Confidence and Overconfidence in Political Economy[*]

Pietro Ortoleva
California Institute
of Technology
ortoleva@caltech.edu
www.hss.caltech.edu/∼ortoleva/

Erik Snowberg
California Institute
of Technology and NBER
snowberg@hss.caltech.edu
www.hss.caltech.edu/∼snowberg/

April 11, 2012

Very Preliminary and Incomplete
do note quote, do not circulate.

Abstract

This paper studies the role of overconfidence in political behavior. We posit a simple model of overconfidence in beliefs that is founded on correlational neglect. The model predicts that overconfidence leads to ideological extremeness, increased voter turnout, and increased strength of partisan identification. Moreover, the model makes many nuanced predictions about the patterns of ideology in society, and across time. We test these predictions using novel survey data that allows for the measurement of overconfidence, and find that the predicted relationships are statistically and substantively important.

1 Introduction

Ideological differences play an important role in many political economy models. As ideology is often interpreted as political preferences, little work in economics has gone into explaining its roots. However, both colloquially and academically, ideology has also been construed as a form of belief, opening the door to examining how these beliefs come about, and how well-known behavioral biases may influence them (McMurray, 2011).

This paper focuses on the role of overconfidence in the formation of political beliefs or ideology, and how this, in turns affects turnout and attachment to parties. The predictions of our model find support in unique survey data from the 2010 Cooperative Congressional Election Survey (CCES).

Overconfidence is a well-know behavioral bias that has been successfully used by behavioral and financial economists to explain many behaviors (see, for example, Odean, 1998; Daniel, Hirshleifer and Subrahmanyam, 1998; Camerer and Lovallo, 1999; Santos-Pinto and Sobel, 2005). Moreover, some economists have shown that overconfidence may be advantageous, especially in political settings (for example Benabou and Tirole, 2002, 2006; Benabou, 2008). We expand on these literatures by analyzing the implications of overconfidence for political behavior.

In our model of ideology formation, citizens passively learn through their experiences. These experiences are correlated signals about an underlying state of the world, and citizens underestimate that correlation to varying degrees. This results in citizens being overconfident about their beliefs, and this overconfidence is correlated with ideology. To make this concrete, suppose that a citizen notices that many people in his neighborhood are unemployed. If he neglects the fact that he lives in Detroit (or any other location with high unemployment), and believes that these are independent signals of the national economic situation, he will be extremely confident that the national economic situation is quite dire, and perhaps favor generous aid to the unemployed and loose monetary policy. If, on the other hand, he realizes that all of these experiences have a common cause, say a factory shutting down, then he will realize that he has comparatively little information about the national economic situation,

and conclude that although the situation is bad, it is not likely to be dire. This prediction, that overconfidence and ideology are correlated, is supported by our data.

The model makes additional predictions about the structure of ideology in society. First, it predicts that older citizens will be more overconfident, and will also, generally, be more ideologically extreme. This prediction is supported by our data. Moreover, the theory predicts that if more overconfident citizens are, on average, more conservative, ideology should be more correlated with overconfidence for conservatives than for liberals. This appears to be the case in more recent years in the US, but poor-quality historical data is consistent with overconfidence being correlated with ideological extremeness on both the left and the right.

To extend the model to voter turnout, we posit that citizens prefer to vote for the party or candidate whose policy is more likely to be better for them, given the true state of the world. If, however, neither party is much more likely to be better for them, they abstain. This is consistent with Matsusaka (1995) and Degan and Merlo (2011), as well as a large literature in psychology and behavioral economics that documents regret and choice avoidance. Similarly, we model strength of partisan identification as the probability a citizen places on their favored party's policy being better for that citizen.

As more overconfident citizens are more likely to believe that one or the other party is likely to have the right policy for them, they are more likely to turnout to vote *even conditional on ideology*. The opposite conditional statement is also true: more ideological citizens are more likely to vote, even conditional on overconfidence. This means that our model matches the well-known regularity that more ideological people are more likely to vote. The more nuanced predictions are also supported: overconfidence is correlated with turning out to vote, even controlling for ideological extremeness.

The paper is structured so that each set of predictions is quickly tested using survey data. Most tests are conducted using a unique dataset, collected as part of the 2010 Cooperative Congressional Election Study. This data allows for a measure of overconfidence that resembles those generally used in psychology. As noted above, the data indicates that

overconfidence is related to ideological extremeness, voter turnout, and strength of partisan attachment. These empirical relationships are both substantively and statistically important, even when controlling for a large number of other factors.

1.1   Literature

Ideology, voting, and partisan attachment are the subject of a massive political science literature. The work here is most closely related to two sub-literatures. First, it is most closely related to the (small) bayesian literature in political behavior. In particular, we modify the normal learning model to incorporate over-precision.[1] This is a novel approach to modeling overconfidence. Moreover, we provide a way to measure overconfidence on surveys.

Second, the paper is related to the literature that strives to understand how political behaviors are tied to personality traits. While recent work in this literature has focused on the "Big Five" personality traits (see, for example Gerber et al., 2010, 2011), overconfidence is also often seen as a personality trait. However, over-confidence (specifically over-precision) is often found to be orthogonal to the "Big Five" personality traits (Moore and Healy, 2007). Thus, the research described here is complimentary to this literature.

In the economics literature, our model of ideology is most closely related to Blomberg and Harrington (2000), which studies a model in which citizens have priors with heterogeneous means and precisions. Citizens all observe public signals of the true state of the world. Those that start with extreme precise beliefs end up retaining those beliefs, while those with extreme imprecise beliefs converge to the center. While similar in some respects to our model, there are substantive differences. For example, Blomberg and Harrington (2000) predicts that older citizens should be less ideologically extreme. However, consistent with our model, older people appear to be more ideologically extreme. Our model of turnout follows Matsusaka (1995) and Degan and Merlo (2011), which explain turnout and roll-off

---

[1] Although the literature is not large, it cannot be completely reviewed here. Early papers include Zechman (1979); Achen (1992), and others. For a recent review, see the introduction of Bullock (2009).

respectively as being due to either regret or choice avoidance. However, these models focus on uncertainty about candidate's policies, rather than uncertainty about the effects of those policies, as in our model.

Finally, by micro-founding behaviors important to political economy, this project tries to bring work on political behavior and institutions closer together. As such, it is part of a small, but emerging, literature described as either behavioral political science or behavioral political economy (Bendor, Diermeier and Ting, 2003; Callander and Wilson, 2006, 2007, 2008; Bendor et al., 2011; Bisin, Lizzeri and Yariv, 2011).

1.2   What is Overconfidence?

Overconfidence describes a few related phenomena that share the general characteristic that a person thinks some attribute of his or hers, usually information or performance, is better than it actually is. Moore and Healy (2007, 2008) divide overconfidence into three categories that are conflated in most studies: over-estimation, over-placement, and over-precision. Over-placement and over-estimation are quite closely related, and encapsulate the classic statement that, "93% of drivers believe that they are better than average."[2]

In this paper we focus on over-precision: the belief that one's information is more precise than it actually is. There are two reasons for this focus. First, while over-estimation and over-placement often suffer from reversals, this is not the case for over-precision.[3] That is to say, (almost) everyone exhibits over-precision (almost) all the time (Moore and Healy, 2007, 2008). Despite our focus on over-precision, following wide-spread (mis-)usage, we generally use the term overconfidence. Second, over-precision has a very natural interpretation in political contexts: it is the result of people believing that their own experiences are more informative about policy and politics than they actually are.

Overconfidence, as described above, can be either modeled directly, or, we show, as

---

[2]Interestingly, this may be a perfectly rational, see Benoît and Dubra (2011).

[3]Extensive research shows that both of these phenomena suffer from reversals: people tend to perceive themselves as better than they really are when a task is easy, and worse than they actually are when the task is difficult (Erev, Wallsten and Budescu, 1994).

the consequence of correlational neglect, a phenomena that has recently been implicated in (poor) financial decision making (Eyster and Weizsäcker, 2011). That is, a citizen has many experiences, which they believe to be independent signals of the state of the world. However, he neglects that these experiences are all happening to him, and thus are highly correlated. The more that a citizen ignores this correlation, the more precise a signal the citizen believes he has received.[4]

It is worth noting that overconfidence is the subject of a large literature, having first been documented (in the form of over-precision) by Alpert and Raiffa (1969/1982).[5] That literature has documented overconfidence in a wide range of contexts, and among people from a wide range of backgrounds and countries. There are two robust findings that are relevant here. First, men are more overconfident than women (Lundeberg, Fox and Punćcohar̃, 1994). Second, older people are more overconfident than younger people.[6] Moreover, previous studies treat overconfidence as something akin to a personality trait, that is, it seems to be generally believed that some people are simply more overconfident than others.

2 Framework

This section, like those that follow, lays out theory and then immediately moves to data corresponding to the theory. We start with a model of correlational neglect and show it is equivalent to a simpler model that posits overconfidence directly. This is followed by a discussion of the data we have available, and how we use it to construct measures of overconfidence, ideology, and the other theoretical variables discussed in the paper.

---

[4]DeMarzo, Vayanos and Zwiebel (2003) study a model of networks where people exchange messages and update their beliefs on the basis of the messages they receive. In their model, people do not take into account that correlation between multiple messages from the same people, and opinion collapses to a single dimension.

[5]Moreover, some international relations scholars have embraced overconfidence as a likely cause of some wars (Fearon, 1995; Johnson, 2004).

[6]Some studies have found that older people are less overconfident than younger people, however, these usually involve only college students. Studies that examine a larger range of ages find that age is positively correlated with overconfidence, see Hansson et al. (2008).

2.1 Theoretical Framework

There is a unit measure of citizens $i \in [0, 1]$. Each citizen i has a utility for actions which depends on the state of the world. Moreover, each citizen has beliefs over the state of the world. These beliefs are determined by a citizen's experiences, and how much they neglect the correlation between those experiences.

Utilities: Each citizen i is endowed with a standard quadratic-loss utility over actions $a_i \in \mathbb{R}$, which depends on the state of the world $x \in \mathbb{R}$, and a preference bias b_i

$$U(a_i, b_i | x) = -(a_i - b_i - x)^2.$$

That is, for a given state of the world, a citizen's utility is maximized by setting $a_i = b_i + x$. The action a_i may be any political action, such as giving a speech, but for all of this paper it is will be treated as the policy that a citizen would like to see implemented by government.

Citizens are uncertain about the actual state of the world. We postpone briefly a discussion of where beliefs over the state of the world come from, and for now assume that beliefs are given by some c.d.f. $F_i(x)$. The citizen wants to set a to maximize utility, that is, they solve:

$$\max_{a_i} \int U(a_i, b_i | x) dF_i(x) \quad = \quad \max_{a_i} \quad -(a_i - b_i - \mathbb{E}_i[x])^2 - \frac{1}{\tau_{F_i}}$$

where τ_{F_i} is the precision of $F_i(x)$, and hence $1/\tau_{F_i}$ is the variance of $F_i(x)$. Note that the policy preferred by citizen i will be $a_i = b_i + \mathbb{E}_i[x]$.[7] We define this quantity as the ideology $\mathcal{I}$ of the citizen, thus

$$\mathcal{I}_i = b_i + \mathbb{E}_i[x]. \tag{1}$$

Note that in this framework revealed preference provides no way to distinguish the role of preferences and beliefs in ideology.

---

[7]The expectation here is taken over the measure F_i. $\mathbb{E}_i[\cdot]$ is an abuse of notation meant to convey this.

Preferences: Citizens differ in terms of their preference bias b_i. This bias is an i.i.d. draw from a normal distribution with mean π_b and precision τ_b. We write this as

$$b_i \sim \mathcal{N}\left[\pi_b, \tau_b\right].$$

For simplicity, we normalize $\pi_b = 0$ throughout.

Experience and Beliefs: The core of the model is the process by which citizens form beliefs over the state of the world. We propose two processes, one where citizens believe that their experiences are more informative about the state of the world than they actually are (overconfidence), and another where citizens are well-calibrated about the informativeness of individual experiences, but underestimate how correlated their experiences are (correlational neglect), and then show that these processes are equivalent.

In both processes each citizen starts with a normal prior $\mathcal{N}\left[\pi, \tau\right]$ over the state of the world, which has a common mean π, and a common precision τ. For simplicity, we normalize $\pi = 0$ throughout.[8]

Approach I: Overconfidence. In the overconfidence model, citizens then have experiences e_i, which are a signal of the state of the world. In particular, $e_i = x + \varepsilon_i$, where $\varepsilon_i \sim \mathcal{N}\left[0, 1\right]$, $\forall i$, where the normalization that the precision of ε_i is without loss of generality. Despite the fact that every citizen's experience is an equally precise signal of the state of the world, citizens vary in how precise they *believe* their signals are. In particular, citizen i believes

$$\varepsilon_i \sim \mathcal{N}\left[0, \kappa_i\right].$$

If the believed precision $\kappa_i > 1$, then a citizen believes his information is more precise than it actually is. Such a citizen is overconfident in his beliefs about the state of the world. Each citizen's $\kappa_i \in \left[\underline{\kappa}, \overline{\kappa}\right]$ is an i.i.d. draw from c.d.f. F_κ.

---

[8] All of our results hold when $\pi_i | \kappa \sim \mathcal{N}[0, \tau_\pi]$ and $\tau | \kappa \sim F_\tau(\cdot)$ over $\left[\underline{\tau}, \overline{\tau}\right] \in (0, \infty)$. We refrain from doing so for simplicity. Overconfidence κ will be defined below.

Approach II: Correlational Neglect. In the correlational neglect model, citizens have multiple experiences over time, $e_{it} = x + \varepsilon_{it}$, $t \in \{1, 2, \ldots, n\}$. The error ε_i is distributed according to a mean-zero multinomial normal with covariance matrix

$$\Sigma_{\varepsilon_i} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} . \quad \text{However, citizen } i \text{ believes that } \Sigma_{\varepsilon_i} = \begin{pmatrix} 1 & \rho_i & \cdots & \rho_i \\ \rho_i & 1 & \cdots & \rho_i \\ \vdots & \vdots & \ddots & \vdots \\ \rho_i & \rho_i & \cdots & 1 \end{pmatrix} .$$

Note that as each ε_{it} has unit variance, the covariance between any two signals is the same as the correlation, ρ.

Intuitively, when $\rho_i < \rho$ a citizen underestimates, or neglects, the amount of correlation between his or her experiences. In particular, when $\rho_i = 0$ and $\rho = 1$, the citizen will have the same experiences over and over again, but believe each experience provides completely new information about the state of the world. Intuitively, if such a citizen has two experiences, they will treat those experiences as independent signals even though the second provides no new information. To an outside observer, it would be as if the citizen had received a single signal, but had interpreted it as being twice as precise as it actually was. Indeed, this intuition generalizes.[9]

Lemma 1 *The correlational neglect model is equivalent to the overconfidence model with* $e_i = \dfrac{1}{n} \displaystyle\sum_{t=1}^{n} e_{it}$ *and* $\kappa_i = \dfrac{n}{1 + (n-1)\rho_i}$.

Proof. All proofs can be found in the appendix. ∎

As the overconfidence model is somewhat simpler, in most cases we will discuss results, and intuition, in terms of that model. However, predictions dealing with age are most

---

[9]Alternatively, we could model the state of the world in a multi-dimensional space with multi-dimensional errors over time, and citizens either underestimate the amount of correlation between dimensions, or across time, or both. This produces similar results throughout, although there is no exactly equivalent overconfidence model, in the style of Lemma 1. For simplicity, we restrict ourselves to a single-dimensional issue space.

naturally thought of in terms of the correlational neglect model.

2.2 Data

The data used in this paper comes from two different sources. The first source, the Harvard and Caltech modules of the 2010 Cooperative Congressional Election Study (CCES) is unique (as far as we are aware) in that it allows a survey-based measure of overconfidence in beliefs. Unfortunately, it is just a single cross-section of responses. Therefore we supplement this data with analyses, in Section 3.4, from the American National Election Study (ANES) cumulative data file, which has greater coverage across time, but only a crude measure of overconfidence. As the ANES data is well-known, this section focuses on the data from the CCES.

The CCES is an annual cooperative survey. Participating institutions purchase a "module" of at least 1,000 responses to 10–15 minutes of survey questions. In addition, every respondent across all modules are asked the same battery of basic demographic and political questions. The complete survey is administered on-line by Knowledge Networks. The survey strives to be nationally representative across each module by balancing on demographics.

2.2.1 Overconfidence

The most important feature of this data is that it allows for a measure of overconfidence. This is constructed from four questions about respondent confidence in their guesses about four factual quantities, adjusted for how accurate a survey respondent's answers to the factual questions are. In particular, as part of another set of studies, respondents were asked their assessment of the current unemployment and inflation rate, and their assessment of what the unemployment and inflation rate would be a year from the date of the survey (Ansolabehere, Meredith and Snowberg, 2010, 2011).

After each factual question, respondents were asked their confidence about their answer to the factual question. In particular, they were asked:

How confident are you in your estimate?

- No confidence at all
- Not very confident
- Somewhat unconfident
- Somewhat confident
- Very confident
- Certain

with the first response coded with the numeric value 1, and the final with a numeric value 6.

We use the first principal component of the answers to these four questions as a measure of confidence.[10] However, a respondent's confidence reflects both their knowledge about the subject area in question, as well as overconfidence. In order to transform our measure of confidence into a measure of overconfidence, we regress the confidence measure on a fourth order polynomial of respondent accuracy in assessing each of the four of the factual questions (16 variables in all).[11] The residual from this regression is thus purged of a respondent's actual knowledge about unemployment and inflation rates, and thus can be used as a measure of overconfidence. The appendix discusses other ways we have tried to measure overconfidence.[12] In keeping with previous research, and, as we will see, a prediction of Proposition 3, this measure strongly correlated with an indicator for the respondent's age (0.20), and with a respondent's gender (0.24).

---

[10]Responses to these questions are highly correlated and all of the results hold using any one of the four questions in isolation. The first principal component weights each of the four questions approximately equally.

[11]In keeping with the treatment of these factual questions in Ansolabehere, Meredith and Snowberg (2010, 2011), we topcode responses to the unemployment and inflation questions at 25. This limits how inaccurate a respondent can be. We are able to assess the accuracy of respondents' assessments of future unemployment and inflation rates as these analyses were done more than a year after the data were collected.

[12]Note that the data we use to elicit overconfidence is similar to that used in psychology. However, there are several important differences. In particular, psychology studies typically elicit actual confidence intervals, and do so for a very large number of factual questions: up to 150 (see, e.g. Alpert and Raiffa, 1969/1982; Soll and Klayman, 2004). Whether confidence intervals can be used on surveys (as opposed to the highly selected, and quite sophisticated subjects in most psychology experiments) is an open question (see, e.g. Juslin, Wennerholm and Olsson, 1999; Rothschild, 2011). The appendix contains some preliminary data on this question. Note that the method we use to control for information of respondents is considerably more conservative than those used in psychology, see Moore and Healy (2007, 2008).

2.2.2 Dependent Variables

The predictions in this paper concern three dependent variables: ideology, turnout, and strength of partisan identification.

Ideology: This study uses three different measures of ideology. The first is the scaled ideology measure of Tausanovitch and Warshaw (2011). This measure is generated by using item response theory (IRT) to scale responses to eighteen issue questions asked on the CCES (for example, questions about abortion and gun control). A similar process generates the nominate scores used to evaluate the ideology of members of Congress (Poole and Rosenthal, 1985). One particular feature of this measure is worth noting here: on questions with more than two possible answers, the possible answers are split into two groups, one group of answers being coded as for, the other against.[13] That is, respondents who give more extreme responses are coded the same way as those who give a more moderate response in the same direction. This eliminates concerns that our results concerning ideological extremeness are driven by respondents who simply like to choose extreme answers on surveys.

We also employ two measures of self-reported ideology. The CCES twice asks respondents to report their ideology: from extremely liberal to extremely conservative. The first elicitation is when the subject agrees to participate in surveys (on a five point scale), and the second when taking the survey (on a seven point scale). We normalize each of these measures to be on the interval $[-1, 1]$, and then average them. Those that report they "don't know" are either dropped from the sample, or treated as moderates (0). Results are presented for both treatments of those that "don't know". These self-reported measures are imperfectly correlated with scaled ideology (0.44).

To generate measures of ideological extremeness, these measures are folded around the midpoint of the ideological scale.

---

[13]If an issues question had an odd number of responses, the middle response is coded as either for or against randomly for all respondents.

Turnout: Turnout is a self-report of whether or not the respondent reported having voted in 2010. The 2010 CCES will eventually contain verified voter turnout indicators, which we plan to incorporate.

Partisan Identification: At the time of the survey, respondents are asked whether they identify with the Republican or Democratic Party, or consider themselves to be an independent. If they report one of the political parties, for example the Democrats, they are then asked if they are a "Strong Democrat" or "Not so Strong Democrat". Those that report that they are an independent are asked if they lean to one party or the other, and are allowed to say that they do not lean towards either party. Those that report they are a strong Democrat or Republican are coded as strong partisan identifiers. Independents (those that do not lean towards either party) are coded as either strong party identifiers, weak party identifiers, or left out of the data. Results are presented for all three resultant measures.

2.2.3 Controls

Our theory makes no predictions about which controls should be included in tests of the propositions that follow. Thus, we follow a "kitchen sink" approach. Although the controls are not theoretically motivated, they are useful in comparing the effect size of overconfidence on various independent variables to demographics.

The CCES provides many demographic controls as categories: for example, rather than providing years of education, it groups education into categories such as "Finished High School". Thus, we introduce a dummy variable for each category in each of the demographic controls. We also include a category for missing data for each variable. The controls, and number of categories they contain, can be found in Table 1.

3 Ideological Extremeness

The first set of theoretical and empirical results concern ideological extremeness.

Table 1: Controls used in statistical tests.

Type	Number of Categories
Union / union member in household	6 categories
Income	15 categories
Education	6 categories
Race	8 categories
State–including DC	51 categories
Employment status	9 categories
Church attendance	7 categories
Age	73 categories
Gender	2 categories
Total	**177 categories**

3.1 Ideological Extremeness

Define ideological extremeness as the ideological distance from the midpoint $\mathcal{E} = |\mathcal{I}|$.[14] Our first prediction is that:

Proposition 2 *Ideological extremeness and overconfidence are correlated ($\rho_{\mathcal{E},\kappa} > 0$).*

The intuition in this case is quite simple: consider two citizens with the same preference bias $b = 0$ and the same experience $e \geq 0$, but two different levels of overconfidence κ_1 and κ_2, with $\kappa_1 > \kappa_2$. The distribution of a citizen's beliefs after his or her experiences, will be distributed according to

$$\mathcal{N}\left[\frac{\kappa e_i}{\tau + \kappa}, \tau + \kappa\right]. \tag{2}$$

That is, after his or her experiences, each of our two citizens citizen will have ideology given by

$$\mathcal{I} = b_i + \mathbb{E}_i[x] = \frac{\kappa e}{\tau + \kappa} \qquad \text{and thus} \qquad \frac{d\mathcal{I}}{d\kappa} = \frac{\tau e}{(\tau + \kappa)^2} \geq 0. \tag{3}$$

---

[14]If $\pi + \pi_b \neq 0$, then ideological extremeness would be defined as $|\mathcal{I} - \pi - \pi_b|$.

That is, the more overconfident citizen (κ_1) is more ideologically extreme. Intuitively, the citizen that believes his experience is a better signal of the state of the world updates more on that signal, becoming more extreme.

In order to show a correlation, however, we need to account for the full range of preference biases and experiences in the population. Focusing on the distribution of beliefs in a population with the same overconfidence κ, we have

$$
\begin{aligned}
\text{Prob}((\mathbb{E}_i[x]|\kappa) < y) = \text{Prob}\left(\frac{\kappa(x+\varepsilon)}{\tau+\kappa} + \pi < y\right) &= \text{Prob}\left(\varepsilon < \frac{\tau+\kappa}{\kappa}\left(y - \pi - \frac{\kappa x}{\tau+\kappa}\right)\right) \\
&= \Phi\left[\frac{\tau+\kappa}{\kappa}\left(y - \pi - \frac{\kappa x}{\tau+\kappa}\right)\right] \\
\Rightarrow \quad \mathbb{E}_i[x]|\kappa &\sim \mathcal{N}\left[\frac{\kappa x}{\tau+\kappa}, \left(\frac{\tau+\kappa}{\kappa}\right)^2\right], \quad (4)
\end{aligned}
$$

where $\Phi[\cdot]$ denotes the c.d.f. of the standard normal distribution. Note that as the precision of this distribution is decreasing in κ, this implies the variance of the distribution is increasing in κ.

This is illustrated in Figure 1. From the figure it is clear that as one moves further away from the ideological center, citizens are more likely to be more overconfident. Thus, ideological extremeness and overconfidence are correlated.

We can immediately test this proposition using survey data. In particular, Table 2 presents the results of regressing ideological extremeness, derived from the scaled ideology measures of Tausanovitch and Warshaw (2011), on our overconfidence measure.

Two patterns are immediately apparent. First, the relationship between ideological extremeness and overconfidence is extremely statistically robust, no matter what additional (non-theoretically motivated) controls are added to the regressions. Second, the controls that lead to the greatest attenuation of the coefficient on over-confidence are gender and age, two factors that previous research has found to be robustly correlated with overconfidence. In particular, including controls for race, education, income, and union membership attenuates the coefficient on overconfidence only slightly, whereas controls for gender and age cuts the

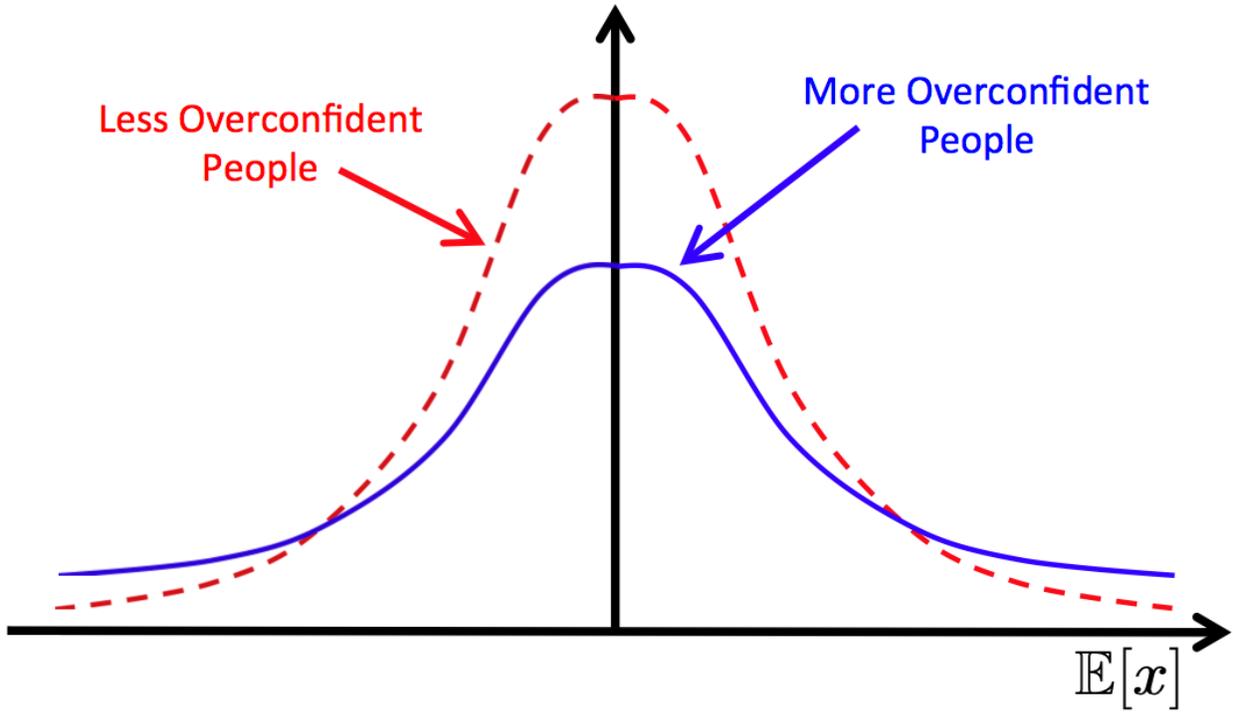Figure 1: Overconfidence and Ideological Extremeness are Correlated

Table 2: Ideological extremeness is robustly related to overconfidence.

Dependent Variable: Scaled Ideological Extremeness					
Overconfidence	0.13***	0.11***	0.10***	0.076***	0.056***
	(.016)	(.015)	(.015)	(.016)	(.016)
Race, Education Income, Union		Y			
Employment, Church, State			Y		
Gender and Age				Y	
All Controls					Y
N			2,803		

Notes: ***, **, * denote statistical significance at the 1%, 5% and 10% level with standard errors in parenthesis.

Table 3: Self-reported ideological extremeness is robustly related to overconfidence.

Treatment of "Don't Know"	Centrist		Missing	
Overconfidence	0.099***	0.078***	0.081***	0.071***
	(.015)	(.016)	(.015)	(.016)
All Controls	N	Y	N	Y
N	2,934		2,774	

Notes: ***, **, * denote statistical significance at the 1%, 5% and 10% level with standard errors in parenthesis.

relationship nearly in half. This last finding is reassuring: the controls that really matter are those that have found to be correlated overconfidence in other work.

In Table 3 we study similar relationships for self-reported ideology. As discussed in Section 2.2.2, there are two measures that depend on whether those that answered they "don't know" their ideological disposition are treated as centrist, or removed from the data. Table 3 considers both measures, and shows that the robust relationship found in Table 2 between ideological extremeness and overconfidence also exists in self-reported measures of ideology.[15]

One other pattern in Table 3 is worth noting: classifying those who report they don't know their ideological disposition as centrist increases the correlation overconfidence and ideological extremeness. This appears intuitive: those that express a low level of confidence about their answer to factual questions are also likely to be relatively less confident about their ideological leanings.

While the relationship between ideological extremeness and overconfidence is clearly statistically robust, is it substantively important? Table 4 suggests the answer is yes. In particular, it asks how a one-standard-deviation change in a number of controls associates with a one-standard-deviation change in ideological extremeness. As the table shows, the

---

[15]In Footnote 14 of Kuklinski et al. (2000) the authors note a strong correlation (0.34) between partisan strength and misinformation. Misinformation in that study is similar to confidence in incorrect opinions. This is the only similar empirical result we have found.

Table 4: Overconfidence is a substantively important predictor of ideological extremeness.

A one standard deviation change in _____ is associated	with a _____ standard deviation change in ideological extremeness.	
	Minimum	Maximum
Income	13%	24%
Education	12%	19%
Race (Black)	12%	17%
Church attendance	7%	11%
Age	18%	22%
Gender	6%	17%
Overconfidence	**8%**	**16%**

Notes: The minimum and maximum effect size come from regressions with no (other) control variables, and all other control variables respectively. In order to compute the effect size of categorical variables, such as education, they are entered linearly in the regressions.

effect size of overconfidence is similar in size to education, race, and gender; slightly smaller than income, and age; and larger than church attendance.[16]

3.2 Age

As noted in the introduction, and the previous section, the correlational neglect model allows us to make predictions about how age, overconfidence and ideology are related. In particular:

Proposition 3 *Older citizens are more overconfident. Further, if $\rho \geq \frac{1+\rho_i\tau}{1+2\tau-\rho_i\tau}$ then ideological extremeness is increasing with age.*

To build intuition, consider the case when $\rho_i = 0$ and $\rho = 1$, that is, when experiences are perfectly correlated, but citizens believe that the signals are uncorrelated. In this case, each signal is identical, so it will make the citizen more confident, without increasing his information. Moreover, each signal makes a citizen more extreme, as his posterior shifts closer and closer to the signal. In general, the condition in Proposition 3 suggests that

---

[16]The range of effect sizes for overconfidence is wider than most demographic controls because we tended to keep controls that attenuated the coefficient on overconfidence.

Figure 2: Age, Overconfidence, and Ideology
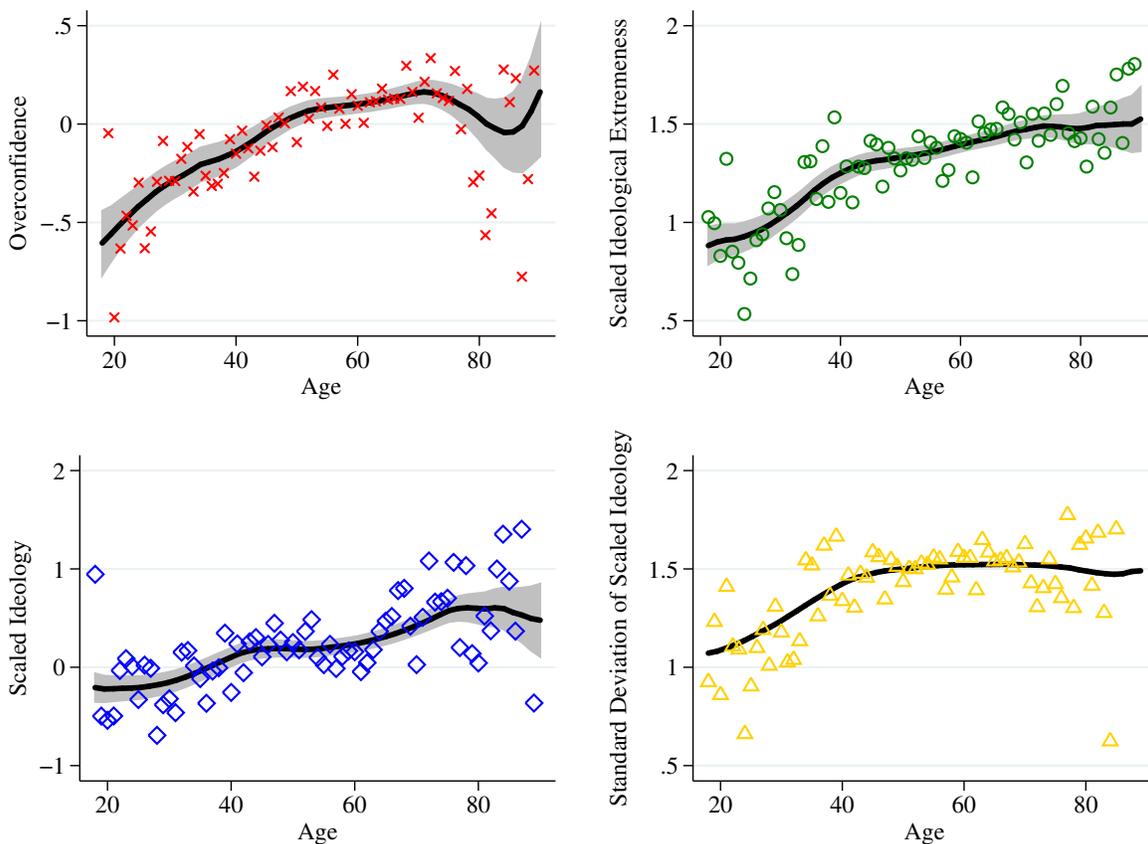
ideological extremeness will increase with age whenever the correlation between signals ρ is large, or, if the correlation is relatively small, when correlational neglect is large (ρ_i is small).

Figure 2 examines patterns in overconfidence and ideology by age. Each panel shows a smoothed, non-parametric, fit of the data, along with 95% confidence intervals, in addition to averages for each year of age. The first panel shows that, in accordance with Proposition 3, overconfidence increases with age, except for possibly among those who are older than 80, which is less than 1% of the data.[17] The second panel shows that ideological extremeness increases with age, consistent with our theory, but counter to that of Blomberg and Harring-

---

[17]As overconfidence is a residual from a regression, about half the respondents have negative overconfidence scores. This should be interpreted as a respondent being less overconfident than average.

ton (2000). The third and fourth panels show that this increase in ideological extremeness is due to both a slight rightward shift in ideology as people age, and an increase in ideological dispersion that levels out at about age 50. Interestingly, the increase in ideological extremeness is more pronounced than the rightward shift in ideology.

3.3   Shifts in Average Ideology

While Proposition 2 holds for all values of the state of the world x, more nuanced predictions are possible by considering specific values of x. In particular, the following conjecture holds in simulations.

Conjecture 4 *If $\mathbb{E}[\mathbb{E}_i[\mathcal{I}]|\kappa]$ is increasing in κ, then ideological extremeness will be more correlated with overconfidence for those with ideologies to the right of median ideology than for those to the left.*

The intuition for the conjecture is displayed in Figure 3(a). The figure is drawn using three different levels of κ to facilitate its comparison with the (smoothed) distribution of data from ideology self-reports, broken down by terciles of over-confidence, in Figure 3(b).[18] To see the intuition behind the conjecture, start at the modal ideology of citizens with middling confidence—which is roughly median ideology. Moving right from this point, towards the mode of the most over-confident tercile, average overconfidence is strictly increasing along with ideological extremeness measured from the median point. Moving to the left, ideological extremeness measured from the median point is also increasing, but average ideology decreases initially. Eventually, average ideology will increase, but this occurs beyond the region of ideology that contains most citizens. Depending on which part dominates—the part close to the mode, or the part far away—the correlation to the left will be either small and negative or small and positive.

---

[18]Scaled ideology produces a somewhat similar picture. However, as Tausanovitch and Warshaw (2011) use techniques to maximize discrimination in the tails, their estimates of ideology are strongly bimodal. While there are ways to make our theory generate bimodal distributions of ideology, we prefer to leave this for future work.
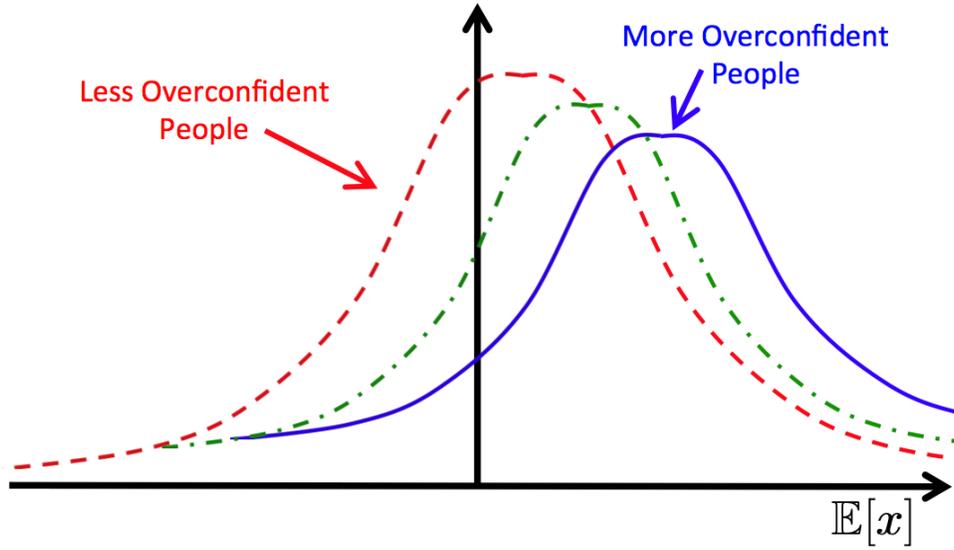
(a) Theory: When average ideology is increasing in overconfidence.
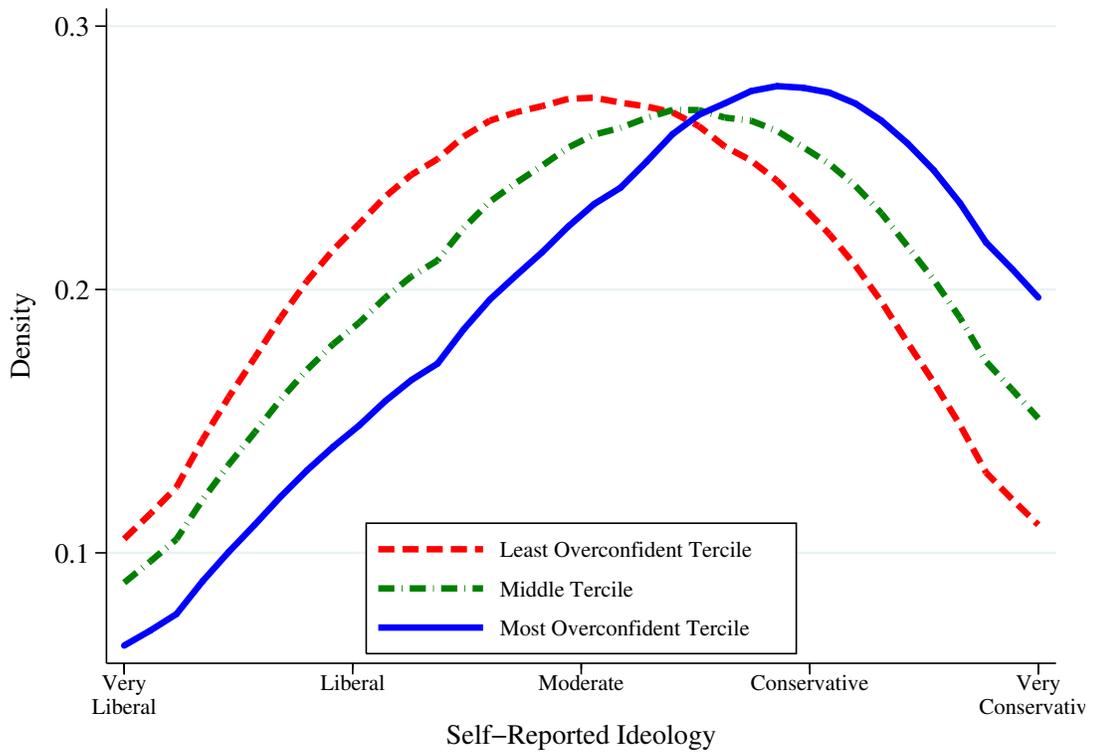
Note: Data is smoother using an Epanechnikov kernel function with bandwidth 0.8.

(b) Data: Distribution of self-reported ideology by tercile of overconfidence.

Figure 3: Theory and Data when more overconfident people are to the right ideologically.

Although Figure 3(b) provides prima facie support for the conjecture, a more rigorous analysis of the data requires that we first establish the hypothesis of the conjecture. Indeed, for all three measures of ideology, those in the middle and highest tercile of overconfidence are significantly further to the right than those in the lowest tercile. The difference between the first and second tercile (with standard error) for the scaled ideology measure is 0.33 (.067), and the difference between the first and third is 0.76 (.067). For the self-reported measure with "don't know" treated as ideologically centrist the corresponding differences are 0.23 (.051) and 0.54 (.051). When treating "don't know" as missing, the differences are 0.23 (.054) and 0.56 (.054). It is clear that for all three measures differences between the terciles are statistically significant.[19]

As the hypothesis of the conjecture is clearly met for all three measures of ideology, Table 5 tests to see whether the conclusion is confirmed by the data. As is readily apparent from the table, ideological extremeness has a substantially higher correlation with overconfidence for those to the right of center than for those to the left of center.[20] We use the Frisch-Waugh-Lovell Theorem to compute partial correlations, after including all of the other controls. Because correlations have poor sampling properties, we then confirm the results, and determine statistical significance, using OLS. For all three measures, overconfidence is statistically unrelated with ideological extremeness for those left of center.

3.4   Historical Data

The results in the previous subsection raise the concern that the correlation between overconfidence and ideological extremeness is only true for those that are right of center. While there is no way to refute this in the CCES data, more data across a greater range of con-

---

[19]Another obvious prediction from Figure 3(a) is that the variance of ideology should be increasing in overconfidence. While the current data do not support this, they do not refute it either. To understand why, note that the measures of ideology used here are ordinal, not cardinal: that is, someone whose ideology is say, -2, is to the left of someone with ideology -1, but they are not actually twice as liberal. Thus, we can find an affine transformation of the ideology measures, in particular one that reduces ideological differences in the center and increases them towards the sides, so that the data would appear to support that prediction.

[20]Ideological extremeness here is measured as the distance from median ideology in the overall population. Measuring ideological extremity from the nominal zero of the ideology scale does not qualitatively affect the results. Nor does measuring ideological extremeness from the median of the population in Tables 2 and 3.

Table 5: Overconfidence is more correlated with ideological extremeness for those right of center than those left of center.

| Measure: | Scaled | | Self-Reported Treatment of "Don't Know" | | | |
| | | | Centrist | | Missing | |
	Left of Median	Right of Median	Left of Median	Right of Median	Left of Median	Right of Median
Partial Correlation with Overconfidence	-0.031	0.12	-0.019	0.11	-0.023	0.12
			OLS Specifications			
Overconfidence	-0.027	0.13***	-0.015	0.11***	-0.038	0.11***
	(.022)	(.023)	(.022)	(.019)	(.024)	(.020)
Difference	0.15***		0.13***		0.14***	
	(.031)		(.030)		(.031)	
All Controls	Y	Y	Y	Y	Y	Y
N	1,402	1,402	1,472	1,470	1,488	1,432

Notes: ***, **, * denote statistical significance at the 1%, 5% and 10% level with standard errors in parenthesis. The coefficient magnitudes in columns 1 and 2 can be compared to Table 2, and those in columns 3–6 can be compared to Table 3. The N of the two regressions may not sum to the N in those other tables due to the fact that those respondents with the median ideology are included in both regressions.

texts would be able to provide a greater understanding of how overconfidence is related to ideology.

Unfortunately, the 2010 CCES is the only survey we are aware of that provides both good measures of political ideology and of overconfidence. Thus, we instead turn to a survey with greater coverage over but has more limited measures of ideology and overconfidence: the ANES. In particular, we follow a strategy based on the fact that many studies over time have found men to be more overconfident then women and use male as a proxy for "more overconfident".[21]

---

[21]Barber and Odean (2001) use male as an instrument for overconfidence in a study of financial risk taking. We have not adopted this strategy with the CCES data as being male is likely correlated with numerous other factors which may also affect the dependent variables we are interested in (Grinblatt and Keloharju, 2009). The curious reader may be interested to know that doing so approximately triples the effect size of overconfidence in the regressions presented in the previous, and subsequent, section. Using age as an

Figure 4: Men became significantly more conservative after 1980.
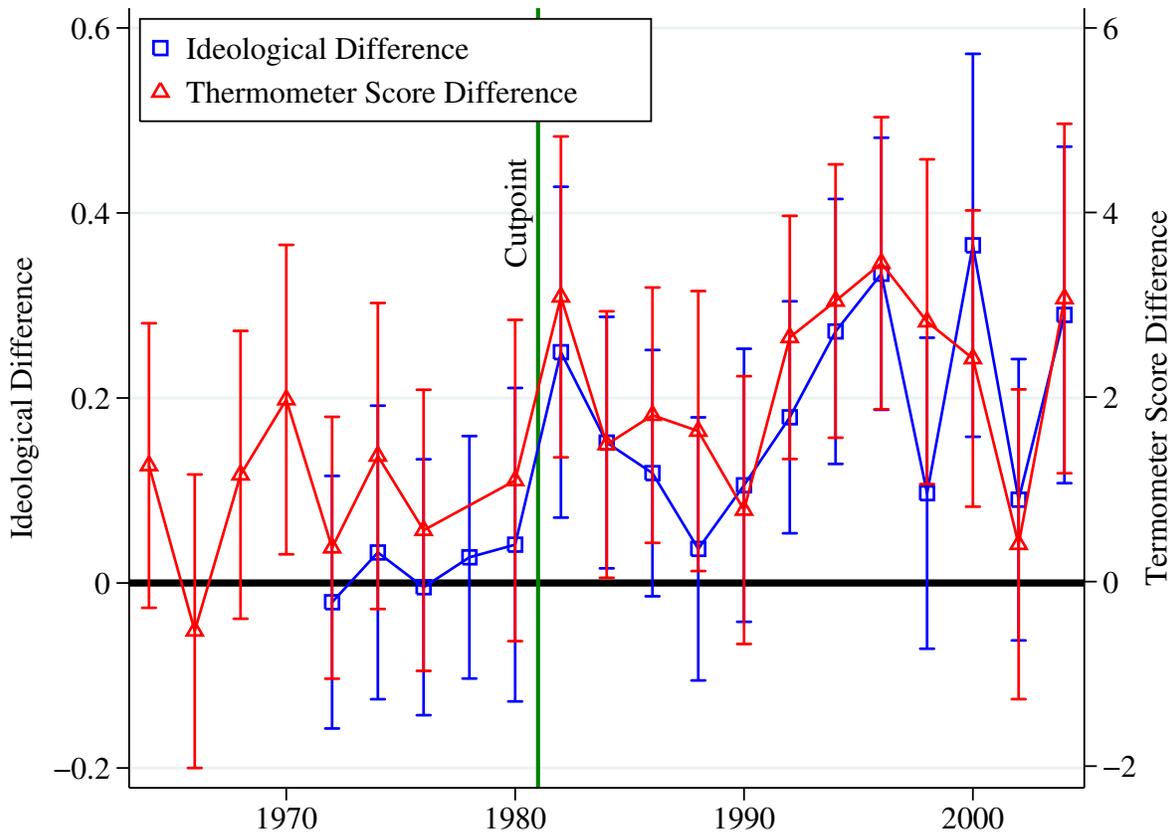
Note: Thermometer scores were not collected in 1978.

To begin the analysis we add a basic result.

Proposition 5 *If more and less overconfident citizens have the same average ideology, then overconfidence is equally correlated with ideological extremeness for both those to the right and to the left of center:* $\rho_{\mathcal{E},\kappa|\mathcal{I}\geq 0} = \rho_{\mathcal{E},\kappa|\mathcal{I}\leq 0}$

Next , we investigate if there is variation over time in the difference between the average ideology of men and women. In particular, we have both self-reported ideology and the difference between respondent's thermometer scores for "liberals" and "conservatives", which is intended as a measure of ideology. Figure 4 plots the difference between men and women on both of these scales over time with 95% confidence intervals in each year we have data.

instrument increases the effect size by a factor of roughly six.

Table 6: Data from the ANES is broadly consistent with Conjecture 4 and Proposition 5.

Time Frame	Up to 1980			1982 and After		
Dep. Variable	Ideology	Extremeness		Ideology	Extremeness	
Sample		Left of Median	Right of Median		Left of Median	Right of Median
Panel A: Self-Reported Ideology						
Male	0.013	0.14***	0.10***	0.18***	0.044**	0.16***
	(.032)	(.027)	(.025)	(.022)	(.019)	(.017)
Difference		0.035			0.11***	
		(.037)			(.025)	
Year Fixed Effects	Y	Y	Y	Y	Y	Y
N	6,880	4,241	5,132	15,183	8,808	11,395
Panel B: Thermometer Scores						
Male	0.88***	0.72***	1.62***	2.17***	-0.092	1.96***
	(.28)	(.24)	(.25)	(.23)	(.19)	(.21)
Difference		0.89**			2.05***	
		(.35)			(.28)	
Year Fixed Effects	Y	Y	Y	Y	Y	Y
N	11,439	6,551	8,709	18,105	10,455	12,992

Notes: ***, **, * denote statistical significance at the 1%, 5% and 10% level with standard errors in parenthesis. The N of the split-sample regressions do not sum to the N of the ideology regression due to the fact that those respondents with the median ideology are included in both regressions.

There is a clear rightward shift for men between 1980 and 1982. We divide the sample into two parts around 1981, and conduct a similar analysis to Table 5. The results can be found in Table 6.

The results in Table 6 are broadly consistent with the patterns predicted by Conjecture 4 and Proposition 5. For self-reported ideology, there is no statistical difference in average ideology between men and women before 1982, yet, consistent with Proposition 5, men are more ideologically extreme, regardless of their ideological direction. After 1982, men are significantly further to the right then women on average, and, consistent with Conjecture 4,

being male exhibits greater correlation with ideological extremeness for those to the right of the population median than for those to the left of the median.[22] For the thermometer scores, the difference in correlation between right and left expands as the ideological difference between men and women increases.

While the results presented in this section are broadly consistent with theory, we would not claim that they provide strong support. In particular, gender is correlated with a multitude of political differences, and the shift in ideology that occurred in the 1980s has many potential explanations that have nothing to do with overconfidence. We believe it is best to note that the available data is consistent with theory, but that better data is clearly needed.

4 Turnout and Partisan Identification

To apply our analysis to other political behaviors, such as turnout and partisan identification, we must specify how citizens make political choices. Consistent with Matsusaka (1995) and Degan and Merlo (2011), we posit that citizens choose whether to vote, and whom to vote for, in part in order to minimize regret.[23] This is motivated by the well-established phenomena that when facing a risky decision, decision-makers often feel regret if they later determine that they made the wrong choice. Moreover, decision-makers are aware of the potential for regret, and this induces them to modify their behavior in an effort to minimize future regret.[24] It is straightforward to see how this would lead to citizens avoiding a choice—whether to turn out to vote, or strongly identify with a party—unless he or she assigns a

---

[22]The magnitudes of the coefficients are similar in magnitude to the coefficient on gender in the analysis of the 2010 CCES in Sections 3.1 and 3.3. After 1988, the self-reported ideological extremeness measure exhibits no statistically significant correlation with gender for those to the left of the median, which is consistent with the analysis in Table 5.

[23]Degan and Merlo (2011) note that as it is unlikely that a citizen will discover the true state of the world, they will not anticipate regretting their decision, and instead discuss their model in terms of choice avoidance. For examples of choice avoidance in other contexts see Iyengar, Huberman and Jiang (2004), Iyengar and Lepper (2000), Boatwright and Nunes (2001), Shah and Wolford (2007), Schwartz (2004), Choi, Laibson and Madrian (2009), DellaVigna (2009), Reutskaja and Hogarth (2009), and Bertrand et al. (2010).

[24]See, Connolly and Zeelenberg (2002), Zeelenberg (1999), Zeelenberg, Inman and Pieters (2001). Models of regret have then been frequently used to explain behavioral patterns which are not compatible with standard, expected-utility, models (Bell (1982), Loomes and Sugden (1982, 1987), Sugden (1993), and Sarver (2008)).

high enough probability to that choice being later revealed as correct—precisely as we model. Indeed, this piece of our model is a direct instantiation of Sugden (1993), applied to politics.

This modeling approach allows for both non-trivial turnout and strong partisan attachment even if the policies proposed by political parties are similar to each other, as seems to be the case (Snowberg, Wolfers and Zitzewitz, 2007a,b). In contrast, this is generally not possible in models that assume citizens turn out to vote only if the difference in expected utility of the platforms proposed is high enough (Riker and Ordeshook, 1968). To make this specific, suppose that both parties propose very similar platforms, and consider an agent who is quite sure that the best policy for her is the one proposed by R. According to our definitions, this agent would strongly support, identify with, and turnout to vote for party R. However, this would not hold if these behaviors were rooted in differences in expected utility: as the two parties suggest similar platforms, for any reasonably smooth utility function there is a small difference in utility between the two parties—and hence no reason to strongly identify with one party or the other, or turnout.

4.1 Formalization

We begin with the canonical rational-choice model of voter turnout from Riker and Ordeshook (1968). Voters turn out to vote if and only if

$$pB_i - C_i + D_i > 0 \tag{5}$$

where p is the probability an individual citizen's vote is pivotal, that is, changes the winner of the election, and B_i is the benefit, to the citizen, of the citizen's favored candidate winning over the other candidate. The remaining terms C_i and D_i are the costs and benefits of voting that are unrelated to the outcome of the election.

Following the discussion above, we focus on these costs and benefits. In particular, we assume that there are idiosyncratic costs and benefits to voting, in addition to an idiosyncratic level of regret $\mathcal{R}_i$ if the citizen votes for a candidate whose platform turns out to be

worse for the citizen, given the state of the world. That is

$$D_i - C_i = D_i - \mathcal{R}_i \mathbb{I}_{\text{vote=wrong}} - C_i'$$

with D_i, R_i and C_i' i.i.d. draws from some (possibly different) distributions. We emphasize that (expected) regret can be seen as either a reduction in the benefit of voting, or an increase in the cost of voting.

Finally, we suppose there are two parties that have committed to platforms L and R, with $L = -R$.[25] Denote $U_L(b_i|x)$ and $U_R(b_i|x)$ the utility that a citizen with bias b_i receives when the state of the world is x from the platform of parties L and R respectively. Thus, a citizen will consider his vote to be "wrong" when he, say, votes for party R, but the state of the world is revealed to be such that $U_L(b_i|x) > U_R(b_i|x)$. This allows for a particularly convenient representation:

Lemma 6 *In large elections, comparative statics on voter turnout are the same as comparative statics on*

$$\left| Prob_i[U_R(b_i|x) > U_L(b_i|x)] - \frac{1}{2} \right| > c_i.\text{[26]} \tag{6}$$

That is, a citizen is more likely to turn out to vote if they believe that one party or the other is very likely to be correct, as this minimizes the chance that the voter will experience regret. For simplicity we consider c_i to be an i.i.d. draw from some distribution with strictly increasing c.d.f. $F_c \in \left(0, \frac{1}{2}\right)$, and define

$$\left| \text{Prob}_i[U_L(b_i|x) > U_R(b_i|x)] - \frac{1}{2} \right| \tag{7}$$

as the level of partisan identification of a citizen.

---

[25]This divergence, symmetric about zero, can be generated from a Calvert (1985) type model with partially policy motivated parties, and our uncertainty about the position of the median voter generated by a random realization of x in each period.

[26]In large elections, the pivot probability will not matter, and thus turnout will be driven by the idiosyncratic costs and benefits of voting, including regret. In small elections the following propositions will hold for the quadratic loss utility function. In large elections the following results hold for any single-peaked utility function.

4.2 Predictions

Our first theoretical result about turnout is empirically well documented:

Proposition 7 *More ideologically extreme citizens are more likely to turnout.*

It is interesting to note that this prediction does not depend on the specific form of the utility function, only on the fact that the utility function is single-peaked. This contrasts with the standard, pivot-probability, formulation which can produce this prediction only for very specific utility functions.

The intuition for this proposition is displayed in Figure 5(a). In this figure, we compare the probability of turning out for two citizens who have the same level of overconfidence κ, and the same preference bias $b = 0$. The probability that one or the other citizen turns out to vote depends on his or her belief about which candidate is more likely to be correct, which will depend on the citizens' posterior beliefs, which are distributed as in (2). In particular, we assume that citizen 1 has a more extreme belief than citizen 2: $\mathbb{E}_1[x] > \mathbb{E}_2[x]$, as in the figure, which implies that $e_1 > e_2$.

The positions of the parties L and R define a cutpoint, labeled in the figure at $x = 0$. If x is to the left of this cutpoint than party L is correct, whereas if x is to the right of this cutpoint than R is correct. It should be clear from the figure that citizen 1 puts a higher probability on R being right than citizen 2. Thus, citizen 1 is more likely to vote than citizen 2.

What then, is the role of overconfidence in turnout? Proposition 2 tells us that more overconfident citizens tend to be more ideologically extreme. Combining this with Proposition 7, this suggests that more overconfident citizens will be more likely to turn out. While this is indeed true in our model, the model makes an even stronger prediction: more overconfident citizens are more likely to turn out to vote even *conditional on ideology.*

Proposition 8 *Conditional on ideology, more overconfident citizens are more likely to turn out to vote. Moreover, conditional on overconfidence, more ideologically extreme citizens are more likely to turn out to vote.*
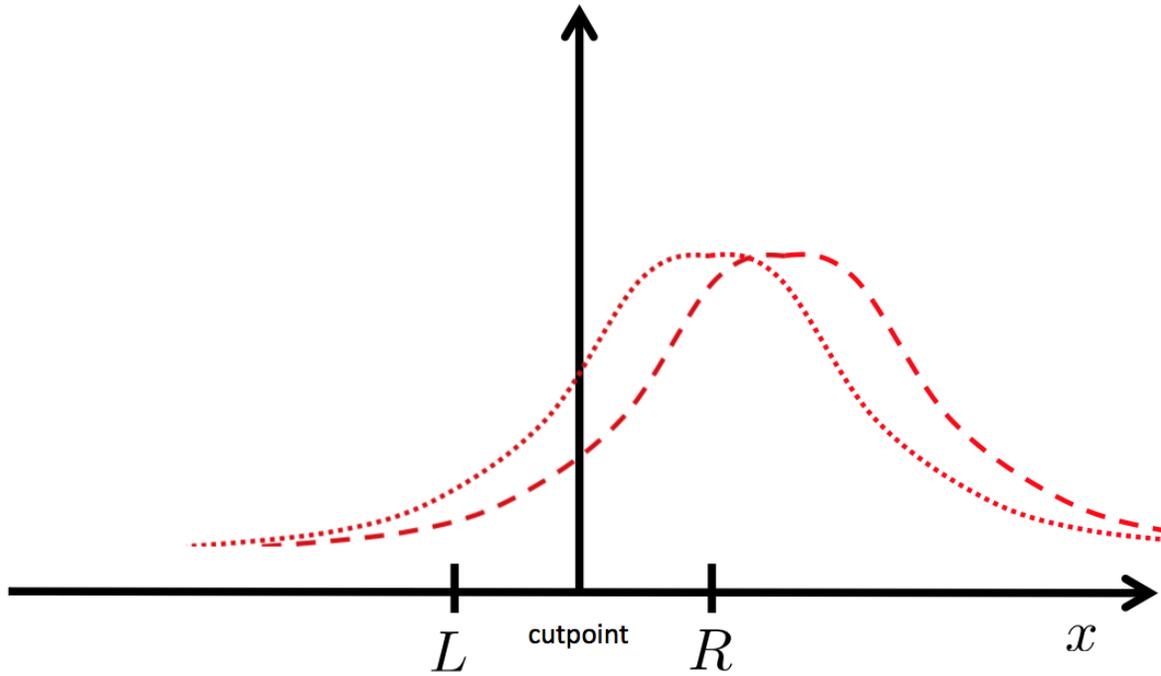
(a) More ideologically extreme citizens are more likely to turnout.
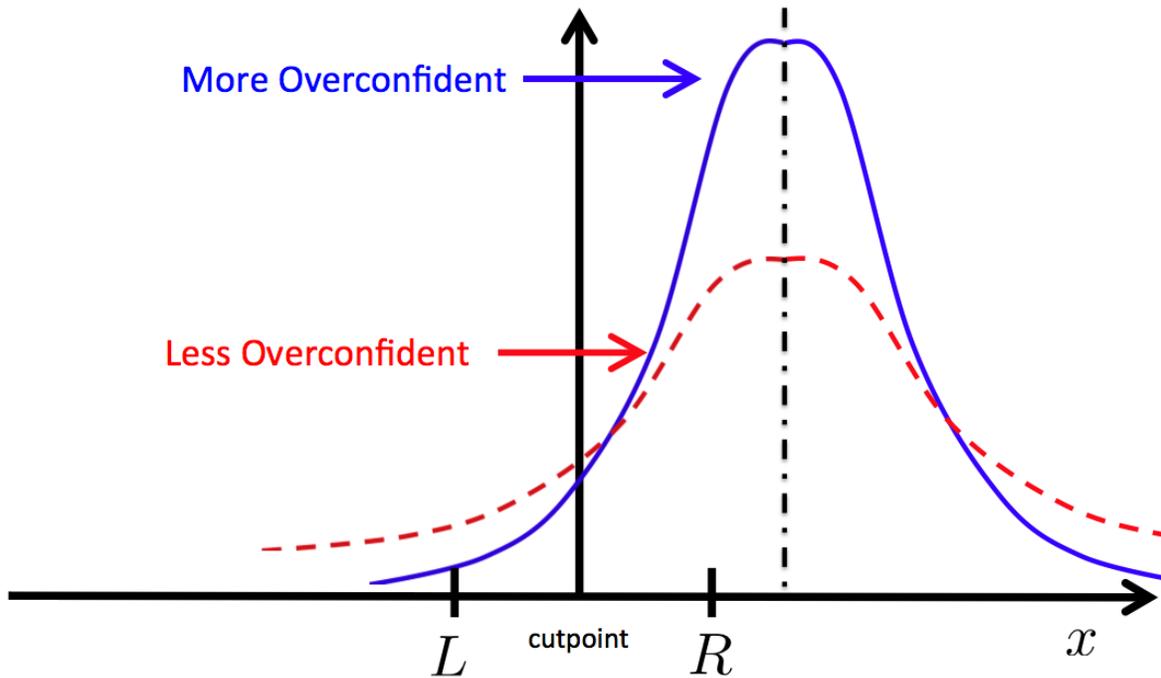
(b) More overconfident citizens are more likely to turnout, conditional on ideology

Figure 5: Intuition for Propositions 7 through 9.

The intuition behind the proposition is also shown in Figure 5. There, we consider two citizens, both with $b = 0$, and the same posterior mean belief $\mathbb{E}_i[x]$. However, citizen 1 is more overconfident than citizen 2, $\kappa_1 > \kappa_2$. Thus, according to (2), the posterior belief of citizen 1 is more precise, as shown in Figure 5(b). Applying the logic from the previous proposition, this implies that citizen 1 is more likely to think that party R is correct, and thus, more likely to turnout.

The last implication of the model we examine in survey data concerns the strength of partisan identification, as defined in (7).

Proposition 9 *Strength of partisan attachment is increasing in overconfidence, both conditional on, and independent of, ideological extremeness.*

Moreover, conditional on overconfidence, strength of partisan attachment is increasing in ideological extremeness.

The intuition is very similar to that underlying Propositions 7 and 8, and indeed one may refer to Figure 5 to understand that intuition. As in that figure, we consider two citizens with the same preference bias $b = 0$, and the same expectation after some initial experience $\mathbb{E}_i[x]$. This means, as in Figure 5, that the more overconfident citizen will place a lower probability on the state of the world being to the left of center, and thus a will express stronger support for the party R, and identify more strongly with that party.

4.3 Empirical Tests

We can test Propositions 7 and 8 using self-reported turnout from the 2010 CCES, as in Table 7. The results provide evidence in support of the propositions. The major concern with using self-reported turnout is that it is well known that survey respondents over-report their propensity to turnout. The CCES tries to correct for this by providing sample weights that try to remove the known biases in over-reporting of turnout. We therefore imple-

Table 7: Turnout is increasing in ideological extremeness and overconfidence, as predicted by Propositions 7 and 8.

	Dependent Variable: Self-Reported Turnout				
Overconfidence	0.11***	0.075***	0.071***	0.057***	0.040***
	(.017)	(.017)	(.017)	(.017)	(.013)
Ideological Extremeness		0.23***	0.23***	0.19***	0.16***
		(.019)	(.019)	(.017)	(.016)
Male			0.036		
			(.029)		
Age / 100				0.87***	
				(.010)	
All Controls					Y
N			2,759		

Notes: ***, **, * denote statistical significance at the 1%, 5% and 10% level with standard errors in parenthesis. All specifications implemented using WLS with CCES sampling weights as weights. The N of the split-sample regressions do not sum to the N of the ideology regression due to the fact that those respondents with the median ideology are included in both regressions.

ment all specifications in Table 7 using WLS and the provided sample weights.[27] The table presents some evidence consistent with the sample weights offering a reasonable solution to our concerns: gender, which is generally not associated with turnout, does not have a statistically significant effect, whereas age, which has been associated with both overconfidence and turnout, has a significant effect on the coefficients.[28]

To get a full accounting of the the effect of overconfidence on turnout, we need to account for the fact that overconfidence is also associated with ideological extremeness. After doing so, a one-standard deviation increase in overconfidence is associated with a 6–10 percentage point increase in the probability that a respondent turns out to vote (versus a baseline reported turnout rate of 69% in the data). This effect is 50–60% of the effect size associated with ideological extremeness, depending on the specification.

To examine Proposition 9 in survey data we construct three measures of partisan attach-

---

[27]The 2010 CCES plans to offer data on verified voter turnout in the future, and we look forward to incorporating this data into our analysis.

[28]The results in Tables 2 and 3 are slightly stronger when using the sample weights.

Table 8: Overconfidence is correlated with strength of partisan identification, even controlling for ideological extremeness.

Treatment of Independents	Strong (1)		Weak (0)		Missing (.)	
Overconfidence	0.034***	0.025**	0.035***	0.025**	0.039***	0.030***
	(.010)	(.011)	(.010)	(.011)	(.023)	(.012)
Ideological Extremeness		0.12***		0.18***		0.17***
		(.015)		(.014)		(.016)
All Controls	N	Y	N	Y	N	Y
N		2,816			2,513	

Notes: ***, **, * denote statistical significance at the 1%, 5% and 10% level with standard errors in parenthesis.

ment. All three code someone who identifies as a "Strong Democrat" or "Strong Republican" as a strong partisan, and most others as weak identifiers. As noted in Section 2.2.2, the three measures differ in how they treat those who identify as "Independent". Although the theory does not ascribe any particular status to such people, it is possible that they are strongly invested in their identity as an independent. Therefore, the three different measures code independents as strong partisan identifiers (1), weak partisan identifiers (0), or drops these respondents all together. Table 8 then regresses these three measures on our overconfidence and ideological extremeness measures.

The results in Table 8 are consistent with theory, no matter which measure of strength of partisan identification is used, although the magnitude of the effect size, and in some cases, level of statistical significance is lower than in the other regressions presented here. Doing the same accounting exercise as in Section 4.2, a one standard-deviation change in overconfidence is associated with a 3–4 percentage point increase in the probability a respondent classifies themselves a a strong party identifier (mean rates of strong party identification are 48%, 38% and 42%, respectively, for the three different measures). This effect is between 38–48% of the effect size of ideological extremeness.

One other pattern in Table 8 is worth noting: ideological extremeness is a better pre-

dictor of strength of partisan attachment when independents are treated as weak partisan identifiers, or left out of the data altogether. This makes intuitive sense: there are very few people who hold extremely conservative or liberal views, but identify as independent.

5 Conclusion

This paper examines some political implications of overconfidence. A novel model of over-confidence, designed for politics, is introduced. Implications for political behavior are drawn, and are tested in survey data. In particular, overconfidence is a substantively and statistically important correlate of such central political attributes as ideology, ideological extremeness, turnout, and strength of partisan attachment. Much of the future work on this paper will be incorporating more data, both from our own surveys, and from any panel data we find that might allow for measures of overconfidence and political behavior.

References

Achen, Christopher H. 1992. "Social Psychology, Demographic Variables, and Linear Regression: Breaking the Iron Triangle in Voting Research." *Political Behavior* 14(3):195–211.

Alpert, M. and H. Raiffa. 1969/1982. A Progress Report on the Training of Probability Assessors. In *Judgement Under Uncertainty: Heuristics and Biases*, ed. Daniel Kahneman, Paul Slovic and Amos Tversky. Cambridge and New York: Cambridge University Press pp. 294–305.

Ansolabehere, Stephen, Marc Meredith and Erik Snowberg. 2010. "Asking About Numbers: Why and How." California Institute of Technology, *mimeo*.

Ansolabehere, Stephen, Marc Meredith and Erik Snowberg. 2011. "Mecro-Economic Voting: Local Information and Micro-Perceptions of the Macro-Economy." California Institute of Technology, *memeo*.

Barber, Brad M. and Terrance Odean. 2001. "Boys will be boys: Gender, overconfidence, and common stock investment." *Quarterly Journal of Economics* 116(1):261–292.

Bell, David E. 1982. "Regret in Decision Making Under Uncertainty." *Operations Research* 30(5):961–981.

Benabou, Roland. 2008. "Ideology." *Journal of the European Economic Association* 6(2–3):321–352.

Benabou, Roland and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *Quarterly Journal of Economics* 117(3):871–915.

Benabou, Roland and Jean Tirole. 2006. "Belief in a Just World and Redistributive Politics." *Quarterly Journal of Economics* 121(2):699–746.

Bendor, Jonathan, Daniel Diermeier, David Siegel and Michael M. Ting. 2011. *A Behavioral Theory of Elections*. Princeton, NJ: Princeton University Press.

Bendor, Jonathan, Daniel Diermeier and Michael M. Ting. 2003. "A Behavioral Model of Turnout." *American Political Science Review* 97(2):261–280.

Benoît, Jean-Pierre and Juan Dubra. 2011. "Apparent Overconfidence." *Econometrica* 79(5):1591–1625.

Bertrand, M., D. Karlan, S. Mullainathan, E. Shafir and J. Zinman. 2010. "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment." *The Quarterly Journal of Economics* 125(1):263–306.

Bisin, Alberto, Alessandro Lizzeri and Leeat Yariv. 2011. "Government Policy with Time Inconsistent Voters." California Institute of Technology, *mimeo*.

Blomberg, S. Brock and Joseph E. Harrington. 2000. "A Theory of Rigid Extremists and Flexible Moderates with an Application to the US Congress." *The American Economic Review* 90(3):605–620.

Boatwright, Peter and Joseph C. Nunes. 2001. "Reducing Assortment: An Attribute-Based Approach." *The Journal of Marketing* 65(1):50–63.

Bullock, John G. 2009. "Partisan Bias and the Bayesian Ideal in the Study of Public Opinion." *Journal of Politics* 71(3):1109–1124.

Callander, Steven and Catherine H. Wilson. 2006. "Context-dependent Voting." *Quarterly Journal of Political Science* 1(3):227–254.

Callander, Steven and Catherine H. Wilson. 2008. "Context-dependent Voting and Political Ambiguity." *Journal of Public Economics* 92(3):565–581.

Callander, Steven and Catherine H.H. Wilson. 2007. "Turnout, Polarization, and Duverger's Law." *Journal of Politics* 69(4):1047–1056.

Calvert, Randall L. 1985. "Robustness of the Multidimensional Voting Model: Candidate Motivations, Uncertainty, and Convergence." *American Journal of Political Science* pp. 69–95.

Camerer, Colin F. and Dan Lovallo. 1999. "Overconfidence and Excess Entry: An Experimental Approach." *The American Economic Review* 89(1):306–318.

Choi, James, David Laibson and Brigitte Madrian. 2009. Reducing the Complexity Costs of 401 (k) Participation Through Quick Enrollment. In *Developments in the Economics of Aging*, ed. David A. Wise. Chicago, IL: University of Chicago Press pp. 57–82.

Connolly, Terry and Marcel Zeelenberg. 2002. "Regret in Decision Making." *Current Directions in Psychological Science* 11(6):212–216.

Daniel, Kent, David Hirshleifer and Avanidhar Subrahmanyam. 1998. "Investor Psychology and Security Market Under– and Overreactions." *The Journal of Finance* 53(6):1839–1885.

Degan, Arianna and Antonio Merlo. 2011. "A Structural Model of Turnout and Voting in Multiple Elections." *Journal of the European Economic Association* 9(2):209–245.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47(2):315–372.

DeMarzo, Peter M., Dimitri Vayanos and Jeffrey Zwiebel. 2003. "Persuasion Bias, Social Influence, and Unidimensional Opinions." *The Quarterly Journal of Economics* 118(3):909–968.

Erev, Ido, Thomas S. Wallsten and David V. Budescu. 1994. "Simultaneous Over– and Underconfidence: The Role of Error in Judgment Processes." *Psychological Review* 101(3):519–527.

Eyster, Erik and Georg Weizsäcker. 2011. "Correlation Neglect in Financial Decision-Making." London School of Economics and Political Science, *mimeo.*

Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49:379–379.

Gerber, Alan S., Gregory A. Huber, David Doherty, Conor M. Dowling and C. Panagopoulos. 2011. "Big Five Personality Traits and Responses to Persuasive Appeals: Results from Voter Turnout Experiments." Yale University, *mimeo.*

Gerber, Alan S., Gregory A. Huber, David Doherty, Conor M. Dowling and Shang E. Ha. 2010. "Personality and Political Attitudes: Relationships Across Issue Domains and Political Contexts." *American Political Science Review* 104(1):111–133.

Grinblatt, Mark and Matti Keloharju. 2009. "Sensation Seeking, Overconfidence, and Trading Activity." *The Journal of Finance* 64(2):549–578.

Hansson, Patrick, Michael Rönnlund, Peter Juslin and Lars-Gören Nilsson. 2008. "Adult Age Differences in the Realism of Confidence Judgments: Overconfidence, Format Dependence, and Cognitive Predictors." *Psychology and Aging* 23(3):531–544.

Iyengar, Sheena S., Grub Huberman and Wei Jiang. 2004. How Much Choice is Too Much? Contributions to 401 (k) Retirement Plans. In *Pension Design and Structure: New Lessons from Behavioral Finance*, ed. Olivia S. Mitchell and Stephen P. Utkus. Oxford University Press chapter 5, pp. 83–95.

Iyengar, Sheena S. and Mark L. Lepper. 2000. "When Choice Is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of Personality and Social Psychology* 79(6):995–1006.

Johnson, Dominic D.P. 2004. *Overconfidence and War: The Havoc and Glovy of Positive Illusions.* Cambridge, MA: Harvard University Press.

Juslin, Peter, Pia Wennerholm and Henrik Olsson. 1999. "Format Dependence in Subjective Probability Calibration." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25(4):1038–1052.

Kuklinski, James H., Paul J. Quirk, Jennifer Jerit, David Schwieder and Robert F. Rich. 2000. "Misinformation and the Currency of Democratic Citizenship." *Journal of Politics* 62(3):790–816.

Loomes, Graham and Robert Sugden. 1982. "Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty." *The Economic Journal* 92(368):805–824.

Loomes, Graham and Robert Sugden. 1987. "Some Implications of a More General Form of Regret Theory." *Journal of Economic Theory* 41(2):270–287.

Lundeberg, Mary A., Paul W. Fox and Judith Punćcohaŕ. 1994. "Highly Confident but Wrong: Gender Differences and Similarities in Confidence Judgments." *Journal of Educational Psychology* 86(1):114–121.

Matsusaka, John G. 1995. "Explaining voter turnout patterns: An information theory." *Public Choice* 84(1):91–117.

McMurray, Joseph C. 2011. "Ideology as Opinion: A Spatial Model of Common-value Elections." Bringam Young University.

Moore, Don A. and Paul J. Healy. 2007. "The Trouble with Overconfidence." Carnegie Mellon University, *mimeo.*

Moore, Don A. and Paul J. Healy. 2008. "The Trouble with Overconfidence." *Psychological Review* 115(2):502–517.

Odean, Terrance. 1998. "Volume, Volatility, Price, and Profit When All Traders are Above Average." *The Journal of Finance* 53(6):1887–1934.

Poole, Keith T. and Howard Rosenthal. 1985. "A Spatial Model for Legislative Roll Call Analysis." *American Journal of Political Science* 29(2):357–384.

Reutskaja, Elena and Robin M. Hogarth. 2009. "Satisfaction in Choice as a Function of the Number of Alternatives: When "Goods Satiate"." *Psychology and Marketing* 26(3):197–203.

Riker, William H. and Peter C. Ordeshook. 1968. "A Theory of the Calculus of Voting." *The American Political Science Review* 62(1):25–42.

Rothschild, David. 2011. "Expectations: PointEstimates, Probability Distributions, Confidence, and Forecasts." Yahoo! Research, *mimeo.*

Santos-Pinto, Luis and Joel Sobel. 2005. "A Model of Positive Self-Image in Subjective Assessments." *American Economic Review* pp. 1386–1402.

Sarver, Todd. 2008. "Anticipating Regret: Why Fewer Options May Be Better." *Econometrica* 76(6):263–305.

Schmidt, Klaus D. 2003. "On the Covariance of Monotone Functions of a Random Variable." Lehrstuhl für Versicherungsmathematik Technische Universität Dresden, *memeo.*

Schwartz, Barry. 2004. *The Paradox of Choice: Why More is Less.* New York, NY: Harper Collins.

Shah, Avni M. and George Wolford. 2007. "Buying Behavior as a Function of Parametric Variation of Number of Choices." *Psychological Science* 18(5):369–370.

Snowberg, Erik, Justin Wolfers and Eric Zitzewitz. 2007a. "Partisan Impacts on the Economy: Evidence from Prediction Markets and Close Elections." *The Quarterly Journal of Economics* 122(2):807–829.

Snowberg, Erik, Justin Wolfers and Eric Zitzewitz. 2007b. "Party Influence in Congress and the Economy." *Quarterly Journal of Political Science* 2(3):277–286.

Soll, Jack B. and Joshua Klayman. 2004. "Overconfidence in Interval Estimates." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30(2):299–314.

Sugden, Robert. 1993. "An Axiomatic Foundation for Regret Theory." *Journal of Economic Theory* 60(1):159–180.

Tausanovitch, Chris and Christopher Warshaw. 2011. "Testing Theories of Representation in Congress, State Legislatures and Cities Using a "Super Survey"." Stanford University, *mimeo.*

Zechman, Martin J. 1979. "Dynamic Models of the Voter's Decision Calculus: Incorporating Retrospective Considerations into Rational-Choice Models of Individual Voting Behavior." *Public Choice* 34(3):297–315.

Zeelenberg, Marcel. 1999. "Anticipated Regret, Expected Feedback and Behavioral Decision Making." *Journal of Behavioral Decision Making* 12(2):93–106.

Zeelenberg, Marcel, J. Jeffrey Inman and Rik G.M. Pieters. 2001. What we do when Decisions go Awry: Behavioral Consequences of Experienced Regret. In *Conflict and Tradeoffs in Decision Making*, ed. Elke U. Weber, Jonathan Baron and Graham Loomes. Cambridge University Press pp. 136–155.

Appendix A   Survey Measures of Overconfidence

Ansolabehere, Meredith and Snowberg allowed us to append some questions to a survey they were running on the 2011 CCES ($n =$1,000). Like the 2010 CCES, this survey included questions on unemployment and inflation, as well as respondent confidence in their answers. These questions differed slightly in that the unemployment questions asked about changes in, rather than levels of, unemployment. There seems to have been some big differences in confidence across these two surveys: In 2010 on a scale of 1–6, the average level of confidence about current inflation was 3.3, and confidence about future inflation was 3.45, whereas in 2011 the means were 2.5 and 2.45, respectively. Perhaps this was due to the fact that 2011 was not an election year in most places. Regardless, we use these questions to construct a measure of overconfidence using the same procedure detailed in this paper.

In addition, we asked four factual questions. These concerned the year the telephone was invented, the population of Spain, the year Shakespeare was born, and the percent of the US population that lives in California. These questions were all examples given in previous research on overconfidence.

We followed these questions with the same six point assessment of the respondent's confidence discussed in the text, and a question designed to elicit confidence as a confidence interval. However, rather than asking for a confidence interval directly, which we felt may have been too challenging for survey respondents, we asked them to give their estimates of the probability that the true answer was in some interval around their answer. So, for example, after giving their best guess as to the date of Shakespeare's birth, respondents were asked:

> What do you think the percent chance is that your best guess, entered above, is within 50 years of the actual answer?

Given a two-parameter distribution, such as a normal, this is enough to pin down the variance of a respondent's belief.

Mean confidence (on a six point scale) on our four questions was 2.3, 1.85, 2.0 and 2.1, respectively. On the 100 point (percent) scale, mean confidence was 71, 46, 48, and 43 respectively. Both the six point and 100 point scales of confidence were used to construct overconfidence indices, using the same method as before. Finally, all 12 confidence questions together were used to construct a combined index of overconfidence. This allows for controlling for many more dimensions of factual knowledge. Thus, we have four measures of overconfidence: economy, trivia (6 point), trivia (100 point), and combined. The economy measure, which is closest to the measure used in this paper, is correlated with the trivia (6 point) measure at 0.51, the trivia (100 point) measure at 0.28, and the combined measure at 0.61. All four measures are correlated with gender.

The 2011 CCES only contains only self-reported ideology. The four measures correlate with self-reported ideological extremeness: economy at 0.12, trivia (6 point) at 0.06, trivia (100 point) at 0.10, and combined at 0.10. Using OLS, these correlations are significant at greater than 1% in all cases but trivia (6 point), which is significant at the 5% level. The relevant comparison for 2010 is a correlation of 0.12, which is comparable to the economy measure in 2011.

While these results are not as robust due to the smaller sample size, and there are some differences between election and non-election years, it does appear that traditional ways of measuring overconfidence produce similar correlations to those observed in the body of this paper.

Appendix B   Proofs

Lemma 1 *The correlational neglect model is equivalent to the overconfidence model with* $e_i = \dfrac{1}{n}\sum_{t=1}^{n} e_{it}$ *and* $\kappa_i = \dfrac{n}{1+(n-1)\rho_i}$.

Proof of Lemma 1: The posterior likelihood in the correlational neglect model is propor-

tional to

$$\mathcal{L}(x|e_i) \propto \mathcal{L}(e_i|x)\mathcal{L}_0(x)$$

$$\propto \exp\left\{-\frac{1}{2}\begin{pmatrix} x-e_{i1} \\ x-e_{i2} \\ \vdots \\ x-e_{in} \end{pmatrix}^T \begin{pmatrix} 1 & \rho_i & \cdots & \rho_i \\ \rho_i & 1 & \cdots & \rho_i \\ \vdots & \vdots & \ddots & \vdots \\ \rho_i & \rho_i & \cdots & 1 \end{pmatrix}\begin{pmatrix} x-e_{i1} \\ x-e_{i2} \\ \vdots \\ x-e_{in} \end{pmatrix}\right\}\exp\left\{-\frac{1}{2}x^2\tau\right\}$$

$$= \exp\left\{-\frac{1}{2}\left(\frac{nx^2 - 2x\sum_{t=1}^{n}e_{it}}{1+(n-1)\rho_i}+C\right)\right\}\exp\left\{-\frac{1}{2}x^2\tau\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\frac{n+\tau(1+(n-1)\rho_i)}{1+(n-1)\rho_i}\left(x - \frac{\sum_{t=1}^{n}e_{it}}{n+\tau(1+(n-1)\rho_i)}\right)^2\right\}$$

where C is constant with respect to x. Thus, defining $e_i = \frac{1}{n}\sum_{t=1}^{n}e_{it}$, the posterior belief of a citizen is distributed according to

$$\mathcal{N}\left[\frac{ne_i}{n+\tau(1+(n-1)\rho_i)}, \frac{n+\tau(1+(n-1)\rho_i)}{1+(n-1)\rho_i}\right].$$

Substituting $\rho_i = \frac{n-\kappa}{(n-1)\kappa}$ the posterior is given by

$$\mathcal{N}\left[\frac{\kappa e_i}{\kappa+\tau}, \kappa+\tau\right],$$

as in (2). Finally, $\mathbb{E}[e_i] = x$, and

$$\text{Var}[e_i] = \left(\frac{1}{n}\right)^2\sum_{t=1}^{n}\text{Var}[\varepsilon_{it}] + 2\left(\frac{1}{n}\right)^2\frac{n(n-1)}{2}\text{Cov}[\varepsilon_{it_1}, \varepsilon_{it_2}]$$

$$= \frac{1}{n} + \frac{n-1}{n}\rho.$$

Thus, when $\rho = 1$, $e_i \sim \mathcal{N}[x, 1]$ as in the overconfidence model. If $\rho < 1$ then $e_i \sim \mathcal{N}\left[x, \frac{n}{1+(n-1)\rho}\right]$. However, as the normalization of the precision of ε_i to 1 in the overcon-

fidence model was without loss of generality, to establish equivalence when $\rho < 1$ requires simply normalizing the precision of ε_i to $\frac{n}{1+(n-1)\rho}$. Thus, the overconfidence and correlational neglect models are equivalent. ∎

Proposition 2 *Ideological extremeness and overconfidence are correlated ($\rho_{\mathcal{E},\kappa} > 0$).*

Proof of Proposition 2: $\rho_{\mathcal{E},\kappa} > 0$ when $\text{Cov}[\mathcal{E}, \kappa] > 0$. The definition of covariance and the law of iterated expectations gives

$$
\begin{aligned}
\text{Cov}[\mathcal{E}, \kappa] &= \mathbb{E}[\mathcal{E} \times \kappa] - \mathbb{E}[\mathcal{E}]\mathbb{E}[\kappa] \\
&= \mathbb{E}[\mathbb{E}[\mathcal{E} \times \kappa | \kappa]] - \mathbb{E}[\mathbb{E}[\mathcal{E}|\kappa]]\mathbb{E}[\kappa] \\
&= \mathbb{E}[\kappa \times \mathbb{E}[\mathcal{E}|\kappa]] - \mathbb{E}[\mathbb{E}[\mathcal{E}|\kappa]]\mathbb{E}[\kappa] = \text{Cov}[\mathbb{E}[\mathcal{E}|\kappa], \kappa]
\end{aligned}
$$

Note that $\mathcal{I}_i|\kappa = b_i + \mathbb{E}_i[x|\kappa]$ is a sum of two independent random normal variables. This implies that the distribution, conditional on κ, of ideology is also a normal random variable with mean given by the sum of means of the two variables, and variance given by the sum of variances of those two random variables. Thus,

$$
\mathcal{I}|\kappa \sim \mathcal{N}\left[\frac{\kappa x}{\tau + \kappa}, \frac{\tau_b(\tau + \kappa)^2}{\tau_b \kappa^2 + (\tau + \kappa)^2}\right] \tag{8}
$$

$$
\text{and,} \quad \frac{d}{d\kappa}\left(\frac{\tau_b \kappa^2 + (\tau + \kappa)^2}{\tau_b(\tau + \kappa)^2}\right) = \frac{2\kappa\tau}{(\kappa + \tau)^3} > 0 \tag{9}
$$

Note that, according to (8), $\mathcal{I}|\kappa$ is distributed according to a normal distribution, and $\mathcal{E} = |\mathcal{I}|$, so $\mathcal{E}|\kappa$ is distributed as a folded normal. When $y \sim \mathcal{N}\left[\mu, \frac{1}{\sigma^2}\right]$, then

$$
\mathbb{E}[|y|] = 2\sigma\,\phi\left[\frac{\mu}{\sigma}\right] + \mu\left(1 - 2\Phi\left[-\frac{\mu}{\sigma}\right]\right)
$$

where ϕ is the standard normal p.d.f. Assume $\mu \geq 0$, then

$$
\begin{aligned}
\frac{d}{d\mu}\mathbb{E}[|y|] &= 1 - 2\Phi\left[-\frac{\mu}{\sigma}\right] \geq 0 \\
\frac{d}{d\sigma}\mathbb{E}[|y|] &= 2\phi\left[\frac{\mu}{\sigma}\right] > 0
\end{aligned}
$$

$\mu \geq 0$ implies $x \geq 0$. Taken together with the fact, shown in (3), that the mean $\mathcal{I}|\kappa$ is weakly increasing in κ, and the variance, as shown in (9), is increasing in κ, this implies that $\mathbb{E}[\mathcal{E}|\kappa]$ is increasing in κ. A symmetric argument establishes this fact when $\mu, x < 0$. Using the result in Schmidt (2003), this implies $\mathrm{Cov}[\mathbb{E}[\mathcal{E}|\kappa], \kappa] > 0$, and thus $\rho_{\mathcal{E},\kappa} > 0$. ∎

Proposition 3 *Older citizens are more overconfident. Further, if $\rho \geq \frac{1+\rho_i\tau}{1+2\tau-\rho_i\tau}$ then ideological extremeness is increasing with age.*

Proof of Proposition 3: In the correlational neglect model, increasing the number of signals increases the precision of a citizens' posterior:

$$
\frac{n+1}{1+n\rho_i} - \frac{n}{1+(n-1)\rho_i} = \frac{1-\rho_i}{(1+(n-1)\rho_i)(1+n\rho_i)} > 0
$$

As long as $\rho_i < \rho$, then this increase in precision will be in excess of the new information transmitted, so older citizens will be more over-confident.

To establish the second part, consider a citizen who observes n signals e_{it} who believes that the correlation between those signals is ρ_i. Following Lemma 1 define $e_i = \frac{1}{n}\sum_{t=1}^{n} e_{it}$, and his mean belief is $\frac{ne_i}{n+\tau(1+(n-1)\rho_i)}$. In turn, e_i is distributed according to $e_i \sim \mathcal{N}\left[x, \frac{n}{1+(n-1)\rho}\right]$. This means that his mean belief is distributed according to a normal distribution with mean $\frac{nx}{n+\tau(1+(n-1)\rho_i)} = 0$ and variance $\frac{1+(n-1)\rho}{n} \cdot \frac{n^2}{(n+\tau(1+(n-1)\rho_i))^2}$.

Following the proof of Proposition 2 it suffices to show that the variance of the ideology of agents who receive $n+1$ signals is higher than the variance of the ideology of agents who

receive n. That is, we need to show

$$\frac{1+n\rho}{n+1} \cdot \frac{(n+1)^2}{(n+1+\tau(1+n\rho_i))^2} - \frac{1+(n-1)\rho}{n} \cdot \frac{n^2}{(n+\tau(1+(n-1)\rho_i))^2} \geq 0 \quad (10)$$

We will argue that the LHS increasing in ρ. We first show that the derivative of the LHS with respect to ρ is positive, that is

$$n(1+n)(n+(1+(n-1)\rho_i)\tau)^2 - (n-1)n(1+n+(1+n\rho_i)\tau)^2 \geq 0$$

which, in turn, is equal to

$$n + n^2 + 2n\tau + 2n^2\tau + 2n\tau^2 - 2n\rho_i\tau^2 + 2n^2\rho_i\tau^2 + n\rho_i^2\tau^2 - n^2\rho_i^2\tau^2 \geq 0.$$

Since $\rho_i \in [0,1)$, $n \geq 1$, and $\tau > 0$, then we must have $2n\tau^2 - 2n\rho_i\tau^2 \geq 0$ and $2n^2\rho_i\tau^2 - n^2\rho_i^2\tau^2 \geq 0$, which implies that that the condition is satisfied.

Since this the LHS of Equation 10 increasing in ρ, and a since we know $\rho \geq \frac{1+\rho_i\tau}{1+2\tau-\rho_i\tau}$, then it suffices to show that this condition holds when $\rho = \frac{1+\rho_i\tau}{1+2\tau-\rho_i\tau}$. Replacing ρ with this value and solving yields

$$\frac{(\rho_i-1)^2\tau^2(1+2n+(2+(2n-1)\rho_i)\tau)}{1+(2-\rho_i)\tau} \geq 0,$$

which is always true since $\rho_i \in [0,1)$, $n \geq 1$, and $\tau > 0$. ■

Proposition 5 *If more and less overconfident citizens have the same average ideology, then overconfidence is equally correlated with ideological extremeness for both those to the right and to the left of center.*

Proof of Proposition 5: Consider two citizens with $\kappa_1 > \kappa_2$. As $\mathbb{E}[\mathbb{E}_i[\mathcal{I}|\kappa]] = \frac{\kappa x}{\tau+\kappa}$, we

have that $\frac{\kappa_1 x}{\tau + \kappa_1} = \frac{\kappa_2 x}{\tau + \kappa_2} \Rightarrow x = 0$. Thus,

$$\mathcal{I}|\kappa \sim \mathcal{N}\left[0, \frac{\tau_b(\tau + \kappa)^2}{\tau_b \kappa^2 + (\tau + \kappa)^2}\right].$$

As this is symmetric about zero for all κ, it implies $\mathrm{Cov}[\mathbb{E}[\mathcal{E}|\kappa, \mathcal{I} \geq 0], \kappa] = \mathrm{Cov}[\mathbb{E}[\mathcal{E}|\kappa, \mathcal{I} \leq 0], \kappa]$ and $\mathrm{Var}[\mathcal{I}|\mathcal{I} \geq 0] = \mathrm{Var}[\mathcal{I}|\mathcal{I} \leq 0]$. Finally, as this implies $f(\kappa|\mathcal{I} \geq 0) = f(\kappa|\mathcal{I} \leq 0) = f(\kappa)$, thus, $\mathrm{Var}[\kappa|\mathcal{I} \geq 0] = \mathrm{Var}[\kappa|\mathcal{I} \leq 0]$. Taken together this implies $\rho_{\mathcal{E},\kappa|\mathcal{I} \geq 0} > \rho_{\mathcal{E},\kappa|\mathcal{I} \leq 0}$. ∎

Lemma 6 *In large elections, comparative statics on voter turnout are the same as comparative statics on*

$$\left| Prob_i[U_R(b_i|x) > U_L(b_i|x)] - \frac{1}{2} \right| > c_i.$$

Proof of Lemma 6: When elections are large $p \to 0$ in (5). Supposing citizen i favors candidate R if he or she were to vote, citizen i will vote if and only if

$$
\begin{aligned}
D_i - \mathcal{R}_i \mathbb{E}[\mathbb{I}_{\text{vote=wrong}}] - C_i' &> 0 \\
\mathrm{Prob}[\text{vote} = \text{wrong}] &< \frac{D_i - C_i'}{\mathcal{R}_i} \\
1 - \mathrm{Prob}[U_R(b_i|x) > U_L(b_i|x)] &< \frac{D_i - C_i'}{\mathcal{R}_i} \\
\mathrm{Prob}[U_R(b_i|x) > U_L(b_i|x)] - \frac{1}{2} &> \frac{1}{2} - \frac{D_i - C_i'}{\mathcal{R}_i} = c_i.
\end{aligned}
$$

The absolute value follows from symmetry of considering the case where i favors candidate L. ∎

Proposition 7 *More ideologically extreme citizens are more likely to turnout.*

Proposition 8 *Conditional on ideology, more overconfident citizens are more likely to turn out to vote. Moreover, conditional on overconfidence, more ideologically extreme citizens are more likely to turn out to vote.*

Proposition 9 *Strength of partisan attachment is increasing in overconfidence, both conditional on, and independent of, ideological extremeness.*

Moreover, conditional on overconfidence, strength of partisan attachment is increasing in ideological extremeness.

Proof of Propositions 7, 8, 9: Consider an individual i with ideology $\mathcal{I}$, overconfidence κ, and preference bias b. Suppose, without loss of generality that $\mathcal{I} > 0$. Notice first that we have $\mathbb{E}[x] = \mathcal{I} - b$. This means that we have $U_R(b|x) > U_L(b|x)$ if and only if $x > -b$. In turn, this means that we have $\text{Prob}_i[U_R(b_i|x) > U_L(b_i|x)] = \text{Prob}_i[x > -b] = 1 - \text{Prob}_i[x < -b]$. By construction this is equal to $1 - \Phi\big[(-b - (\mathcal{I} - b))\sqrt{\tau + \kappa}\big] = \Phi\big[\mathcal{I}\sqrt{\tau + \kappa}\big]$.

We now turn to show that $\mathcal{I}'' > \mathcal{I}' > 0$, then the distribution of κ conditional on $\mathcal{I}''$ first order stochastically dominates the distribution of κ conditional on $\mathcal{I}'$. Assume $b = 0$. Recall $\mathcal{I} = \frac{\kappa e}{\tau + \kappa}$, hence $\kappa = \frac{\tau \mathcal{I}}{e - \mathcal{I}}$. Now fix $\bar{\kappa} \in \mathbb{R}$ and notice that, for a given $\mathcal{I}$, the probability that κ is above $\bar{\kappa}$, $\text{Prob}(\kappa > \bar{\kappa}|\mathcal{I})$ must be equal to the probability that e is *below* $\frac{\mathcal{I}(\tau + \bar{\kappa})}{\bar{\kappa}}$, i.e. $\text{Prob}(e < \frac{\mathcal{I}(\tau + \bar{\kappa})}{\bar{\kappa}}|\mathcal{I})$. Since this threshold is strictly increasing in $\mathcal{I}$, we must have that for all $\bar{\kappa}$, the probability that κ is above $\bar{\kappa}$ conditional on $\mathcal{I}''$ is strictly higher than the one conditional on $\mathcal{I}'$, proving first order stochastic dominance.

We now show that for any κ'', κ' such that $\kappa'' > \kappa'$, the distribution of $\mathcal{I}$ conditional on κ'' and on $\mathcal{I} > 0$ first order stochastically dominates the distribution of $\mathcal{I}$ conditional on κ' and on $\mathcal{I} > 0$. Assume again that $b = 0$. Recall that $\mathcal{I} = \frac{\kappa e}{t + \kappa}$, and that we have already show that $\frac{\mathrm{d}\frac{\kappa e}{\tau + \kappa}}{\mathrm{d}\kappa} \geq 0$. Fix some κ and some $\bar{\mathcal{I}}$ in $\mathbb{R}$, and notice that $\text{Prob}(\mathcal{I} \geq \bar{\mathcal{I}}|\kappa, \mathcal{I} > 0) = \text{Prob}(e \geq \frac{\bar{\mathcal{I}}(\tau + \kappa)}{\kappa}|\kappa, e > 0)$. Note that $\frac{\mathrm{d}\frac{\bar{\mathcal{I}}(\tau + \kappa)}{\kappa}}{\mathrm{d}\kappa} \leq 0$, which means that, for any $\kappa'' > \kappa'$, we have $\text{Prob}(e \geq \frac{\bar{\mathcal{I}}(\tau + \kappa)}{\kappa}|\kappa = \kappa'', e > 0) > \text{Prob}(e \geq \frac{\bar{\mathcal{I}}(\tau + \kappa)}{\kappa}|\kappa = \kappa', e > 0)$, hence $\text{Prob}(\mathcal{I} \geq \bar{\mathcal{I}}|\kappa = \kappa'', \mathcal{I} > 0) > \text{Prob}(\mathcal{I} \geq \bar{\mathcal{I}}|\kappa = \kappa', \mathcal{I} > 0)$.

The two arguments above prove that, assuming $\mathcal{I} > 0$ and $b = 0$, $\mathcal{I}\sqrt{\tau + \kappa}$ must be strictly increasing in: κ; $\mathcal{I}$; κ conditional on $\mathcal{I}$; $\mathcal{I}$ conditional on κ. The same must therefore hold for $\Phi\big[\mathcal{I}\sqrt{\tau + \kappa}\big]$, and hence for $\text{Prob}_i[U_R(b_i|x) > U_L(b_i|x)]$ and (7). Note that specular

results would hold conditional on $\mathcal{I} < 0$. This means that we can replace $\mathcal{I}$ with $\mathcal{E} = |\mathcal{I}|$, and obtain the same results.

Note that the above holds if $b \neq 0$ as the distribution of the bias is independent from that of the other measures. Thus, the above proves Proposition 9. To extend this to Propositions 7 and 8, note that in (6) $F_c(\cdot)$ is a c.d.f. and thus increasing in its argument.

$\blacksquare$