP will decrease the number of manufacturing jobs by more than 9.3 million jobs cannot be as-

signed to such prediction distance values which cases the number of manufacturing jobs to be zero or negative. For individuals who would have impossible predictions, the prediction distance is instead capped at the minimum value. I drop these respondents (n = 67) from the analysis and re-estimate the models. The results are shown in Table 3.

Table 3: Dropping Potential Floor Observations

	Updating						Credibility			
	Pre-Registered			Post Hoc			Pre-Registered		Post Hoc	
Distance	-0.09	-0.35	-0.36	-0.11*	-0.57*	-1.20*	-0.16***	-0.67***	-0.14**	-0.62**
	(0.05)	(0.21)	(0.52)	(0.05)	(0.23)	(0.59)	(0.04)	(0.18)	(0.05)	(0.19)
$Distance^2$		0.09	0.09		0.15*	0.64		0.17**		0.15*
		(0.07)	(0.40)		(0.07)	(0.44)		(0.06)		(0.06)
$Distance^3$			-0.00			-0.11				
			(0.09)			(0.09)				
Prior = Prediction (Dummy)				-0.54	-0.79*	-0.96*			0.44	0.19
				(0.33)	(0.35)	(0.38)			(0.26)	(0.28)
AIC	19727	19727	19729	19725	19723	19724	18610	18603	18973	18605
Observations	3960	3960	3960	3960	3960	3960	3965	3965	3965	3965

Note: OLS estimates with robust standard errors within parentheses. Higher values for updating implies manufacturing job increases and higher values for credibility means more credible predictions.
$^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$

The table shows that the findings are robust from dropping these individuals.

C.8 Robustness: Controlling for Prior Beliefs

As a robustness check, I re-estimate the models including fixed effects for prior beliefs. The results are shown in Table 4.

Table 4: Controlling for Prior Beliefs

	Updating						Credibility			
	Pre-Registered			Post Hoc			Pre-Registered		Post Hoc	
Distance	-0.08	-0.43*	-0.32	-0.11*	-0.73**	-1.39*	-0.16***	-0.58***	-0.14**	-0.54**
	(0.05)	(0.21)	(0.51)	(0.05)	(0.22)	(0.57)	(0.04)	(0.18)	(0.05)	(0.19)
$Distance^2$		0.12	0.02		0.20**	0.72		0.14*		0.13*
		(0.07)	(0.39)		(0.07)	(0.43)		(0.06)		(0.06)
$Distance^3$			0.02			-0.11				
			(0.09)			(0.09)				
Prior = Prediction (Dummy)				-0.72*	-1.04**	-1.23***			0.35	0.14
				(0.31)	(0.32)	(0.35)			(0.26)	(0.28)
AIC	20057	20049	20057	20053	20046	20047	18957	18953	18957	18954
Observations	4027	4027	4027	4027	4027	4027	4032	4032	4032	4032

Note: OLS estimates with robust standard errors within parentheses. Higher values for updating implies manufacturing job increases and higher values for credibility means more credible predictions.
$^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$

The table shows that the results are robust to controlling for prior beliefs. If anything, the results speak clearer in favor of the quadratic regression.

D Experimental Design

D.1 Formalization of Hypothesis Tests

Table 5 shows the formal criteria for identifying if and what type of confirmation bias respondents exhibit from the statistical model. For the expectations, I assume that the predictions are negative (i.e. a decrease in the number of manufacturing jobs) but that the prediction distances are operationalized as absolute values. The dots refer to non-significant coefficients, while $+$ and $-$ refer to significant positive and negative coefficients. For all hypotheses tests, I let $\alpha = .05$. When deriving the hypotheses, I assume that the treatment domain includes extreme enough values for any underlying discounting to emerge.

Table 5: Formalization of Hypotheses for Updating for Negative Predictions

	No discounting	Linear discounting	Exponential discounting
x^1	-	-	-
x^2	·	+	+
x^3	·	·	-

Note: x is the *prediction distance* variable. The outcome variable is updating.

Note that, in the above table, the formal expectations are valid for negative predictions. If predictions are instead positive, the expectations are shown in Table 6, which is the same expectations as presented in the paper.

Table 6: Formalization of Hypotheses for Updating for Positive Predictions

	No discounting	Linear discounting	Exponential discounting
x^1	+	+	+
x^2	·	−	−
x^3	·	·	+

Note: x is the *prediction distance* variable. The outcome variable is the updating.

In Table 7, I show the corresponding formal criteria testing the mechanism, i.e., perceived credibility of the prediction. Higher values means more credible predictions. Under no discounting, there should be no significant effect of prediction distance on perceived credibility of the prediction. Under linear discounting, the relationship should be linear and under exponential discounting, the relationship should be non-linear. I differentiate between linear and exponential discounting by the magnitude and sign of prediction distance squared.

Table 7: Formalization of Hypotheses for Perceived Credibility of Prediction

	No discounting	Linear discounting	Exponential discounting
x^1	·	−	·
x^2	·	·	−

Note: x is the *absolute prediction distance* variable. The outcome variable is perceived credibility where higher values means more credible predictions. Higher values mean less credible predictions.

In addition to the signs of the estimated coefficients, the magnitude of the coefficients matter. Since the functional form is determined by the relative strength of the coefficients it is hard to *a priori* specify how the different forms of discounting bound the coefficients. I evaluate this *a posteriori* by plotting the estimated relationship over the range of the treatment variables for both dependent variables.

D.2 Survey Flow and Administration of Treatment

Survey flow. The different parts of the survey is presented to respondents in the following manner:

1. Introduction to survey
2. Screener
3. Description of TPP. Measurement of priors.
4. Measurement of partisanship, ideology and nationalist attitude.
5. Treatment exposure. Measurement of outcome variables.
6. Debriefing
7. End of survey

Treatment administration. The treatment is administered in the following manner. First, the prior of the respondent is measured with a survey question. This answer is recorded as the variable $\$prior$. After this, the respondent is randomly assigned to a $\$prediction\ distance$ treatment. The treatment is administered to the respondent as a prediction. The prediction is defined as $\$prior + \$prediction\ distance = \$prediction$. The vignette corresponding to $\$prediction$ is then constructed and presented to the respondent.

For example, the respondent has a $\$prior$ that joining the TPP will decrease the number of manufacturing jobs by 0.5 million jobs. The respondent is assigned to a $\$prediction\ distance$ of 0.5 million jobs. This makes the corresponding $\$prediction$ a decrease of 1 million jobs.

The *prediction distance* variable ranges from 0 to 3 million jobs lost in increments of 0.1 million jobs. This yields 31 possible treatment values.

D.3 Respondent Restrictions, Screeners, Missing Values.

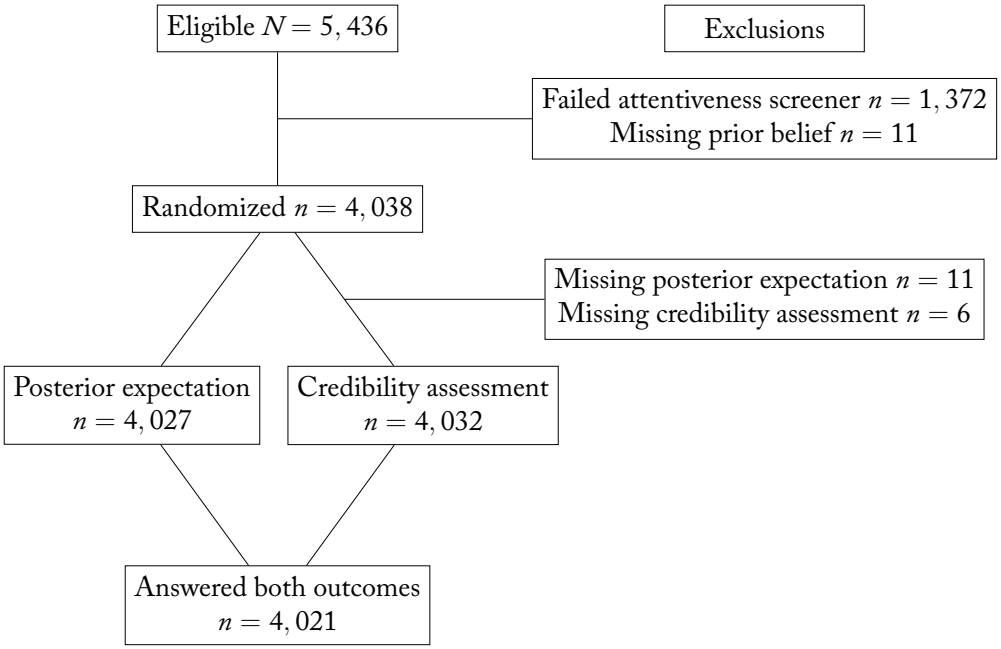
Eligitibility, screeners and restrictions. The population of interest in this study are all US adults. Following the recommendation in Berinsky, Margolis and Sances (2014), I include two screeners. The purpose of screeners is to reduce noise by ensuring that respondents are paying attention to the survey. If respondents do not pay attention to the treatments or the questions, this will lead to attenuation bias. The screeners are included in a battery of agree-disagree statements. The screeners ask the respondents to choose a specific response alternative for data quality purposes. I will only include respondents which successfully pass the screeners in the analysis. The screeners are shown in Table 9.

Out of bound predictions. Since the number of manufacturing jobs cannot be negative, the lowest possible prediction is a decrease of 12.3 millions jobs. Thus, if the prediction, defined as the prediction distance subtracted from the prior, is a negative value, the prediction is capped at a decrease of 12.3 million jobs.

Missing values. Missing values will be handled by listwise deletion. This means that respondents that do not answer the question about prior or posterior beliefs are dropped from the analysis.

Response Rates and CONSORT flow chart. Since Lucid does not use probability sampling, and do not provide information to how many potential respondents they offer to take the survey, calculating the response rate is not possible. Figure 9, however,

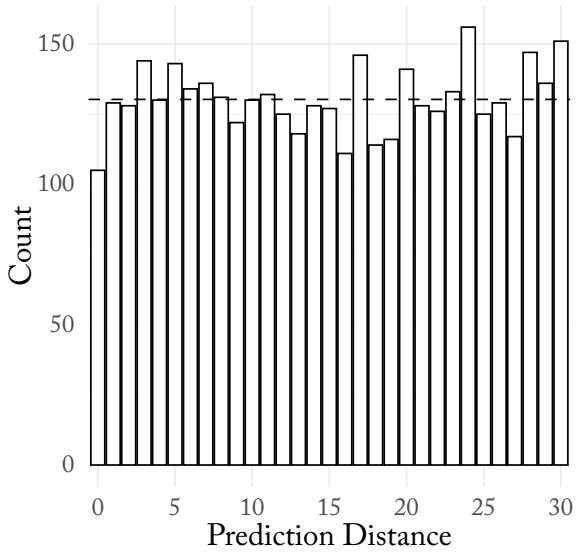
Figure 9: CONSORT Flow Chart



Note: Posterior expectation refers to the posterior belief variable, which is needed for determining the respondent's updating. Credibility assessment refers to the respondent's assessment of the credibility of the prediction.

provides a CONSORT flow chart of how missingness and sample exclusions. 5,436 individuals started taking the survey. Of these, 1,372 individuals failed one or both pre-attentiveness screener, and the survey was terminated. 11 individuals failed to answer the question about prior beliefs. Thus, 4,038 individuals were assigned to treatment. Of these, 4,027 respondents provided a posterior belief and 3,032 individuals also provided a credibility assessment. 4,021 respondents answered both outcomes of interest. Since the treatment is numeric, and thus difficult to visualize in Figure 9, I instead plot the counts for the respective treatment values in Figure 10 below. The figure suggests that the randomization assigned respondents to all treatment values with equal probabilities. A chi-squared test also shows that we cannot reject the null of independence ($\chi^2 = 31.87$, df = 30, $p = 0.37$).

Figure 10: Treatment Assignment Distribution



Note: The figure shows the number of assigned respondents to each treatment value.

D.4 Additional Analyses: Heterogeneity and Robustness

Heterogeneity: Partisanship Since the representative sending the prediction is randomly assigned to be Democrat or Republican, we can examine whether partisans discount predictions differently when they come from an in-group representative compared to an out-group representative. I model this heterogeneity in two ways.

First, I interact a dummy for whether the partisanship of the respondent is aligned with the party membership of the representative with the *prediction distance* variables. That is, Democratic respondents are coded as co-partisans when the sender is a Democratic representative and Republican respondents are coded as co-partisans when the sender is a Republican representative. Independents are never classified as co-partisans.

Second, I interact a dummy for Republican sender with the partisanship of the respondent (using Independents as a reference category) and the *prediction distance*. This

model is less efficient compared to the above model, but allows me to explore whether the partisanship effects differ between Democratic and Republican respondents.

Respondent partisanship is measured with the question in Table 14. I classify leaners as Independents in the main analysis but will explore how the results change when leaners are classified as partisans.

The sample size of the experiment is not calibrated for the heterogeneity analysis, since this does not form the primary analysis in the paper. However, the sample size is large enough to identify small treatment effects. Assuming that 1/3 of the sample receive a co-partisan prediction and that respondents who receive predictions from out-group representatives do not update at all, the experiment has 80% statistical power for expected treatment effects of .15 SDs (no discounting and exponential discounting) and .35-.45 SDs (linear discounting) with 4000 respondents. In section E.4, I present a more detailed power analysis for the main analysis in the paper.

Heterogeneity: Certainty I will examine whether discounting differs between respondents with strong and weak priors. I define respondents with strong priors as respondents who answer that they are "somewhat certain", "certain" or "highly certain" when measuring their priors. All other respondents are defined as having weak priors. I will test for heterogeneity by interacting the treatment variables with the strength of priors variable in the polynomial regressions. I expect that discounting is stronger among respondents who have strong priors.

Note that the number of respondents included in the experiment is not calibrated for the heterogeneity analysis. The heterogeneity analysis is thus exploratory in nature.

Robustness As a robustness exercise, I will drop all respondents who would receive a prediction below the pre-defined floor of the outcome variable when assigned the maximum *prediction distance*. The data from the pilot survey measuring the respondent priors indicates that the share of affected respondents ranges about 1% for the TPP reform.

As a second robustness exercise, I estimate the main models in the paper while adding fixed effects for prior beliefs.

D.5 Vignettes

Table 8: Prior Measurement and Treatment Vignettes

Prior
To increase economic growth, some people think that the U.S. should join the Trans-Pacific Partnership. The Trans-Pacific Partnership is a free trade agreement aiming to increase trade between the U.S. and a number of countries surrounding the Pacific Ocean.

Joining the agreement will make it easier for firms in the U.S. to export their products, but it will also increase competition with firms in other countries.

In the public debate, people disagree on how this will affect manufacturing jobs in the U.S. Some people believe that this will increase the number of manufacturing jobs. Other people believe that this will decrease the number of manufacturing jobs.

The figure shows the number of manufacturing jobs over the last decade. During this period, the lowest number was 11.5 million and the highest number was 13.9 million. The current number is 12.3 million.

Treatment
Consider, once again, the proposal to join the Trans-Pacific Partnership.

[Democrats/Republicans] in Congress predict that joining the Trans-Pacific Partnership, due to increased export opportunities and international competition, will [increase/decrease/leave] the number of manufacturing jobs [by prediction/unaffected].

D.6 Question Wordings

Table 9: Screener Battery

Please indicate what you think about the following statements:
I have a pretty good understanding of the important political issues facing the U.S. today
Generally speaking, most people can be trusted
I'm satisfied with the way democracy works in the U.S. today
Please select "somewhat disagree" for data quality purposes
Please select "strongly agree" for data quality purposes
[Strongly disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Strongly agree]

Table 10: Prior Belief Measurement

Suppose that the U.S. would join the Trans-Pacific Partnership. Do you think that the number of manufacturing jobs would increase, stay the same or decrease?
[Increase, Stay the same, Decrease]
If you had to guess, by how many million jobs do you think it would change?
As a reminder, approximately 12.3 million people are employed in manufacturing today in the U.S.
[slider from -12.3 to + 12.3 million]
How certain do you feel about this guess?
[Highly uncertain, Uncertain, Somewhat uncertain, Neither uncertain nor certain, Somewhat certain, Certain, Highly Certain]

Table 11: Perceived Credibility Measurement

According to the prediction, the number of manufacturing jobs would [decrease by *prediction*/increase by *prediction*/not change] if the U.S. were to join the Trans-Pacific Partnership. How credible do you find this prediction?
[Not at all credible] 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 [Very credible]

Table 12: Updating Measurement

Once again, suppose that the U.S. would join the Trans-Pacific Partnership. By how many million jobs do you think manufacturing employment would increase or decrease?
As a reminder, approximately 12.3 million people are employed in manufacturing today in the U.S.
[slider from -12.3 to +12.3 million]

Table 13: Support for TPP

Consider the proposal to join the free trade agreement the Trans-Pacific Partnership. Do you favor or oppose the proposal?
[Strongly oppose] 1, 2, 3, 4, 5, 6, 7 [Strongly favor]

Table 14: Partisanship

Do you think of yourself as a Democrat, Republican, Independent or something else?
[Democrat, Republican, Independent, Something else]
If Republican or Democrat: Would you consider yourself a strong [Republican/Democrat] or a not very strong [Republican/Democrat]? [Strong, Not very strong]
If Independent or something else: Do you think of yourself as closer to the Republican or Democratic party? [Republican Party, Democratic Party, Neither]

Table 15: Ideology and Nationalist Outlook

Some people think that the U.S. should act less in international terms and concentrate more on its own national problems. Other people think that the U.S. should act more in international terms and help other countries deal with their problems.
Which is closer to the way you feel?
Please tick a box on the scale, where the value 1 means 'U.S. should focus on own problems' and the value 7 means 'U.S. should focus on international problems.'
[1 U.S. should focus on own problems, 2, 3, 4, 5, 6, 7 U.S. should focus on international problems]
Where would you place yourself on the scale below ranging from 'extremely liberal' to 'extremely conservative'?
[Extremely liberal, Liberal, Somewhat liberal, Neither liberal nor conservative, Somewhat conservative, Conservative, Extremely conservative]

D.7 Questionnaire

UNIVERSITY OF GOTHENBURG

Default Question Block

Thank you for your interest in this survey. It should take you about 4 minutes to complete. This study is part of a research project from the University of Gothenburg with the aim of forming a better understanding of people's views in the U.S. today.

It is important to the success of our research that you answer the questions as fully as possible. We check responses in order to make sure that people have read the instructions and responded carefully. There will be some very simple questions in what follows that test whether you are reading the instructions. If you get these wrong, we may not be able to use your data.

All information that you provide will be kept confidential, and no identifiable information will be passed on to a third party.

preamble_attentioncheck

We will start out by asking you a few quick questions to get a sense of your general views and preferences.

attention_check

I have a pretty good understanding of the important political issues facing the U.S. today

Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

Generally speaking, most people can be trusted

Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

I'm satisfied with the way democracy works in the U.S. today

Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

Please select 'somewhat disagree' for data quality purposes

Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

Please select 'strongly agree' for data quality purposes

Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree

reform_introduction

We will now ask you a few questions about a trade reform which is being discussed in the U.S. today.

Please take your time to read the description of the reform carefully before answering the questions. Some of these questions might be difficult to answer, but we would be grateful if you would answer every question to the best of your ability.

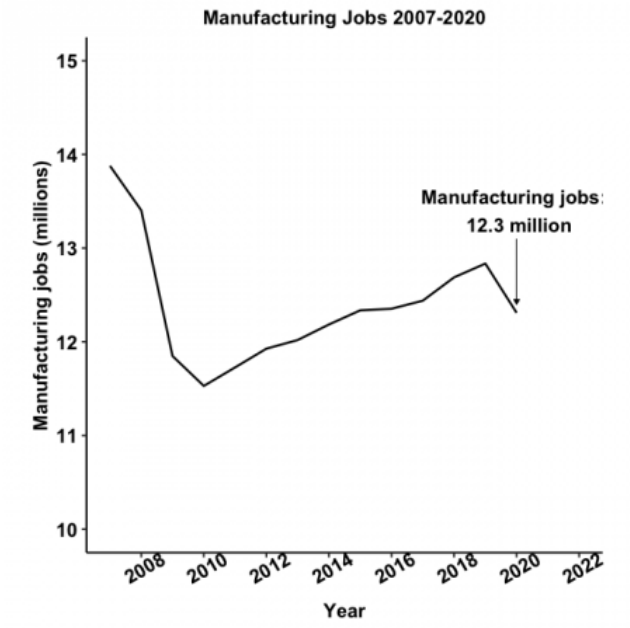
tpp_prior

To increase economic growth, some people think that the U.S. should join the Trans-Pacific Partnership. The Trans-Pacific Partnership is a free trade agreement aiming to increase trade between the U.S. and a number of countries surrounding the Pacific Ocean.

Joining the agreement will make it easier for firms in the U.S. to export their products, but it will also increase competition with firms in other countries.

In the public debate, people disagree on how this will affect manufacturing jobs in the U.S. Some people believe that this will increase the number of manufacturing jobs. Other people believe that this will decrease the number of manufacturing jobs.

The figure shows the number of manufacturing jobs over the last decade. During this period, the lowest number was 11.5 million and the highest number was 13.9 million. The current number is 12.3 million.



If the U.S. would join the Trans-Pacific Partnership, do you believe that the number of manufacturing jobs would increase, stay the same, or decrease?

Increase Stay the same Decrease

If you had to guess, by how many million jobs do you think it would change?

As a reminder, 12.3 million people are employed in manufacturing today.

-12.3 -9.2 -6.2 -3.1 0 3.1 6.2 9.2 12.3

	-12.3	-9.2	-6.2	-3.1	0	3.1	6.2	9.2	12.3
change in millions of jobs									

How certain do you feel about this guess?

Highly uncertain	Uncertain	Somewhat uncertain	Neither uncertain nor certain	Somewhat certain	Certain	Highly certain

preamble_demographics

Next are a few questions about yourself and your political views.

demographics

Generally speaking, do you usually think of yourself as a Democrat, a Republican, an Independent or something else?

Republican	Democrat	Independent	Something else

Would you consider yourself a strong ${q://QID20/ChoiceGroup/SelectedChoices} or a not very strong ${q://QID20/ChoiceGroup/SelectedChoices}?

Strong ${q://QID20/ChoiceGroup/SelectedChoices}	Not very strong ${q://QID20/ChoiceGroup/SelectedChoices}

Do you think of yourself as closer to the Republican Party or to the Democratic Party?

Closer to the Republican Party	Closer to the Democratic Party	Neither

Where would you place yourself on the scale below ranging from 'extremely liberal' to 'extremely conservative'?

Extremely liberal	Liberal	Somewhat liberal	Neither liberal nor conservative	Somewhat conservative	Conservative	Extremely conservative

Some people think that the U.S. should act less in international terms and concentrate more on its own national problems. Other people think that the U.S. should act more in international terms and help other countries deal with their problems.

Which is closer to the way you feel?

Please tick a box on the scale, where the value 1 means ‘U.S. should focus on own problems’ and the value 7 means ‘U.S. should focus on international problems.’

1 - U.S. should focus on own problems	2	3	4	5	6	7 - U.S. should focus on international problems

tpp_treatment

Consider, once again, the proposal to join the Trans-Pacific Partnership.

${e://Field/party} predict that joining the Trans-Pacific Partnership, due to increased export opportunities and international competition, ${e://Field/treatment_phrase}

According to the prediction, the number of manufacturing jobs would ${e://Field/treatment_direction} if the U.S. were to join the Trans-Pacific Partnership.

How credible do you find this prediction?

0 - Not at all credible	1	2	3	4	5	6	7	8	9	10 - Very credible

Once again, suppose that the U.S. would join the Trans-Pacific Partnership. By how many million jobs do you think that manufacturing employment would change?

As a reminder, 12.3 million people are employed in manufacturing in the U.S. today.

	-12.3	-9.2	-6.2	-3.1	0	3.1	6.2	9.2	12.3
change in millions of jobs									

Do you favor or oppose the proposal to join the Trans-Pacific Partnership?

1- Strongly oppose	2	3	4	5	6	7- Strongly favor

final_question

Your feedback is very important to us.

Please let us know if you had any issues in filling out the survey or if something was difficult to understand.

debriefing

E Statistical Power and Analysis

In this section, I present the power and model selection analyses. For the analyses, I simulate data using the Bayesian learning model with a normal prior and likelihood with known variance. For simplicity, I set $\hat{\mu}_0 = 0$. This allows for the prediction x to represent the distance between the prior and the prediction and simplifies the posterior of $\hat{\mu}_1$ to

$$\hat{\mu}_1(x) = x\left(\frac{\sigma_0^2}{\sigma_0^2 + g(x)}\right) \tag{29}$$

where

$$\sigma_1 = \frac{\sigma_0^2 g(x)}{\sigma_0^2 + g(x)} \tag{30}$$

and $g(x)$ is the specific discounting function determining the variance of x. In all analyses below, I am assuming that the variance of the priors is $\sigma_0^2 = 25$ ($\sigma_0 = 5$). I cap the extremity of the prediction x to a 20 unit shift, i.e. 4 standard deviations. Assuming that the priors are normally distributed, a confidence interval of four standard deviations captures 99.994% of the data. Consequently, the set of messages used in the power analysis contains both extreme and non-extreme values. For simplicity, I only focus on positive distances in the analyses below.

E.1 Characterizing and Visualizing Discounting

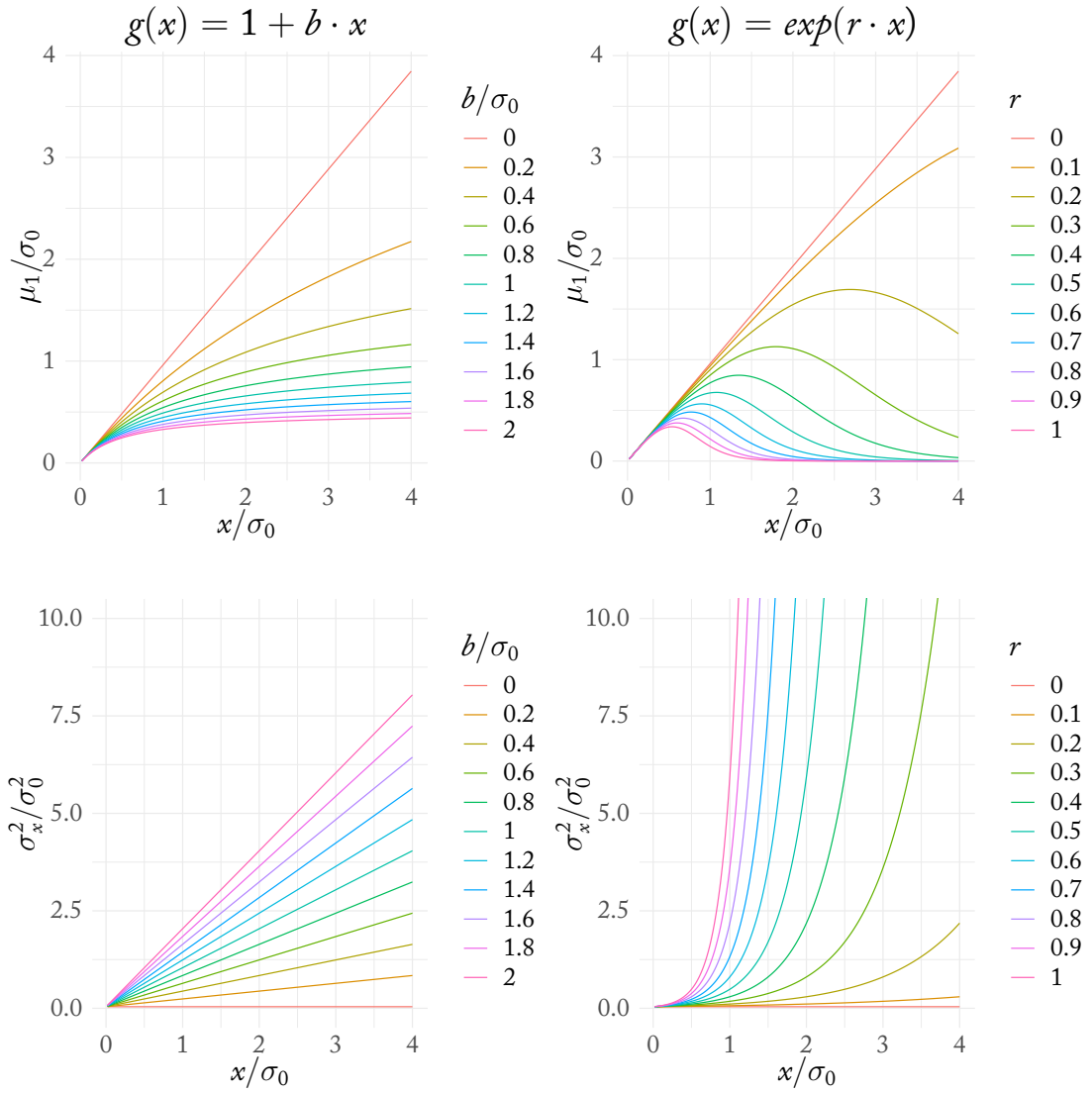
Non-linear treatment effects cannot be summarized in one measure without loss of information. Therefore, I plot the treatment effects for a range of different parameter values for both linear and exponential discounting. I then present a summary measure, the *expected effect of the treatment*, which I use to characterize different treatment effects in the power analysis.

Let us first examine linear discounting, such as $g(x) = a + bx$ for any $a, b \in \mathcal{R}^+$. Linear discounting implies constant marginal discounting. For $x_0 < x_1 < x_2$ and $x_1 - x_0 = x_2 - x_1 \implies g(x_1) - g(x_0) = g(x_2) - g(x_1)$. A one unit shift in the distance between the prior and the prediction increases uncertainty by the same amount regardless of whether the shift is from 1 to 2 or 2 to 3. In the left top panel of Figure 11, I plot $\hat{\mu}_1$ as a function of x for different values of b, all standardized by the standard deviation of the prior. The Figure shows that, even for low levels of prior discounting, e.g. $b = 1$, there is a substantial difference in updating compared to the case of no discounting as represented by the line $b = 0$. The magnitude of updating is almost twice as large under the case of no discounting compared to the case of $b = 1$.[1] The marginal effect of persuasion decreases more rapidly for stronger levels of discounting. This is reflected in the increasing flatness of the curves as b grows. Even for relatively weak discounting, a non-linear pattern emerges. In the bottom left panel, I plot the perceived variance of x for different levels of b. The y-axis shows the perceived variance of x divided by the variance of the prior, σ_x^2/σ_0^2. The perceived variance increases linearly at a constant rate. Even for the weakest form discounting, the perceived variance of the message is approximately twice as high as the variance of the prior.

Let us now examine the exponential discounting function, $g(x) = e^{rx}$ for any $r \in \mathcal{R}^+$. Exponential discounting implies increasing marginal discounting. That is, for $x_0 < x_1 < x_2$ and $x_1 - x_0 = x_2 - x_1 \implies g(x_1) - g(x_0) < g(x_2) - g(x_1)$. This means that the marginal increase of perceived uncertainty is increasing with the *prediction distance*. In the top right panel of Figure 11, I plot the level of updating, μ_1 for different levels of r. The non-monotonic pattern first appears in the interval $r \in [0.1, 0.2]$. When r is

[1] For clarity, $b = 1$ implies that the uncertainty of a prediction, as measured by the standard deviation, increases by one standard deviation of the prior for every five units shift. That is, the message of $x = 5$ is perceived as twice as credible as the message of $x = 10$.

Figure 11: Effect Size and Variances as a Function of Discounting Parameters



Note: Equation 29 is used to determine the magnitude of updating. In all figures, $\sigma_0^2 = 25$ and, consequently, $\sigma_0 = 5$.

below this level, updating is strictly non-negative.[2] The greater the r, the earlier negative marginal persuasion arises. Interestingly, for predictions close to the prior, the rate of discounting is very similar for markedly different values of r. What separates the different rates of exponential discounting is the perceived credibility of predictions further from the prior. This is reflected in the bottom right panel, where I plot the perceived variance of the prediction divided by the variance of the prior as a function of the distance between the prior and the prediction. Initially, the variances are quite similar for different levels of r, but as r becomes greater than 0.2, the variance blows up even for small values of x.

For both linear and exponential discounting, the response to predictions far from the prior belief separates strong from low levels of discounting. Importantly, the difference between the case of no discounting and linear and exponential discounting appears markedly already for low values of discounting.

E.2 Summarizing Non-Linear Treatment Effects

In experimental research, treatment effects are often summarized using Cohen's d. This is defined as the difference in means between two experimental arms divided by the standard deviation of the outcome. However, when the treatment effects are not constant they cannot be summarized into one measure without loss of information. For example, under exponential discounting, the treatment effect is a function of the distance between the prior and the prediction. It is still useful to compute an analogue of Cohen's d for non-linear treatment effects to summarize the strength of the treatment effects for the power analysis. I do this by computing the *expected effect of the treatments* and dividing this by

[2]For clarity, $r = 0.2$ implies that the extreme prediction $x = 20$ is approximately half as credible as the prior belief, while $r = 0.3$ implies that the perceived prediction of the prediction $x = 13$ is approximately half as credible as the prior belief.

the standard deviation of the outcome. I define the expected effect of the treatment as

$$\int_{\Omega} h(x)f(x)dx \tag{31}$$

where Ω is the treatment space of x, $h(x)$ is the treatment effect function, determining the treatment effect of x and $f(x)$ is the pdf of x. For this particular case

$$h(x) = x\left(\frac{\sigma_0^2}{\sigma_0^2 + g(x)}\right) \tag{32}$$

where $g(x)$ is the particular discounting function. Assuming that the prediction is drawn from a uniform distribution, $x \sim Unif(0, 20)$, we can express the expected treatment effect of x as

$$\frac{\sigma_0^2}{20}\int_0^{20} \frac{x}{\sigma_0^2 + g(x)}dx.^{3} \tag{33}$$

In words, equation 31 summarizes the average treatment effect by computing the treatment effect for each possible treatment and weighting the treatment effect by its probability of assignment. It thereby uses information from the whole distribution of treatment effects when summarizing the treatment effect.[4]

E.3 Identifying and Modeling Non-Monotonic Effects

I map the patterns of updating to three distinct functional forms, which I model using polynomial regression. First, if voters do not discount predictions, marginal persuasion

[3] If the treatment variable only takes on discrete, for instance, integer values, we can simplify this to $\frac{\sigma_0^2}{20}\sum_{i=0}^{n=20}\frac{x}{\sigma_x^2+g(x)}$

[4] An alternative measure is the effect of the expected treatment. Assuming that $x \sim Unif(0, 20)$ this implies that $\mathbb{E}[x] = 10$ and the effect of the expected treatment is thus $h(10)\cdot 10$. However, if the marginal treatment effect is, for instance, non-monotonic in x, this measure risks being misleading since it does not drawn on information from the whole distribution of treatments when summarizing the treatment effects.

is constant. This relationship is modeled by a simple linear regression

$$f(x) = \sum_{k=0}^{1} \beta_k x^k + \epsilon. \tag{34}$$

If voters discount predictions linearly, the relationship is modeled by including a square term of *prediction distance*

$$f(x) = \sum_{k=0}^{2} \beta_k x^k + \epsilon. \tag{35}$$

If voter discounting is exponential, this implies a non-monotonic marginal effect. I model this by adding a cubic term of *prediction distance*

$$f(x) = \sum_{k=0}^{3} \beta_k x^k + \epsilon. \tag{36}$$

E.4 Power Analysis: Identifying Effects

In this section I present the power analysis. I generate data using the following equation

$$\hat{\mu}_1 = x\left(\frac{\sigma_0^2}{\sigma_0^2 + g(x)}\right) + \epsilon = x\left(\frac{25}{25 + g(x)}\right) + \epsilon \tag{37}$$

where $g(x)$ is the discounting function and $\epsilon \sim \mathcal{N}(0, \sigma^2 = 25)$. In the paper, the principal outcome of interest is the distance updated $\hat{\mu}_1 - \hat{\mu}_0$. In the analysis, I let $\hat{\mu}_0 = 0$ such that the dependent variable is

$$\hat{\mu}_1 - \hat{\mu}_0 = x\left(\frac{25}{25 + g(x)}\right) + \epsilon. \tag{38}$$

Basing the power analysis on the Bayesian learning model may produce a conservative estimate of the statistical power. Specifically, in the analysis, I let the measured $\hat{\mu}_1$ be a draw from the posterior distribution, in line with Zaller and Feldman (1992), rather than the expected value. If respondents, however, do not randomly sample from the pos-

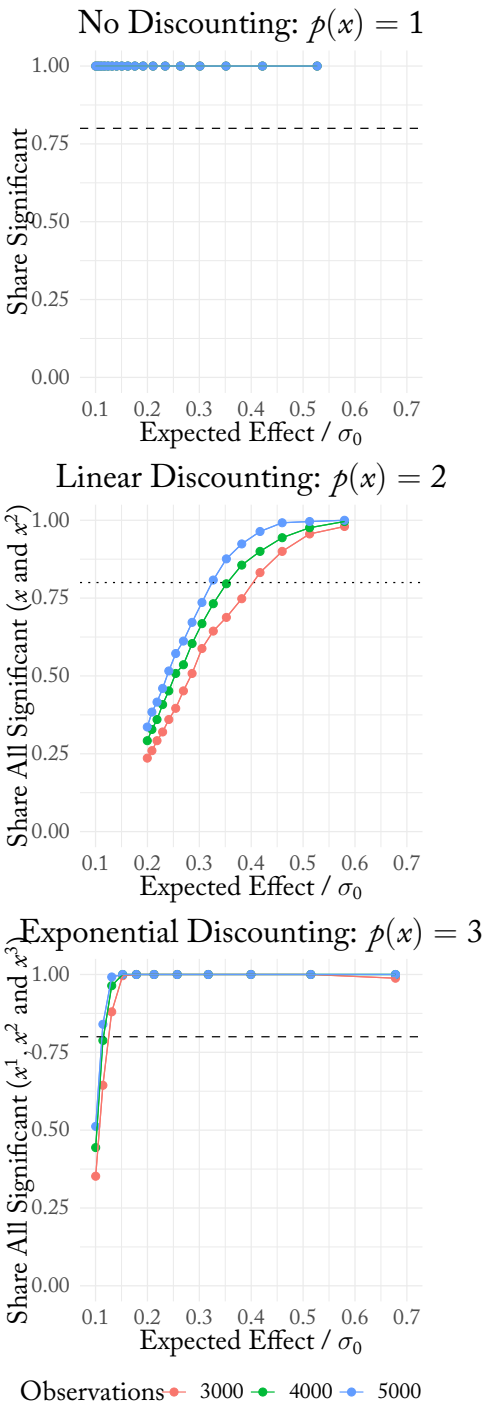
terior distribution but provide the expected value, the variance of the random variable $\hat{\mu}_1 - \hat{\mu}_0$ will be smaller and statistical power will be greater than the power analysis shows. The same is true if the errors in the measurements are correlated. That is, if providing an answer greater than $\hat{\mu}_0$ for the first belief question also increases the probability that the same respondent provides an answer greater than $\hat{\mu}_1$ for the second belief question.[5]

For what treatment effect, summarized as the *expected effect of treatment* divided by the standard deviation of the outcome, do we reach 80% power? I only examine the power of the "correct model", defined as a one, second and third degree polynomial for the case of no discounting, linear discounting and exponential discounting, respectively. I vary the treatment effects by changing the variance of x for the no discounting case and by varying the b and r parameters for the linear discounting and exponential discounting cases. I examine the power for 3000, 4000 and 5000 respondents. The results are shown in figure 12. A significant effect is defined as $p < .05$ for all estimated coefficients excluding the intercept.

For the case of linear discounting, shown in the panel, the model reaches power of almost 100% even for effect sizes smaller than .1 standard deviations. There is little difference between the models with different number of observations. Second, in the center panel, I plot the power for the linear discounting case. This model reaches 80% power for treatment effects of approximately 0.32-0.4 standard deviations. For the case of exponential discounting, shown in the bottom panel, 80% power is reached for a treatment effect of about .12 standard deviations for the third degree polynomial. The reason that the power is lower for the linear discounting case is that it requires a lot of information to distinguish decreasing but positive marginal persuasion from strictly increasing marginal persuasion, which can often be well approximated with a first order

[5] This follows from $Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$.

Figure 12: Power Analysis



Note: Based on 250 simulations per parameter value. I define a significant effect as $p < .05$ Linear discounting is of the form $g(x) = 1 + bx$ and exponential discounting $exp(rx)$ for $b, r > 0$. $p(x)$ refers to the order of the polynomial regression.

polynomial.

Following Cohen (2013), effect sizes of .2 standard deviations are characterized as small and effect sizes of .5 standard deviations as medium. With 4000 observations, the experiment has enough power to identify expected very small treatment effects for the no discounting case, small treatment effects for the exponential discounting case and small to medium size effects for the case of linear discounting. From a political strategy perspective, it is, however, most important to be able to separate exponential discounting from the other two cases, since it is only under exponential discounting that extreme messages are self-defeating. The experiment is well-powered to do this.

E.5 Power Analysis: Model Selection

The above analysis shows that with 4000 respondents the experiment has enough power to identify even small treatment effects. However, the research question requires us to not only identify significant effects but also to examine what model fits the data best.

I examine this using the Bayesian Information Criterion (BIC) (Schwarz 1978) and Akaike's Information Criterion (AIC) (Akaike 1998). Compared to, for instance, the R^2, these statistics penalizes model complexity when evaluating model fit.

I address whether we can use the BIC and AIC to select between different models by generating data according to the same data generating process as the power analysis above and by fitting a set of models to the simulated data. In addition to the linear, quadratic, cubic regressions, I estimate a dummy variable regression where each category corresponds to one standard deviation of the outcome.

I simulate data for a range of different parameter values for the the case of no, linear and exponential discounting. I compute the average BIC and AIC value per model

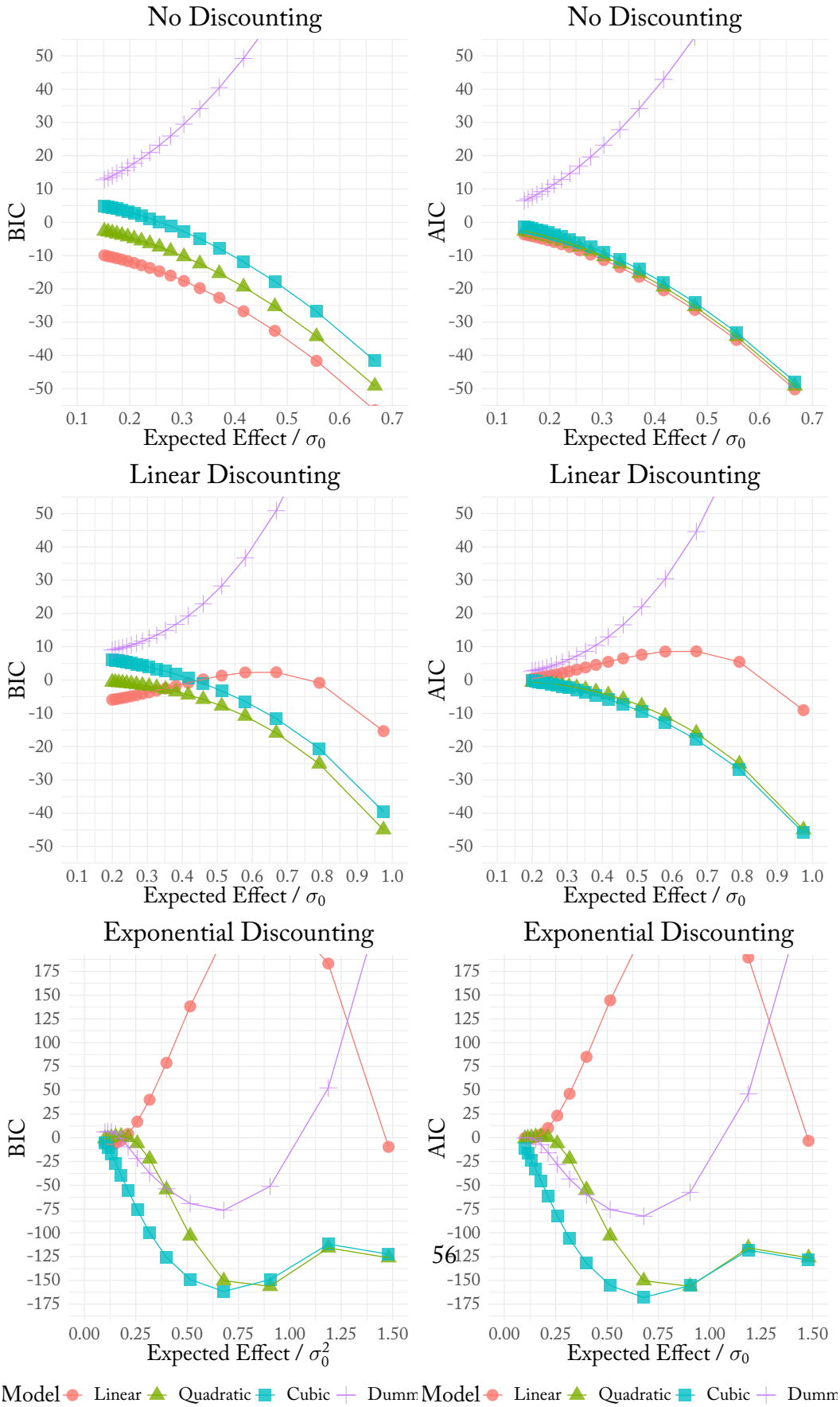
and parameter value for 250 simulations. To facilitate the comparison across different discounting parameters I subtract the mean BIC and AIC score of the models for each parameter value. I use 4000 observations for each simulation.

In Figure 13, I plot the results of the simulations for the three different cases. The y-axis shows the BIC and the AIC value as a deviation from the average of all models for the expected effect value and the x-axis shows the expected effect divided by the standard deviation of the outcome.

The figure highlights three things. First, both the BIC and AIC always correctly identify the correct model when statistical power is sufficiently large. Second, when the treatment effect is very small, the ratio of noise to signal is so great that the model selection effectively is based on the number of parameters in the model. Comparing the left to the right panel, we see that the BIC penalizes additional parameters more strongly than the AIC. This is evident below an expected treatment effect of .15 standard deviations. Third, the BIC and AIC are particularly useful for selecting models when the functional form is non-linear. This is because the linear bivariate regression cannot effectively approximate non-linear effects.

In sum, the analysis shows that it is possible to use the BIC and the AIC to effectively choose between models for small treatment effects. Together with the functional form implied by the estimated effects, I will be able to answer not only whether the respondents exhibit confirmation bias, but also what form of discounting they engage in. In the paper, I will present the AIC. The AIC provides very similar values for the different models when statistical power is low. This prevents us from drawing the false conclusion that, for instance, respondents do not exhibit confirmation bias when in fact the lower BIC score for the linear bivariate regression merely reflects the smaller number of parameters.

Figure 13: Model Selection using BIC and AIC



Note: $N = 4000$ for all simulations. AIC and BIC values are demeaned averages of the models within the expected treatment effect of 250 simulations.

References

Akaike, Hirotogu. 1998. Information Theory and an Extension of the Maximum Likelihood Principle. In *Selected Papers of Hirotugu Akaike*. Springer pp. 199–213.

Berinsky, Adam J, Michele F Margolis and Michael W Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58(3):739–753.

Cohen, Jacob. 2013. *Statistical Power Analysis for the Behavioral Sciences*. Routledge.

Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." *The Annals of Statistics* 6(2):461–464.

Zaller, John and Stanley Feldman. 1992. "A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences." *American Journal of Political Science* 36(3):579–616.