Abstract

It has been known for some time that voters use social and political groups to infer politicians’ positions on various issues, but relatively little is known about what factors moderate the use of cues such as a candidate’s race, gender, or political party. In this paper I develop a formal information-processing model based on categorization psychology to investigate the impact of factors such as the frequency of female politicians on the organization of voters’ belief systems. The model’s central predictions, that increased balance between genders, races, or even political parties reduces voters’ reliance on these cues and that citizens rely on these cues only to the extent that they are correlated with policy positions, are borne out by data on Congressional elections from the 1978-2004 American National Election Studies.

1 Introduction

It was not entirely surprising when, during the 2008 presidential campaign, a substantial number of citizens believed that Barack Obama was a Muslim and would not support Israel (Barbaro 2008; Kantor 2008). After all, Obama is black and was raised in part in a Muslim country, and a substantial body of literature shows that cues such as these influence voters’ evaluations, memories, and perceptions of political candidates (e.g., Granberg 1985; Kahn 1994; Koch 2000, 2002; Lodge and

*I thank John Bullock, Dan Hruschka, and Matt Levendusky for many helpful comments.
Hamill 1986; Rahn 1993). Furthermore, such effects persist even among well-informed, politically sophisticated voters (Cutler 2002; Lau and Redlawsk 2001) and even if voters have direct, easy access to the candidates’ policy positions (Kahn 1994; Lodge and Hamill 1986). The basic explanation is that cues such as race, gender, and party take less time to collect and less cognitive effort to process than policy information (Cutler 2002; Downs 1957; Koch 2000; Lupia 1994; Popkin 1991; Sniderman 2000), so voters wishing economize on time and effort use them to infer candidates’ policy positions rather than focus on policy information itself. Put another way, voters are likely to use categories (or schema or stereotypes) oriented around political parties or demographic features such as race and gender in order to make their inferences (Lodge and Hamill 1986; Kahn 1994; Koch 2000; Rahn 1993). Intuitively, people who care more about an election or know more about politics or a given candidate are less likely to rely on these categories (see, e.g., Wyer and Ottati 1993, 280-284), though they still do to some extent.

It therefore seems that we know quite a bit about how and why people use simple cues such as race and gender to make inferences about candidates, but this observation makes one omission all the more surprising: we know very little about how our experiences shape our beliefs and our use of simple cues. For instance, it is intuitive that if we know more black people, we should rely less on stereotypes about blacks, yet past research has not taken this idea into account. This is a significant omission because, if we want voters to rely less on stereotypes, we are not likely to succeed by trying to get them to care more or know more about politics. Increasing and therefore more balanced exposure to women and minorities in politics is more likely and, as I show in this paper, reduces reliance on social categories that may not have meaningful policy content.

This paper draws on past research on the use of social and political groups for inference but departs from that research in several ways. First, I focus on environmental factors, such as the frequency with which voters encounter female politicians, and show that these influence the likelihood with which we rely on cues such as race and gender versus policy information. While it is intuitive that the environment influences how people think about politics (e.g., Jackman and Sniderman 2002), this idea has not been explored in a general and systematic way that encompasses many different cues, from a candidate’s race to her political affiliation. Second, I focus on endogenously-defined categories. If we want to understand who will use social groups to structure their thinking and how this depends on experiences, we can not simply assume that everyone thinks
in terms of, say, gender categories. Rather, categories arranged on gender, race, or even party lines must emerge endogenously. Thus, by investigating how the world influences our belief systems, the model also provides an explanation for where such constructs as partisan schema and gender stereotypes come from in the first place.

For that reason, this paper departs from the literature methodologically as well. I develop a mathematical model of categorization (which I describe in detail in the Appendix) derived from cognitive psychology principles to study the formation of political belief systems. From this formal and fairly general model of categorization, I derive three essential observations. First, once voters' beliefs are organized around, e.g., gender, it is essentially impossible to change this organization.

Second, the more a cue such as gender is informative — i.e., is correlated with candidates' policy positions — the more likely it is that voters will use this cue to organize their thinking. Furthermore, voters will reach different conclusions about men versus women only to the extent that gender is informative. This is in constrast to the usual assumption that voters believe women, for example, to be more liberal because they associate women with issues such as welfare and education that are in turn more associated with liberal politicians (e.g., Kahn 1994).

Third, the more infrequent women (or blacks or Catholics) are in politics, the more likely voters are to organize their thinking around gender (or race or religion). This is somewhat counterintuitive, since it means that voters are more likely to rely on low-variance cues, when we might suspect high-variance cues to have more value for distinguishing between different politicians. However, recent American National Election Studies data supports the prediction.

The results I derive in this paper have important normative implications for the debate over the value of heuristics and cognitive shortcuts in political decision making. For example, Lupia (1994) and others have argued that heuristics allow voters to make good decisions with less effort than those who are fully informed. However, voters who use heuristics may rely on stereotypes, and ignore important, relevant information. This paper suggests a return to a more intuitive approach: when a nominal cue such as race is actually informative about policy, it may be good to use it, but very often such cues are misleading and unhelpful, and regardless using such cues invariably means making generalizations while largely ignoring individual characteristics.

These are the essential concepts. The rest of this paper develops these ideas first in terms of a formal and fairly general model of categorization and later in terms of a specific model that has had
success in laboratory studies of categorization. In the next section, I give a conceptual overview of the model and develop it formally. I then state the main results using a mix of mathematical propositions and computer simulations. I show that each of the three mechanisms I described above plays an important role in making us more likely to pay attention to race, gender, and other nominal cues even when policy information is available. I then use data from the 2004 American National Election Studies to test the main comparative statics prediction, that the relative frequency in politics of men versus women, whites versus blacks, and even Democrats versus Republicans impacts the influence a candidate’s gender, race, and party has on voters’ perceptions of the candidate’s policy position. I conclude with a summary and a brief discussion of the normative implications of the model.

2 Theory

The basic idea underlying this paper is that people develop a set of categories based on their experiences and that they place candidates in these categories for the purpose of making inferences about them. In particular, they use the set of categories they have developed to make inferences about candidates’ policy positions, and as I discuss below, they are for several reasons more likely to organize their categorization schemes around simple cues such as race and gender. Categorization is a basic feature of human psychology (Estes 1994; Smith 1990) that has proved useful for studying a variety of problems in economics and political science even outside of political psychology (Fryer and Jackson 2008; Hong and Page 2001; Mullainathan, Schwartzstein, and Shleifer 2006). Thus, this paper focuses on how people categorize and what the implications are for political perception and behavior. In this section, I outline the model’s assumptions and state the main conclusions. For the sake of brevity, I save a formal statement of the model and statements of propositions for the Appendix.¹

The most important distinction in the theory is between nominal and continuous cues. A nominal cue is a piece of information such as gender, race, or party that can have only a few values and whose values inherently have some distance between them. For instance, for the most part it

¹It is worth pointing out, however, that the hypotheses I develop below were developed using the formal model. Indeed, without the aid of this model some of the conclusions may seem like a bit of a stretch. Readers interested in the technical details may find them in the Appendix.
is easy to tell men from women, so gender is a nominal cue. The name “nominal” reflects that such cues are not in themselves low-level perceptual cues (as, for example, colors or sounds would be). Rather, nominal cues may be thought of as low-level categorization decisions that happen so quickly and automatically that they function as cues. Again using gender as an example, we notice and integrate such features body shape and voice characteristics so quickly that we immedately think “man” or “woman” rather than treating these characteristics separately. A continuous cue, on the other hand, is a piece of information such as a policy position that takes many values that may be arbitrarily close to each other and hard to tell apart. For example, although it may be easy to tell John Kerry’s positions from John Boehner’s, it is harder to tell Robert Byrd’s from Norm Coleman’s.

The next important concept is that of a **category**. In this paper, a category is abstractly a collection of politicians that a person groups together and about which a person makes the same inferences, e.g., that they have the same policy positions. Although a person may be aware on a perceptual level that candidates’ policy positions may vary even if the candidates are in the same category, she uses the category to make these inferences. In this sense, a category is the basis for such concepts as partisan schema or gender stereotypes (cf. Lodge and Hamill 1986; Koch 2000). Categories are represented by their **prototypes**, which themselves represent the mix of policy positions, genders, races, etc. of the candidates in the category. Note that these prototypes may represent mixes of nominal cue values, e.g., men and women. If they could not, we would have assumed that people place men and women in different categories, which is essentially the phenomenon we are trying to understand.

The third important concept and the one that guides the process of building categories and placing politicians in them is **similarity**. In particular we are interested in how similar a given politician is to an existing category. Similarity has two key properties. First, similarity decreases as the distance between a politician and a category (actually, between the politician and the category’s prototype) increases along a dimension. So, a liberal politician is less similar to a category representing conservative politicians than a moderate politician is. Second, similarity depends on how sensitive a person is to specific cues. If a person is more sensitive to gender than to policy positions, then this person’s judgements of similarity depend more heavily on gender than policy. As a result, he will have a **gender-organized categorization scheme** and rely more heavily on candidates’
gender to make inferences about their policy positions than a person who has a policy-organized categorization scheme.

The central issue, then, is what cues — race, gender, party, policy, and so forth — voters are most likely to be sensitive to. I assume that people will tune their sensitivity to the typical within-category variance of a cue (cf. Love, Medin, and Gureckis 2004). Doing so allows them do a better job sorting future candidates into different categories. This has two consequences. The first consequence is that people tend to become more sensitive to low-variance cues. If the variance of cue stays roughly fixed, this corresponds to a sort of rescaling of the perceptual space of a cue so that typical cue values span the space. If the variance changes over time — in particular, if it increases — this feature of the model has additional consequences, which I explore below. The second consequence is that people tend to pick out one cue as more “useful” than others for making categorization decisions (cf. Gluck, Shohamy, and Myers 2002). A useful cue in this sense does not mean the cue most useful for making good categorization decisions. Rather, a cue is useful if it makes the process of differentiating between different candidates — grouping them into categories — easy, and simple cues such as race, gender, and party fit that description perfectly.

This result has important consequences for the way we think about voters’ use of cues such as race and gender. Implicit in much of the literature on this subject is that voters rely on these cues because they provide information above and beyond what policy information can provide, or at least that they do so at lower cost in time and effort. The present approach indicates that voters will pay attention to them simply because they appear to be more useful, and they appear more useful simply because they are simple and easy to grab on to. This conclusion becomes all the more unsettling when one realizes that spurious correlations or biased samples may lead voters to different conclusions about different demographic groups even in the absence of real correlations.

The foregoing discussion leads to three distinct predictions. The most basic is that people are on average more sensitive to nominal cues such as race, gender, and party than they are to continuous cues such as policy positions. Because men and women, for example, are easily differentiated, it is comparatively easy to create separate categories for male and female politicians. This idea is illustrated in Figure 1(a). Note, however, that this does not mean people will reach different conclusions about male and female politicians. Rather, they will reach independent conclusions. In order to reach different conclusions, citizens must have observed some correlation between a nominal
cue such as gender and policy positions. This does not necessarily mean that gender is actually informative of policy positions — the correlation may be spurious or based on a biased sample of politicians. However, because citizens have only their observations to rely on, such correlations will nonetheless affect their beliefs.

It is not surprising that people would rely on cues such as gender. More interesting is what is likely to modulate the extent to which people rely on such cues. First, people are more likely to rely on these cues when they are informative, e.g., when there is a real correlation between candidates’ genders and their policy positions. Essentially, the natural cognitive separation between men and women is further accentuated by an associated separation of policy positions. (See Figure 1(b).)

Second, as I mentioned above, people are likely to be more sensitive to low-variance cues. A consequence of this is that people are likely to be more sensitive to lopsided nominal cues, e.g., they will be more sensitive to gender when they are relatively fewer women than men in politics. Figure 1(c) illustrates this idea. As long as there continue to be few women in politics, this does not make much difference. However, when women become more frequent — as they have over the past several decades — the fact that voters are more sensitive to gender makes them more likely to organize their thinking around gender. Essentially, as women become more frequent, citizens perceive them as new and different and therefore make separate judgements about them.

3 Observational Implications and Empirical Observations

The last conclusion above indicates that people who have been exposed to a disproportionate number of men (or whites or Democrats) are more likely to organize their political beliefs around gender (or race or party). Specifically, the model predicts — as have other theoretical approaches — that voters should use candidates’ parties, genders, etc., to infer their policy positions, but in addition the model predicts that this effect should be (1) stronger when correlations between the cue and policy positions are stronger — and absent when these correlations are absent — and (2) stronger when voters are exposed to many more Democrats than Republicans, men than women, etc.. Let the probability of observing a Democrat be $P_d$. Simulation results (see the Appendix) indicate that this effect should manifest itself as an interaction between candidate gender and
(P_d - 0.5)^2$. I will therefore test whether nominal cue-policy correlations and lopsidedness (i.e., (P − 0.5)^2) alter the influence of nominal cues on voters’ beliefs about candidates’ policy positions.

**Data and Model.** I used data on U.S. House incumbents from the American National Election Studies Cumulative Data File (CDF) in conjunction with NOMINATE measures of the incumbents’ policy positions on a general liberal-conservative dimension (i.e., the first DW-NOMINATE dimension). The dependent variable was the respondent’s perception of an incumbent’s position on a seven-point liberal-conservative scale (C ∈ {1, 2, . . . , 7}). The primary explanatory variables of interest are the respondent’s perception of the candidate’s party’s position on this scale (P) and the candidate’s gender (G, 0 for men and 1 for women) along with measures of party-lopsidedness (P_d − 0.5)^2 and the absolute value of gender-policy correlations κ_G computed using DW-NOMINATE data and statistics from the Web site of the Clerk of the House of Representatives.² (See the discussion below for why I focus on these variables.) I measured P_d as the fraction of Democrats in the U.S. House and Senate over the the span of a person’s voting life. Party lopsidedness varies from zero to about 0.03. I also used a lifetime-averaged measure of gender-policy correlations κ_G, which varies from about −0.2 to zero. To control for the politicians’ actual policy positions, I included House DW-NOMINATE scores N_1 on the first, liberal-conservative dimension (see Poole and Rosenthal 1997, and http://voteview.com).³ To control for possible projection effects, I included the respondent’s position on a seven-point ideology scale (R) interacted with a feeling thermometer for the candidate in question (T). The full model is:

$$C = \beta_0 + \beta_N N_1 + \beta_R R + \beta_T T + \beta_{RT} RT + \beta_P P + \beta_{Pd} P (P_d - 0.5)^2 + \beta_G G + \beta_G \kappa_G.$$  \hfill (1)

In addition to the model in Equation (1), I estimate several restricted models to determine whether the addition of the correlation and lopsidedness measures improves model fit.

Because party lopsidedness and gender-policy correlations should not have any independent effects, I only include terms P(P_d - 0.5)^2 and Gκ_G; controlling for baseline lopsidedness and policy correlation does not yield statistically significant improvements in model fit. Although party-policy correlations and gender lopsidedness play a role in the model, I do not include them. The party-policy correlations are uniformly high (≥ 0.8) and as a result do not seem to add much to the model in terms of explanatory power. Likewise, the gender lopsidedness measures are

³Regressions making use of Americans for Democratic Action scores yield essentially the same results.
uniformly high (between about 0.13 and 0.23, near the maximum possible value of 0.25) and are very highly correlated with gender-policy correlations (the coefficient is 0.92, i.e., as gender lopsidedness decreases, women become more liberal), so even if lopsidedness is independently important, it will be nearly impossible to discriminate its effects from gender-policy correlations.

The CDF data has some advantages and disadvantages. The main drawback — common to all available datasets — is that it does not record respondents’ beliefs about where women as a group stand (as it does for parties, i.e., NES interviewers ask respondents to place the major parties on a seven-point ideology scale). Also, not all variables were available in all years; as a result, I used data from 1978 to 2004, with a few years in between missing because the survey did not include a required question. The major advantage of the ANES CDF data is that using data from many different elections allows one to control for election-year and age effects. If I were to focus on one election, the lopsidedness measures (for instance) would be difficult to distinguish from age effects, and of course any results would apply only to the election under consideration. Using data from many elections reduces the lopsidedness-age correlations and the results are clearly apply beyond any given election. Having reduced concern about age and election-specific effects, it is worth pointing out that lopsidedness effects are cohort effects. That is, the average lopsidedness depends exactly on when a voter was born and when the election in question takes place.

Hypotheses. The most basic hypotheses are that $\beta_N > 0$, $\beta_P > 0$, and $\beta_{RT} > 0$, i.e., perceptions of a candidate’s position should be positively correlated with perceptions of the candidate’s party’s position, with the candidate’s actual position, and with the respondent’s position if the respondent likes the candidate. (Note that increasing $C$, $N_1$, $P$, and $R$ all correspond to increasingly conservative positions.) In addition, we expect $\beta_{PD} > 0$, i.e., that the correlation between perceptions of a candidate and her party should increase the more lopsided the distribution of parties the respondent has been exposed to are.

The gender terms are somewhat more ambiguous. Previous research has found $\beta_G < 0$, i.e., respondents perceive women as more liberal than men. Once we include the interaction term with gender-policy correlation, however, $\beta_G$ should go to zero: if there are no correlations between gender and policy, then gender should have no effect. The sign of the interaction term, however, is somewhat less clear because gender lopsidedness is highly negatively correlated with a shift toward more liberal female politicians. Because lopsidedness measures are generally quite high, it is likely
that the correlation effects will dominate. In that case, because I use the absolute value of the
gender-policy correlation, the interaction term should have the same sign as the gender term, i.e.,
$\beta_{G\kappa} < 0$, since these correlations should increase the effect of gender on respondents’ beliefs. To
summarize, we expect $\beta_G = 0$ and $\beta_{G\kappa} < 0$.

[Table 1 about here.]

Results. I present the results of the OLS regressions in Table 1. First, while the improvements
in $R^2$ are small, each additional variable added to Model 1 improves the model fit. Adding the
party-policy correlation term improves fit ($F(1, 5379) = 10.53, p = 0.001$), and so does adding the
lopsidedness term ($F(1, 5379) = 2.99, p = 0.084$). Additional regressions indicate, as expected, that
adding baseline effects of lopsidedness or policy correlations does not improve model fit. Regarding
the parametric hypotheses, there are positive effects of actual candidate positions and projection
effects. Party has a positive, statistically significant effect as expected. Furthermore, there is a
positive, statistically significant effect of lopsidedness. The effective party coefficient increases from
about 0.31 to about 0.4, a roughly thirty percent increase, as party lopsidedness increases from its
minimum to its maximum.

The most remarkable effect is that of gender-policy correlations. As expected, $\beta_G$ is indistin-
guishable from zero and $\beta_{G\kappa}$ in the full model, meaning that respondents perceive female politicians
as more liberal only to the extent that they have observed female politicians to be more liberal, as
measured by their voting histories. This goes against a basic intuition in political psychology, that
voters perceive women as more liberal because they are perceived to care about issues associated
with liberal causes. The present result indicates, instead, that voters make this generalization only
to the extent that women are, on average, more liberal.

4 Conclusion

There are three main theoretical results. First, generally speaking, citizens in the model pay more
attention to the nominal cue when it is more informative. This is an intuitive result, since part of
the point of categorization is to focus on those traits that are most useful for categorizing candidates
and making inferences. Second, even when the nominal cue is minimally informative, a substantial
fraction of the simulated citizens still focus most of their attention on the nominal cue, so that they
will likely infer policy information from this cue. This effect is driven by the relative ease of telling men from women, blacks from whites, and Democrats from Republicans. The model therefore explains why people would focus on nominal cues even when they have policy information easily available to them (Lodge and Hamill 1986; Kahn 1994). Furthermore, the model explains this phenomenon without postulating the existence of structures such as a partisan schema or gender stereotype — indeed, the model offers an explanation of where such constructs come from.

Third, the relative prevalence of a particular value of a nominal cue makes a difference. In concrete terms, when there is an imbalance in the representation of, say, Democrats and Republicans, voters are more likely to pay attention to a candidate’s party. This is perhaps easier to understand in the case of gender: the fact that the vast majority of politicians are men leads citizens to be quite sensitive to a candidate’s sex and to pay attention to it when thinking about political figures and their policy positions. That is, observing a female politician is a somewhat unusual event, and because it is unusual, a politician’s sex can become significant in figuring out the politician’s beliefs. However, even relatively small imbalances have an effect.

The hypothesis I put forward in this paper are well-supported and lead to a decidedly more cognitive interpretation of gender and party effects. Data from U.S. Congressional elections show citizens’ perceptions of candidates’ policy positions depend more on the candidates’ parties when citizens have been exposed to a less balanced distribution of political parties, e.g., more Democratic representatives than Republican on average. Furthermore, gender-policy correlations have a moderating effect on citizens’ use of gender to determine candidate positions, and in the absence of these correlations citizens do not use gender at all. These results indicate that, while citizens do use stereotypes and generalize about genders and parties, they do so on the basis of their experiences.

These results have clear normative implications. In the past two decades, much research has focused on the use of cognitive heuristics to make better decisions. For example, Lupia (1994) showed that voters who knew who had endorsed several different car insurance initiatives but did not know much about the content of the initiatives made essentially the same voting decisions as those who did understand the initiatives’ content. Prior to the emergence of this literature, the use of heuristics such as these was largely viewed negatively: voters who used shortcuts were thought to be making suboptimal decisions, at least from the normative point of view.

The present paper suggests something different: to the extent that simple, nominal cues are
informative — as they presumably were in the car insurance case — it is normatively favorable to use them. If a nominal cue is not informative and, as this paper’s model predicts, people still use it to infer policy information, then it is normatively unfavorable. This is particularly so because voters are not likely to observe large, representative samples of any particular demographic group. Thus, they are likely by dint of small or biased samples to reach different conclusions about the policy positions of, say, blacks and whites despite there being no valid evidence in favor of these conclusions. The theory and evidence I present here then suggest something at best more equivocal: stereotypes and categories might occasionally be useful shortcuts, but their use should be viewed with suspicion, and all the more so when citizens does not pay much attention to politics.

Finally, the results I developed here have broader theoretical implications. One plausible underlying explanation of citizens’ reliance on cues such as race and gender is that they are members of social groups organized around race and gender, e.g., their neighborhoods are largely segregated by race. The theory and empirical observations above suggest, however, that it is experience with elite politics rather than everyday experiences that matter and therefore that cognitive factors rather than social or group effects play the more central role in structuring our belief systems. On the other hand, it may be theoretically and empirically difficult to differentiate these effects. A possible way forward is to investigate these phenomena in the context of more basic cognitive and social psychology experiments.

A  Appendix

In this section I present a mathematical statement of the theory I presented above. This section will be much more technical than the earlier discussion, but while the concepts are the same it is important to realize that the predictions I made above derive from the formal model I present here — indeed, some of them would have been difficult to derive without this model.

Definitions

I begin by defining the main concepts.

**Definition 1.** A *continuous dimension* is a continuous set of points $\mathbb{P} \subset \mathbb{R}$ with the usual metric, i.e., $\|a - b\| = |a - b| \forall a, b \in \mathbb{P}$. A *continuous cue* is a point $p \in \mathbb{P}$.
For present purposes, the most important kind of continuous cue is a *policy cue*, i.e., a candidate’s statement of his policy position.

**Definition 2.** A *nominal dimension* is a set of probability distributions $\Sigma$ with elements $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$, each of which has support on a finite set of points $S = \{s_1, s_2, \ldots, s_n\}$. The metric on $\Sigma$ is

$$
\|\sigma_a - \sigma_b\| = \frac{1}{2} \sum_{s \in S} |\sigma_{a,s} - \sigma_{b,s}|.
$$

A *nominal cue* is a point $s \in S$.

The distributions $\sigma$ represent mixes of nominal-cue values. For instance, if the dimension is gender, $S = \{s_{male}, s_{female}\}$, and a given point (i.e., distribution) on the nominal dimension represents a particular mix of men and women. Because we wish to understand why people might use different categories for male and female politicians, it is important to allow the possibility that they do not use different categories for men and women, i.e., to allow categories that are mixes of men and women. The metric on a nominal dimension reflects this fact, i.e., it is a measure of distance between distributions.

**Definition 3.** The *candidate space* is $\mathcal{C} = \prod_i \mathcal{P}_i \times \prod_j S_j$. A candidate is a point $c \in \mathcal{C}$.

In other words, a candidate is a collection of cues such as policy positions, race, gender, party, and so forth.

**Definition 4.** The *category space* (alternatively, the *feature space*) is $\mathcal{K}$. Let a voter’s set of categories be $\mathcal{K} \subset \mathcal{K}$. A *category prototype* is a point $k \in \mathcal{K}$.

I will typically use *category* to mean category prototype, i.e., I will refer to a voter’s category $k \in \mathcal{K}$. However, I will refer at times to a voter placing a candidate in category $k$. In this usage, “category” refers to the collection of candidates represented by the category prototype $k$.

In order to simplify the following discussion, I make some basic assumptions. First, I normalize $\mathcal{P} = [0, 1]$ and focus on two-valued nominal cues such as gender, i.e., $S = \{s_0, s_1\}$. Second, I focus on feature space with one policy dimension and one nominal dimension, i.e., $\mathcal{K} = \mathcal{P} \otimes \Sigma$ and $\mathcal{C} = \mathcal{P} \otimes S$. Category prototypes thus have positions $(p_k, \sigma_k) \in \mathcal{K} \subset \mathcal{K}$, where $\sigma_k$ is the probability that a candidate in $k$ takes nominal cue value $s_0$. Candidate $i$ has position $(p_i, s_i) \in \mathcal{C}$.
Assumptions

A formal model is in essence a set of mathematical assumptions about the behavior of its fundamental objects — in this case, categories and candidates. Here, I state the core assumptions concerning categorization and learning.

The purpose of categorization is to place each candidate in a category similar to that candidate. Here, I define similarity and the algorithm for categorizing candidates.

Definition 5. A similarity function is a function $H: \mathbb{R} \times C \rightarrow [0, 1]$.

Similarity functions are often called activation functions in the cognitive psychology literature; I use similarity function for clarity.

Assumption 1. $H(k, c)$ has the following properties.

1. $H(k, c)$ depends on $\|p_k - p_c\|$ and $\|\sigma_k - \sigma_c\|$, rather than on $k$ and $c$ separately, and on sensitivity parameters $\lambda_p$ and $\lambda_\sigma$ that control the weight placed on policy and nominal cues.

2. $H(k, c)$ is symmetric in the sense that if we exchanged $\|p_k - p_c\|$ with $\|\sigma_k - \sigma_c\|$ and $\lambda_c$ with $\lambda_\sigma$, $H(k, c)$ would be unchanged.

3. $H(k, c)$ is monotonically decreasing in $\|p_k - p_c\|$ and $\|\sigma_k - \sigma_c\|$ and achieves its maximum value, normalized to be one, at $k = c$.

4. Let $d_i$ be a displacement vector that takes value $d$ on dimension $i$ and zero on every other dimension. $H(k, k + d_i) > H(k, k + d_j)$ if and only if $\lambda_i < \lambda_j$.

The first two assumptions are symmetry properties that imply there is no dimension or point in the feature space that the model treats differently. The second assumption is based on a strong empirical regularity observed in similarity judgements (see Shepard 1987). The fourth assumption means that similarity is most sensitive to dimensions on which attention strength is greatest.

Next I define a categorization algorithm. The idea is that citizens each observe (possibly different) sequences of candidate statements and categorize candidates based on these statements. Roughly, we require that a citizen place each candidate she observes in the most similar category unless she has some reason not to, in which case she should create a new category for this apparently new kind of candidate. This reason might come in the form of feedback, e.g., a signal that an
inference was incorrect. In the present case, there is no such feedback, so we require an internal standard: if there is no sufficiently similar category, create a new one.

**Assumption 2.** The categorization algorithm is as follows:

1. The set of categories $\mathbb{K}$ is initially empty.

2. If $\mathbb{K}$ is empty, create a first category $k_1 = c$ after the first candidate observation $c$. Now, $\mathbb{K} = \{k_1\}$.

3. If there is some category $k$ for which $H(k, c) > \tau$, where $\tau$ is the exogenous threshold similarity, place $c$ in the category $k^*$ most similar to it, i.e., in $k^* = \arg \max_k H(k, c)$.

4. If no category satisfies $H(k, c) > \tau$, create a new category $k' = c$. Now $\mathbb{K} \rightarrow \mathbb{K} \cup \{k'\}$.

This categorization algorithm is essentially the same as that used in SUSTAIN (Love et al. 2004), a categorization model that has proved successful in explaining data from laboratory experiments.

Although I use $k^*$ here to mean the category most similar to a candidate, I will typically use it to mean the category that a person places a candidate in, i.e., a new category or the most-similar category if the similarity is larger than $\tau$.

As a person observes candidates and makes decisions about what categories to put them in, she learns two things: what position the category should have and what cues are most useful for making these decisions. We can represent these observations with two learning rules, one for category positions $k$ and one for the sensitivities $\lambda_i$.

**Assumption 3.** Let the position of a category after a voter places a candidate in it be $(k^*)'$. Then, $(k^*)' = (1 - \eta_k)k^* + \eta_k c$.

A category’s position $k$ is therefore approximately the mean position of the candidates a person has placed in it, though with more weight given to candidates more recently placed in the category.

Now consider the attention update. Let $\Delta \lambda_i = \lambda_i' - \lambda_i$, $i = p, \sigma$ be the change in attention following a candidate observation.

**Assumption 4.** $\Delta \lambda$ has the following properties.

1. $\max \Delta \lambda_i = \eta_\lambda > 0$, where $\eta_\lambda$ may in general be different from $\eta_k$. 

15
2. $\Delta \lambda_i$ achieves its maximum when the distance along dimension $i$ is zero, i.e., $(k^* - c)_i = 0$, where $k^*$ is either a new category or the most-similar category.

Note that $\Delta \lambda_i$ may be negative for some (though not all) values of $k^* - c_i$.

Persistence of Nominally-Organized Categorization

This section’s first goal is to explain why citizens would pay attention to cues such as a candidate’s gender in order to infer the candidate’s policy positions even when they have relevant policy information. In the language laid out above, this is centrally a question of why citizens organize their categorization schemes more around nominal cues and less around policy cues. I now present some results that address this issue in terms of the fundamental psychological geometry of policy and nominal cues. So the discussion is concrete I assume the nominal cue is gender, i.e., $S = \{s_m, s_f\}$.

Keep in mind that the discussion extends to other nominal cues such as race and party as well.

Again, the central question is why and to what extent citizens would organize their thinking about candidates around nominal cues. To make these ideas precise, I state the following definitions.

**Definition 6.** A categorization scheme is the set of categories $K$ and the set of attention strengths $\lambda = \{\lambda_c, \lambda_n\}$.

**Definition 7.** A categorization scheme is nominally organized if and only if all categories have the form $(k, 0)$ or $(k, 1)$ and $\lambda_n > \lambda_c$. A categorization scheme is policy organized if and only if $\lambda_\sigma < \lambda_\rho$ at least one category is of the form $(p_k, \sigma_k)$ such that $0 < \sigma_k < 1$.

A person with a nominally-organized categorization scheme thinks first and foremost in terms of a nominal cue — in this case, gender. Because a person with a gender-organized categorization scheme places more weight on gender, gender plays a larger role than the policy cue in making categorization decisions. Because all categories are either all-male or all-female — as opposed to categories that represent a mix of both — a voter with a gender-organized categorization scheme makes independent inferences for men and women.

Some additional notation will be useful. Let $\Delta_c(K, C)$ be the distance along the continuous dimension between two points in the feature space, and let $\Delta_n(K, C)$ be the distance along the nominal dimension. In some cases, I will construct points that are actually outside of the feature space as defined earlier. If this causes concern, note that we can extend the space to include such
points without necessarily allowing candidates (and hence categories) to exist outside the original space.

Lemma 1. Suppose there are two categories, $K_0 = (k_0, 0)$ and $K_1 = (k_1, 1)$, and $\lambda_n > \lambda_c$. Without loss of generality, let the next observation be $C = (c, 0)$. Then, $H(K_0, C) > H(K_1, C)$.

Proof. Let $d_c = (1, 0)$ and $d_n = (0, 1)$. Then, $\Delta_n(K_0, C) = \Delta_n(C + d_c, C) = 0$ and $\Delta_c(K_0, C) = |k_0 - c| \leq \Delta_c(C + d_c, C) = 1$ (since $k_0, c \in [0, 1]$). Therefore $H(K_0, C) \geq H(C + d, C)$, since $H$ is decreasing in the distance along any dimension — in this case, along the continuous dimension. Similarly, $H(K_1, C) \leq H(C + d_n, C)$ since $\Delta_n(K_1, C) = \Delta_n(C + d_n, C)$ and $\Delta_c(K_1, C) = |k_1 - c| \geq \delta_c(C + d_n, C) = 0$. By assumption, $\lambda_n > \lambda_c$, so from the basic similarity assumptions, $H(C + d_n, C) < H(C + d_c, C)$. Then,

$$H(K_0, C) \geq H(C + d_c, C) > H(C + d_n, C) \geq H(K_1, C) \Rightarrow H(K_0, C) > H(K_1, C),$$

which completes the proof. \(\square\)

Lemma 2. Suppose there are two categories, $K_0 = (k_0, 0)$ and $K_1 = (k_1, 1)$, and $\lambda_n > \lambda_c$. Without loss of generality, let the next observation be $C = (c, 0)$. Then, $\lambda_n + \Delta \lambda_n > \lambda_c + \Delta \lambda_c$.

Proof. From Lemma 1, a person either places $C$ in $K_0$, since it is more similar than $K_1$, or a new category. Suppose that $C$ is placed in $K_0$. Since $\|k^*_c - c_c\| = |k_1 - c| \geq 0$ and $\|k^*_n - c_n\| = 0$, it follows from the attention updating assumptions that $\Delta \lambda_n = \eta$ and $\Delta \lambda_c \leq \eta$.

Suppose instead that $C$ is placed in a new category. Then $\Delta \lambda_n = \Delta \lambda_c = \eta$. In both cases, $\lambda_n + \Delta \lambda_n > \lambda_c + \Delta \lambda_c$. \(\square\)

Proposition 1. Suppose that a person’s categorization scheme is nominally organized, and suppose that there is at least one category of the form $(k, 0)$ and at least one of the form $(k, 1)$. Then the categorization scheme is nominally organized forever.

Proof. First suppose that there are only two categories, $(k, 0)$ and $(k', 1)$. Then, it follows from Lemmas 1 and 2 and the fact that any newly created category will have the form $(c, 0)$ or $(c, 1)$ that, after the next candidate observation, the categorization scheme will be nominally organized.

It is easy to see that Lemma 1 generalizes to a larger number of categories: the similarity of a candidate to any category of the form $(k, 0)$ will always be higher than the similarity of a category
of the form \((k', 1)\) if \(\lambda_n > \lambda_c\). Thus, a person will always place \((c, 0)\) in a category of the form \((k, 0)\) or in a new category, and a similar statement holds for \((c, 1)\). Therefore, Lemma 2 still applies, so after the next candidate observation the categorization scheme will be nominally organized. The result follows by induction, i.e., each subsequent candidate will be placed in a category with matching nominal-cue value, and this preserves nominally-organized categorization.

The importance of this proposition stems from the fact that policy-organized categorization schemes do not have this property. That is, one can construct sequences of candidates such that at one time the categorization scheme is policy organized but at a later time is not.\(^4\) Thus, because of the special psychological geometry of nominal cues, the general categorization model I laid out above is biased toward nominally-organized categorization.

This result has important consequences for the way we think about voters’ use of cues such as race and gender. Implicit in much of the literature on this subject is that voters rely on these cues because they provide information above and beyond what policy information can provide, or at least that they do so at lower cost in time and effort. Proposition 1 shows that voters will pay attention to them simply because they appear to be more useful, and they appear more useful simply because they are simple and easy to grab on to. This conclusion becomes all the more unsettling when one realizes that spurious correlations or biased samples may lead voters to different conclusions about different demographic groups even in the absence of real correlations.

Remark. One can show that if the distribution of candidates on the continuous dimension (conditional on a particular value of the nominal cue) is everywhere positive, then the statement in Proposition 1 holds with probability one if \(\lambda_n = \lambda_c\). To see this, consider the proof of Lemma 2. With probability one, \(|k^*_c - c_c| > 0\), so that with probability one \(\Delta \lambda_n > \Delta \lambda_c\).

\(^4\)Here is an example. Specialize to the SUSTAIN model with \(\tau = 0.5\), \(\eta = 0.1\), and \(r = 10\); these are typical of experimentally-determined values (Love et al. 2004). Then, consider the sequence of candidates \((0, 0)\), \((0, 1)\), \((1, 0)\), \((0.5, 0)\), and \((0.6, 0)\). (Recall that the first value is the policy position and the second in the nominal cue value.) After the second candidate, the categorization scheme is policy organized, but after the fifth, \(\lambda_\sigma > \lambda_p\), so that the categorization scheme is not policy organized.
Prevalence and the Frequency of Social and Political Groups

The previous result is essentially focused on the individual and does not depend on the political environment. The next two results focus on how the political environment and in particular the distribution of social and political groups in elite politics shape voters’ belief systems. The proof of the next result requires a slightly stronger version of Assumption 1.4.5

**Assumption 5.** Suppose a voter has $\lambda_\sigma \geq \lambda_p$, and consider a category $k$ and a point $b = k + \delta$ where $\delta(0, \delta_\sigma)$. Suppose that after categorizing some candidate, the category moves to $k'$; let $b' = k' + \delta$. If $\Delta \lambda_\sigma \geq \Delta \lambda_p$, then $H(k, b) > H(k', b')$.

Assumption 5 that if the weight a voter places on a nominal cue such as gender increases relative to the weight she places on policy, then the similarity of a point at a fixed distance from the category along the nominal dimension decreases. This assumption leads to the following proposition.

**Proposition 2.** Suppose that a voter observes a sequence of $N$ male candidates prior to observing the first female candidate. Further suppose the support of the distribution of candidates is a continuous subset of the feature space and that the distribution of female candidates is uniform on the policy space. Then, if the $(N + 1)$th candidate is a man who the voter would place in an existing category, the probability the voter will always have a nominally-organized categorization scheme increases.

*Proof.* Because all candidates are men, $\lambda_\sigma = N \eta$ after the first $N$ candidates, and since the distribution of candidates has continuous support on the policy dimension, $\lambda_p < N \eta$ with probability one. Then, because the first $N$ candidates are men, each category is an all-male category. Therefore, the voter has a nominally-organized categorization scheme after $N$ candidate observations. The probability that the categorization scheme remains nominally organized forever is the probability that the voter creates a new category for the first female candidate the voter encounters. This probability is increasing in the fraction of points $c = (c_p, 0)$ satisfying $H(k, c) < \tau$ for some $k \in \mathbb{K}$.

Now, suppose the $(N - 1)$th candidate is a man. With probability one, $|k_p - c_p| > 0$, so $\Delta \lambda_p < \Delta \lambda_\sigma$ almost surely. Let $b$ and $b'$ be defined as in Assumption 5. Then, for each category, $H(k, b) > H(k', b')$. Since the number of categories stays fixed, the number of points $c$ satisfying $H(k, c) \geq \tau$ changes.
decreases. (Note that this is true whether or not a given category moves after categorization, but not necessarily true if the voter creates a new category — hence the assumption that the voter places the candidate in an existing category.) Then, the probability of creating a new category for the first female candidate increases and so too does the probability the voter has a nominally-organized categorization scheme forever.

While this proposition has a number of caveats, its meaning is quite intuitive: if it is more likely voters observe male rather than female candidates, it is generally more likely that they will organize their categories — i.e., their beliefs — around gender. (I say generally because the proposition does not extend to cases in which voters create new categories. As I will show, however, the result extends, at least in expectation, beyond cases that are amenable to formal proof.) This result is somewhat counterintuitive, since if there is no variation in gender, gender would not seem to enter into anyone’s thinking. The key concept is that when a voter finally comes across a female politician, it will be a surprising event, and gender will become directly relevant to the voter’s thinking. Furthermore, the longer a person has gone without encountering women in politics, the more surprising it will be, and the more candidate gender will matter.

Because a formal proof of the proposition requires stronger assumptions than we would prefer, I augment the mathematical approach with a computational one. I simulated voters learning about politicians using the unsupervised-learning components of the SUSTAIN model of categorization (Love et al. 2004). Each of 1000 simulated voters observed 25 politicians, and these hypothetical politicians’ positions were normally distributed with mean 0.5 (the center of the policy space) and variance either 0.1 or 0.01. The distribution was truncated so that it had zero support outside the policy space [0, 1]. Gender and policy position were uncorrelated. I varied the probability $P_m$ that a candidate was a man and computed the fraction of simulated voters with nominally-organized categorization schemes — i.e., the fraction with beliefs organized primarily around gender. I chose $P_m$ randomly from the interval [0.5, 1] (note that the results for $P_m \in [0, 0.5]$ will be the reflection of the results I report). Following (Love et al. 2004), I initialized $\lambda_p = \lambda_\sigma = 1$ and set $\tau = 0.5$, $\eta = 0.2$, and $r = 10$. The $\lambda$ initializations mean that voters do not initially make any strong distinctions between different politicians, while the choice of $\tau$ is somewhat arbitrary but strikes a balance between very fast learning that generates new categories for almost every candidate — and would therefore be cognitively very taxing — and very slow learning in which many voters would
make the same inferences about every candidate. The values of \( \eta \) and \( r \) were generated by fitting the model to a variety of laboratory experiments on categorization (see Love et al. 2004). Thus, they may or may not be the best fit to real-world situations, but they are at least a useful baseline.

Figure 2 presents the results of the simulations. As expected, the probability that voters in the model have gender-organized categorization schemes — i.e., organize their beliefs around gender — increases as the frequency of observing male politicians increases from \( P_m = 0.5 \) to 1. Thus, although the assumptions of Proposition 2 do not always hold, the conclusion very likely holds. Several other observations are worth mentioning. First, the probability of gender-organized belief systems increases with the variability of policy positions. This happens because highly variable policy positions imply high variability within categories arrayed along the policy dimension, which reduces the policy dimension’s usefulness for making categorization decisions and therefore reduces the weight a typical voter places on policy. Second, the shape of the function is quadratic: a function of the form \( a + (1 - a)(P_m - 0.5)^2 \) accounts for 98 percent of the variance. I use this fact in the regressions I report above.

### Prevalence and Informativeness

Finally, I examine the extent to which real correlations between policy positions and gender or other demographic information influence the probability that voters will organize their beliefs around gender or other such distinctions. Note that, while gender-organized beliefs imply voters make independent inferences about men and women, within the context of this model correlations between policy and gender are necessary for voters to reliably reach different inferences about men and women. Voters would not be entirely misguided in supposing there are correlations. For instance, of the 16 female U.S. Senators at the end of 2008, 11 were Democrats. On the other hand, the model does not require that there be real correlations between, say, race and policy positions. Spurious correlations or correlations based (for whatever reason) on biased samples are sufficient.

Because of the analytical complexity of this issue, I will focus on the computational approach developed above, except this time I fix \( P_m = 0.5 \) and assume different distributions of policy positions for men and women. I assume that both have normally-distributed positions with common
variance $\sigma^2$ but means that differ by $\delta x$. One can show that the correlation between their policy positions is $|\delta x/2\sqrt{\sigma^2 + \delta x^2/4}|$. I chose $\sigma^2$ uniformly from $[0,0.1]$ and $\delta x$ uniformly from $[0,1]$. Figure 3 presents the probability of gender-organized categorization as a function of this correlation. Although the results are clearly more varied, the probability of gender-organized categorization also clearly increases with the gender-policy correlation.

[Figure 3 about here.]

References


Figure 1: Factors that increase sensitivity to nominal cues such as gender. (a) Nominal cues such as gender divide more naturally than policy positions into groups. (b) If nominal cues are informative of policy positions, the division becomes even more natural. (c) If men are much more frequent than women, voters will become more sensitive to gender because it is a low-variance cue. (d) With two policy cues, there is no natural way to divide politicians into groups.
Figure 2: The probability of nominally-organized categorization as a function of the probability of observing male candidates in the SUSTAIN model.
Figure 3: The probability of nominally-organized categorization as a function of correlation between policy and gender in the SUSTAIN model.
### Effects of Party Lopsidedness and Gender-Policy Correlations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>5.168</td>
<td>5.170</td>
<td>5.169</td>
</tr>
<tr>
<td>(0.150)</td>
<td>(0.150)</td>
<td>(0.150)</td>
<td></td>
</tr>
<tr>
<td>DW-NOMINATE Dimension 1</td>
<td>0.671</td>
<td>0.687</td>
<td>0.684</td>
</tr>
<tr>
<td>(0.044)</td>
<td>(0.044)</td>
<td>(0.044)</td>
<td></td>
</tr>
<tr>
<td>Respondent Ideology</td>
<td>-0.599</td>
<td>-0.599</td>
<td>-0.599</td>
</tr>
<tr>
<td>(0.030)</td>
<td>(0.030)</td>
<td>(0.030)</td>
<td></td>
</tr>
<tr>
<td>Thermometer</td>
<td>-0.046</td>
<td>-0.046</td>
<td>-0.046</td>
</tr>
<tr>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>R Ideology × Thermometer</td>
<td>0.0115</td>
<td>0.0115</td>
<td>0.0114</td>
</tr>
<tr>
<td>(0.0005)</td>
<td>(0.0005)</td>
<td>(0.0005)</td>
<td></td>
</tr>
<tr>
<td>Party Placement</td>
<td>0.339</td>
<td>0.313</td>
<td>0.315</td>
</tr>
<tr>
<td>(0.010)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td></td>
</tr>
<tr>
<td>Party Placement × Lopsidedness</td>
<td>2.953</td>
<td>2.761</td>
<td></td>
</tr>
<tr>
<td>(0.910)</td>
<td>(0.917)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-0.222</td>
<td>-0.218</td>
<td>-0.056</td>
</tr>
<tr>
<td>(0.052)</td>
<td>(0.052)</td>
<td>(0.107)</td>
<td></td>
</tr>
<tr>
<td>Gender × Gender-Policy Corr.</td>
<td></td>
<td></td>
<td>-2.444</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.414)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.454</td>
<td>0.455</td>
<td>0.455</td>
</tr>
<tr>
<td>$F$</td>
<td>10.53</td>
<td>2.99</td>
<td></td>
</tr>
<tr>
<td>$p = 0.001$</td>
<td>$p = 0.084$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Unstandardized coefficients for both models; standard errors are in parentheses; substantively important results are highlighted in boldface type. $F$-tests compare each model with the next-smallest model. $N = 5386$. 

28