# Stereotypes

## Supplementary Material For Online Publication

P. Bordalo, K. Coffman, N. Gennaioli, A. Shleifer

## A Proofs

**Proposition 1.** By Definition 1, the representativeness ranking of types for $G$ is the opposite of that for $-G$. Thus, if $t^{max,G} = argmax_t \frac{\pi_{t,G}}{\pi_{t,-G}}$ is the most representative type for $G$, then it is also the least representative type for $-G$.

Suppose now that $\pi_{t,G} > \pi_{t',G}$ if and only if $\pi_{t,-G} > \pi_{t',-G}$ (case i)). Then, both groups share the same modal type $t_{mod}$. Because $\pi_G \neq \pi_{-G}$, it follows that not all types are equally representative. Because the representativeness ranking is opposite for the two groups, $t_{mod}$ can coincide with the most representative type for at most one of the groups.

Consider now the case where $\pi_{t,G} > \pi_{t',G}$ if and only if $\pi_{t,-G} < \pi_{t',-G}$ (case ii)). Then, it also follows that $\pi_{t,G} > \pi_{t',G}$ if and only if $\pi_{t,G}/\pi_{t,-G} > \pi_{t',G}/\pi_{t',-G}$ so that likelihood and representativeness rankings coincide for each group. In particular, the most representative type coincides with the modal type for each group. ■

**Proposition 2.** Index the types $t \in \{1, \ldots, T\}$ according to the underlying cardinal relation. Suppose first the likelihood ratio $\pi_{t,G}/\pi_{t,-G}$ is monotonically decreasing in $t$. Then it follows that $\pi_{t,-G}$ first order stochastically dominates $\pi_{t,G}$, so that $\mathbb{E}(t|G)$ is lower than $\mathbb{E}(t|-G)$, and therefore lower than the unconditional mean, $\mathbb{E}(t|G) < \mathbb{E}(t)$ (recall that $\mathbb{E}(t) = \mathbb{E}(t|\Omega)$ where $\Omega = G \cup -G$). Moreover, the ordering of types by representativeness coincides with the cardinal ordering of types, so that, for any function $h_t$, we have

$$h_t(\pi_G/\pi_{-G}) \leq h_{t'}(\pi_G/\pi_{-G}) \text{ iff } t > t'$$

where the first inequality is strict for at least some types. Consider now the likelihood ratio

between the stereotypical distribution $\pi_G^{st}$ and the undistorted distribution $\pi_G$:

$$\frac{\pi_{t,G}^{st}}{\pi_{t,G}} = \frac{h_t(\pi_G/\pi_{-G})}{\sum_{s\in T}\pi_{s,G}\cdot h_s(\pi_G/\pi_{-G})}$$

This likelihood ratio is (weakly) monotonically decreasing in $t$, implying that $\pi_G$ f.o.s.d. $\pi_G^{st}$, and in particular that $\mathbb{E}^{st}(t|G) < \mathbb{E}(t|G)$.

If the the likelihood ratio is monotonically increasing in $t$, the same logic yields $\mathbb{E}(t|G) > \mathbb{E}(t)$. Moreover, the ordering of types by representativeness coincides with the inverse of the cardinal ordering of types, so that now $\pi_G^{st}$ f.o.s.d. $\pi_G$. It follows that $\mathbb{E}^{st}(t|G) > \mathbb{E}(t|G)$. ■

**Proposition 3.** Let the set of types be $T = \{0, \ldots, T\}$ and consider for concreteness the case where $T$ is even (the same proof goes through for $T$ odd). Note first that the assumption that $\pi_{t,G}$ and $\pi_{t,-G}$ are symmetric around the midpoint $t = \frac{T}{2}$, namely that $\pi_{t,G} = \pi_{T-t,G}$, implies that representativeness $\pi_{t,G}/\pi_{t,-G}$ is also symmetric. This property ensures that the means $\mathbb{E}^{st}(t|G)$ and $\mathbb{E}^{st}(t|-G)$ are correctly estimated at $T/2$. Writing the weighting function as $f_t = h_t(\pi_G/\pi_{-G})/\sum_s \pi_{s,G}h_s(\pi_G/\pi_{-G})$ we find

$$\mathbb{E}^{st}(t|G) - \mathbb{E}(t|G) = \sum_t t\cdot\pi_{t,G}\,(f_t - 1) = \sum_{T-t}(T-t)\cdot\pi_{T-t,G}\,(f_{T-t} - 1) = \frac{T}{2}\sum_t \pi_{t,G}\,(f_t - 1) = 0$$

because $\mathbb{E}(f_t|G) = 1$. In contrast, variances are systematically distorted. To see this, write:

$$Var^{st}(t|G) - Var(t|G) = \sum_t \left(t - \frac{T}{2}\right)^2 \cdot \pi_{t,G}\,(f_t - 1) = cov\left(\left(t - \frac{T}{2}\right)^2, f_t - 1\right)$$

If $Var(t|G) > Var(t|-G)$, so that representativeness is U-shaped, then $cov\left(\left(t - \frac{T}{2}\right)^2, f_t - 1\right) > 0$ and variance gets exaggerated, $Var^{st}(t|G) > Var(t|G)$. Conversely, if $Var(t|G) < Var(t|-G)$ this implies $Var^{st}(t|G) < Var(t|G)$. ■

**Proposition 4.** 1) From Definition 2, we have

$$\mathbb{E}^{st}(t|G) = \mathbb{E}(t|G) - \sum_t t\cdot\pi_{t,G}\cdot(1 - f_t)$$

where $f_t = h_t(\pi_G/\pi_{-G})/\sum_s \pi_{s,G} h_s(\pi_G/\pi_{-G})$ is a differentiable weighing function. The factor $(1-f_t)$ is increasing in the representativeness of type $t$ for group $-G$, and equals zero when $\pi_{t,-G}/\pi_{t,G} = 1$. We will expand $(1-f_t)$ to first order around $\pi_{t,-G}/\pi_{t,G} = 1$.

To do so, denote by $r_s = \frac{\pi_{s,G}}{\pi_{s,-G}}$ and by $x_s = 1/r_s$. The first order expansion around $\mathbf{x} = \mathbf{1}$ reads,

$$f_t(\mathbf{x}) \sim 1 + \sum_s \partial_{x_s} f_t|_{\mathbf{x}=\mathbf{1}} (x_s - 1)$$

We now characterise $\partial_{x_s} f_t$ around $\mathbf{x} = \mathbf{1}$. Denote $h_t(\mathbf{1}) = h$, and $\partial_{r_t} h_t|_{\mathbf{x}=\mathbf{1}} = \theta_1$ and $\partial_{r_s} h_t|_{\mathbf{x}=\mathbf{1}} = \theta_2$ for $s \neq t$, where by assumption $\theta_2 < 0 < \theta_1$. We have:

$$\begin{aligned}
\partial_{x_t} f_t|_{\mathbf{x}=\mathbf{1}} &= -\left.\frac{\partial_{r_t} h_t \cdot (\sum_s \pi_{s,G} h_s) - h_t (\sum_s \pi_{s,G} \partial_{r_t} h_s)}{(\sum_s \pi_{s,G} h_s)^2}\right|_{\mathbf{x}=\mathbf{1}} \\
&= -\frac{\theta_1 h - h(\theta_1 \pi_{t,G} + \theta_2(1-\pi_{t,G}))}{h^2} \\
&= -\frac{(\theta_1 - \theta_2)(1-\pi_{t,G})}{h} < 0
\end{aligned}$$

and

$$\partial_{x_s} f_t|_{\mathbf{x}=\mathbf{1}} = -\frac{(\theta_2 - \theta_1)\pi_{s,G}}{h} > 0$$

Inserting back into the first order expansion, we find

$$f_t(\mathbf{x}) \sim 1 - \frac{(\theta_1 - \theta_2)}{h} \cdot \left[(x_t - 1) - \sum_s \pi_{s,G}(x_s - 1)\right]$$

Note that, by construction, the average departure $x_s - 1$ is zero, namely $\sum_s \pi_{s,G}(x_s - 1) = \sum_s (\pi_{s,-G} - \pi_{s,-G}) = 0$. So we find $f_t(\mathbf{x}) \sim 1 - \frac{(\theta_1-\theta_2)}{h}(x_t - 1)$. Plugging this approximation back into the expression for $\mathbb{E}^{st}(t|G)$ we find

$$\begin{aligned}
\mathbb{E}^{st}(t|G) &\approx \mathbb{E}(t|G) - \sum_t t \cdot \pi_{t,G} \cdot \frac{(\theta_1 - \theta_2)}{h}(x_t - 1) \\
&= \mathbb{E}(t|G)\left(1 + \frac{(\theta_1 - \theta_2)}{h}\right) - \mathbb{E}(t|-G)\left(\frac{\theta_1 - \theta_2}{h}\right)
\end{aligned}$$

which yields the result, with $\kappa = \frac{\theta_1 - \theta_2}{h} > 0$.

2) Distinguish types into the right tail, $H = \{T-2, \ldots, T\}$, and the left tail, $L = \{1, \ldots, T-3\}$. For $X = L, H$, write $\pi_{X,G} = \sum_{t \in X} \pi_{t,G}$, $\bar{t}_X = \frac{\sum_{t \in X} t \cdot \pi_{t,G}}{\pi_{X,G}}$, and $f_X = \frac{\sum_{t \in X} \pi_{t,G} f_t}{\pi_{X,G}}$, so that $\pi_{L,G} f_L + \pi_{H,G} f_H = 1$ (we omit the group index from $f_t$ for simplicity). We then have

$$\mathbb{E}^{st}(t|G) = \mathbb{E}(t|G) + \sum_{L} t \cdot \pi_{t,G} \cdot (f_t - 1) + \sum_{H} t \cdot \pi_{t,G} \cdot (f_t - 1)$$

Under the approximation that $f_t$ varies little within each tail, $f_t \approx f_X$ for $t \in X$, this becomes

$$\mathbb{E}^{st}(t|G) = \mathbb{E}(t|G) + (f_L - 1)\, \pi_L \bar{t}_L + (f_H - 1)\, \pi_H \bar{t}_H = \mathbb{E}(t|G) + (f_H - 1)\, \pi_H \left(\bar{t}_H - \bar{t}_L\right)$$

because $(f_L - 1)\, \pi_L = -(f_H - 1)\, \pi_H$.

Adapting the notation from point 1), we set $r_H = \frac{\sum_H \pi_{t,G}}{\sum_H \pi_{t,-G}}$ and $r_L = \frac{\sum_L \pi_{t,G}}{\sum_L \pi_{t,-G}}$. Expanding $f_H$ around $r_H = 1$ we find

$$\begin{aligned} f_H(\mathbf{r}) &\approx 1 + \partial_{r_H} f_H|_{\mathbf{r}=\mathbf{1}} (r_H - 1) + \partial_{r_L} f_H|_{\mathbf{r}=\mathbf{1}} (r_L - 1) \\ &= 1 + (r_H - 1) \left[ \partial_{r_H} f_H|_{\mathbf{r}=\mathbf{1}} - \partial_{r_L} f_H|_{\mathbf{r}=\mathbf{1}} \frac{\pi_{H,-G}}{\pi_{L,-G}} \right] \end{aligned}$$

Recall that by assumption $\partial_{r_H} f_H|_{\mathbf{r}=\mathbf{1}} > 0 > \partial_{r_L} f_H|_{\mathbf{r}=\mathbf{1}}$. Inserting this back into the approximate expression for $\mathbb{E}^{st}(t|G)$ we get

$$\mathbb{E}^{st}(t|G) \approx \mathbb{E}(t|G) + \lambda_G (r_H - 1)$$

where $\lambda_G = \pi_H \left(\bar{t}_H - \bar{t}_L\right) \left[ \partial_{r_H} f_H|_{\mathbf{r}=\mathbf{1}} - \partial_{r_L} f_H|_{\mathbf{r}=\mathbf{1}} \frac{\pi_{H,-G}}{\pi_{L,-G}} \right]$ is positive. Thus, the believed mean $\mathbb{E}^{st}(t|G)$ increases in the representativeness $r_H$ of the right tail for $G$, and is exaggerated when the right tail is more representative of $G$ than the left tail. Replacing $r_H$ with $R_H^{cons}$ yields equation (6). An analogous calculation, where we expand in $1/R_H^{cons}$, yields equation (7). ■

# B Unordered Types

In many settings, decision makers must assess groups in terms of their distributions over unordered type spaces. For instance, one may be interested in the distribution of occupations, or of political views, or of beliefs of different social groups. Our model applies directly to these settings, provided the type space is specified, or at least implied, by the problem at hand. While there is no notion of "extreme" types in unordered type spaces, the central insight about how representativeness and likelihood combine to determine stereotype accuracy continues to hold: when groups are very similar, representative differences tend to be relatively unlikely, while when groups are different representative differences tend to be likely, and thus generate more accurate stereotypes.

To illustrate this logic in the context of unordered types, consider the formation of the stereotypes "Republicans are creationists" and "Democrats believe in Evolution". In May 2012, Gallup conducted a public opinion poll assessing the beliefs about Evolution of members of the two main parties in the US. The results on the beliefs of Republicans and Democrats, largely unchanged in the three decades over which such polls have been conducted, are presented below:[34]

| | *Creationism* | *Evolution* | *Evolution guided by God* |
|---|---|---|---|
| *Republicans* | 58% | 5% | 31% |
| *Democrats* | 41% | 19% | 32% |

The table shows that being a creationist is the distinguishing feature of the Republicans, not only because most Republicans are creationist but also because more Republicans are creationists than Democrats. In this sense, stereotyping a Republican as a creationist yields a fairly accurate assessment. Formally, $t = Creationism$ maximizes not only $\Pr(Republicans|t)/\Pr(Democrats|t)$ but also $\Pr(t|Republicans)$.

On the other hand, the distinguishing feature of the Democrats is to believe in the "standard" Darwinian Evolution of humans, a belief four times more prevalent than it

[34]The three options were described as "God created Humans in present form in the last 10,000 years", "Humans evolved, God has no part in process" and "Humans evolved, God guided the process". See http://www.gallup.com/poll/155003/Hold-Creationist-View-Human-Origins.aspx for details.

is among Republicans. However, and perhaps surprisingly, only 19% of Democrats believe in Evolution. Most of them believe either in creationism (41%) or in Evolution guided by God (32%), just like Republicans do. Formally, $t = \textit{Evolution}$ maximizes $\Pr(\textit{Democrats}|t)/\Pr(\textit{Republicans}|t)$ but not $\Pr(t|\textit{Democrats})$. Evolution is not the most likely belief of Democrats, but rather the belief that occurs with the highest relative frequency. A stereotype-based prediction that a Democrat would believe in the standard evolutionary account of human origins, and would not believe in Creationism, is highly inaccurate.

Another example in this spirit is as follows. Suppose the DM must assess the time usage of Americans and Europeans. For the sake of simplicity, we consider only two types, namely $T = \{\text{time spent on work, time spent on vacation}\}$. The Americans work 49 weeks per year, so the conditional distribution of work versus vacation time is $\{0.94, 0.06\}$. In contrast, the Europeans work 47 weeks per year, with work habits $\{0.9, 0.1\}$. In both cases, work is by far the most likely activity. However, because the Americans' work habits are more concentrated around their modal activity, the stereotypical American activity is work. Because Europeans have fatter vacation tails, their stereotypical activity is enjoying the dolce vita. This stereotype is inaccurate, precisely because the vast majority of time spent by Europeans is at work. Still, due to its higher representativeness, vacationing is the distinctive mark of Europeans, which renders the image of holidays highly available when thinking of that group.

# C Multidimensional Types

In the real world, the types describing a group are multidimensional. Members of social groups vary in their occupation, education and income. Firms differ in their sector, location and management style. While in some cases only one dimension is relevant for the judgment at hand, in other cases multiple dimensions need to be considered. In these judgments, forming an appropriate model requires DM's to properly weigh the different dimensions. Representativeness has significant implications for this process. In particular, in many cases, the "kernel of truth" logic carries through to the case of multiple dimensions. Stereotypes are formed along the dimensions in which the groups differ most, although the DM focuses on proportional differences rather than absolute differences. As in the unidimensional case, stereotypes are context dependent in the sense that the dimensions along which a group is stereotyped depends on the other group it is compared to.

We focus on the special case in which there are two dimensions. A type consists of a vector $(t_1, t_2)$ of two dimensions, where $t_i \in T_i$ for $i = 1, 2$. Denote by $\pi_{(t_1,t_2),G}$ and $\pi_{(t_1,t_2),-G}$ the joint probability densities in groups $G$ and $-G$, respectively, which are defined over the set of types $T = T_1 \times T_2$. The representativeness of $(t_1, t_2)$ for group $G$ is given by:

$$R_G(t_1, t_2) \equiv \frac{\pi_{(t_1,t_2),G}}{\pi_{(t_1,t_2),-G}} = \frac{\pi_{t_1,G}}{\pi_{t_1,-G}} \cdot \frac{\pi_{t_2,(G,t_1)}}{\pi_{t_2,(-G,t_1)}}. \tag{8}$$

where $\pi_{t_2,(G,t_1)} = \Pr(t_2|G, t_1)$. In light of Equation (8), then, we can immediately observe:

**Lemma 1** *Suppose that $d < |T_1| \times |T_2|$ and that $\pi_{t_1,G} \neq \pi_{t_1,-G}$ for some $t_1 \in T_1$.*

*i) If $\pi_{t_2,(G,t_1)} = \pi_{t_2,(-G,t_1)}$ for all $t_1$ and $t_2$, then the stereotype for group $G$ selects a subset of values for $t_1$ while allowing for all possible values of $t_2$.*

*ii) If instead $\pi_{t_2,(G,t_1)} \neq \pi_{t_2,(-G,t_1)}$ for some $t_1$ and $t_2$, then the stereotype for group $G$ selects a subset of the most representative values of $t_1$ and $t_2$.*

**Proof.** If $\pi_{t_2,(G,t_1)} = \pi_{t_2,(-G,t_1)}$ for all $t_1$ and $t_2$, as in case i), it follows from Equation 8 that $R_G(t_1, t_2) = R_G(t_1)$ (and similarly, $R_{-G}(t_1, t_2) = R_{-G}(t_1)$) for all $t_1, t_2$. However, because $\pi_{t_1,G} \neq \pi_{t_1,-G}$ for some $t_1$, it must be that $R_G(t_1) > R_G(t_1')$ for some $t_1, t_1'$. As a consequence, for $d$ sufficiently small, the stereotype of $G$ consists of a truncation $T_1^{st} \times T_2$,

where $T_1^{st}$ includes only the types $t_1$ that have sufficiently high $R_G(t_1)$. The type space $T_2$ is not truncated because ties are included in the stereotype.

If instead $\pi_{t_2,(G,t_1)} \neq \pi_{t_2,(-G,t_1)}$ for some $t_1$ and $t_2$, Equation 8 implies a strict representativeness ranking in at least a subset of types in $\{t_1\} \times T_2$. Thus, there exists $d < N$ such that some type in $\{t_1\} \times T_2$ is truncated and others are not. Similarly, because $\pi_{t_1,G} \neq \pi_{t_1,-G}$ for some $t_1$, for given $d$ some types in $T_1$ are truncated. Together, these observations imply that the stereotype for $G$ generically implies truncations along both dimensions. ■

This result shows how the kernel of truth logic extends to multiple dimensions. When groups only differ along one dimension, namely when the distribution of $t_2$ is identical across groups conditional on $t_1$ (case i), the stereotype is formed along that dimension, in the sense that it highlights group differences in $t_1$ only. Suppose for instance that $t_1$ indexes education while $t_2$ indexes welfare status. If all groups are equally likely to be on welfare conditional on education, stereotypes exaggerate educational differences but the welfare status is correctly representated (conditional on education types that come to mind).[35]

When instead groups differ along both dimensions (case ii), stereotypes highlight differences along both dimensions. In the context of the previous example, if the less educated group is *also* conditionally more likely to be on welfare, then it is stereotyped as "uneducated and on welfare", while the other group is stereotyped as "educated and not on welfare". Again, there is a kernel of truth in these stereotypes, but also an exaggeration of the correlation between education and being on welfare: people neglect that most elements of the less educated group are not on welfare, as well as the fact that a non-trivial share of the more educated, and possibly larger, group are in fact on welfare.

Multidimensional stereotypes also raise new aspects of context dependence. Consider the stereotype of the red-haired Irish. This stereotype arises from comparing the Irish to a population (e.g., Europeans) with a much lower share of red haired people. Our model predicts that this stereotype should change when the Irish are compared to a group with a similar share of red-haired people, such as the Scots. When compared to the Scots, a more

[35]Here the stereotype allows for all possible values of $t_2$ because of the tie breaking assumption in Definition 2. The result that in case *i*) stereotypes are not organized along $t_2$ would continue to hold under the alternative assumption of random tie breaking. Even in this case, in fact, there would be no systematic selection of values of $t_2$ in the stereotypes of different DMs.

plausible stereotype for the Irish is "Catholic" because religion is the dimension along which Irish and Scots differ the most.

Formally, suppose that groups are characterized by two dimensions: hair color (red $r$, other $o$), and religion (catholic $c$, other $\hat{o}$). The Irish have a share $r_i$ of red haired people and a share $c_i$ of catholics. Europeans have a share $r_e$ of red haired people and a share $c_e$ of catholics. Critically, the Irish have a much higher share of red haired people, $r_i > r_e$, while catholics are similarly prevalent along the two groups, namely $c_i = c_e$. Hair color and religion are statistically independent in both populations.

Consider the stereotypes formed by comparing the Irish to Europeans. Lemma 1 implies stereotypes depend on the joint distribution of these variables. Because $c_i = c_e$, the representativeness of different types for the Irish is then given by:

$$\begin{aligned} R_i(r,c) = \frac{r_i \cdot c_i}{r_e \cdot c_e} = \quad & \tfrac{r_i}{r_e} \quad = \frac{r_i \cdot (1-c_i)}{r_e \cdot (1-c_e)} = R_i(r,\hat{o}) > \\ > R_i(o,c) = \frac{(1-r_i) \cdot c_i}{(1-r_e) \cdot c_e} = \quad & \tfrac{1-r_i}{1-r_e} \quad = \frac{(1-r_i) \cdot (1-c_i)}{(1-r_e) \cdot (1-c_e)} = R_i(o,\hat{o}). \end{aligned}$$

The inequality follows because $r_i > r_e$ implies that $\frac{r_i}{r_e} > \frac{1-r_i}{1-r_e}$. As a consequence, when $d = 1$, the stereotype for the Irish contains the two equally representative types of (red haired, catholic) and (red haired, other). The stereotype differentiates the Irish from the Europeans along the color of hair dimension.

Suppose now that the Irish are compared to the Scots, who have a share $r_s$ of red haired people and a share $c_s$ of catholics. The Scots have a similar share of red haired people, $r_i = r_s$, while they have a much lower share of catholics, namely $c_i > c_s$. Consider the stereotype formed by comparing the Irish to the Scots. In this case, the representativeness of different types for the Irish is:

$$\begin{aligned} R_i(r,c) = \frac{r_i \cdot c_i}{r_s \cdot c_s} = \quad & \tfrac{c_i}{c_s} \quad = \frac{(1-r_i) \cdot c_i}{(1-r_s) \cdot c_s} = R_i(o,c) > \\ > R_i(r,\hat{o}) = \frac{r_i \cdot (1-c_i)}{r_s \cdot (1-c_s)} = \quad & \tfrac{1-c_i}{1-c_s} \quad = \frac{(1-r_i) \cdot (1-c_i)}{(1-r_s) \cdot (1-c_s)} = R_i(o,\hat{o}) \end{aligned}$$

Note that now $c_i > c_s$ implies that $\frac{c_i}{c_s} > \frac{1-c_i}{1-c_s}$. As a consequence, when $d = 1$, the stereotype for the Irish contains the two equally representative types of (red haired, catholic)

and (other, catholic). The dimensions along which the Irish stereotype is formed has changed: it differentiates the Irish from the Scots along the religion dimension, not along hair color.

In summary, because stereotypes are centered along the types for which the groups differ the most, the kernel of truth logic survives when types are multidimensional. The features that are perceived as characteristic of a group depend on the comparison group.

# D Extension to Continuous Distributions

Many distributions of interest in economics can be usefully approximated by continuous probability distributions. Here we show how our results extend to this case. For simplicy, we only consider rank-based truncation, but the model is easily extended to smooth weighing.

## D.1 Basic Setting

Let $T$ be a continuous variable defined on the support $\overline{T} \subseteq R^k$. Denote by $t \in \overline{T}$ a realization of $T$ which is distributed according to a density function $f(t) : \overline{T} \to R_+$. Denote by $f(t|G)$ and $f(t|-G)$, the distributions of $t$ in $G$ and $-G$, respectively. In line with Definition 1, we define representativeness as:

**Definition 4** *The representativeness of $t \in \overline{T}$ for group $G$ is measured by the ratio of the probability of $G$ and $-G$ at $T = t$, where $-G = \Omega \backslash G$. Using Bayes' rule, this implies that representativeness increases on the likelihood ratio $f(t|G)/f(t|-G)$.*

In the continuous case, the exemplar for $G$ is the realization $t$ that is most informative about $G$. For one dimensional variables, the exemplar for $G$ is $\sup(\overline{T})$ if the likelihood ratio is monotone increasing, or $\inf(\overline{T})$ if the likelihood ratio is monotone decreasing, just as in Proposition 2.

The DM constructs the stereotype by recalling the most representative values of $t$ until the recalled probability mass is equal to the bounded memory parameter $\delta \in [0, 1]$. When $\delta = 0$, the DM only recalls the most representative type. When $\delta = 1$ the DM recalls the entire support $\overline{T}$ and his beliefs are correct. When $\delta$ is between 0 and 1, we are in an intermediate case.

**Definition 5** *Given a group $G$ and a threshold $c \in R$, define the set $\overline{T}_G(c) = \left\{ t \in \overline{T} \mid \frac{f(t|G)}{f(t|-G)} \geq c \right\}$. The DM forms his beliefs using a truncated distribution in $\overline{T}_G(c(\delta))$ where $c(\delta)$ solves:*

$$\int_{t \in \overline{T}(c(\delta))} f(t|G)dt = \delta.$$

The logic is similar to that of Definition 2, with the only difference that now the memory constraint acts on the recalled probability mass and not on the measure of states, which would be problematic to compute when distributions have unbounded support. This feature yields and additional (and potentially testable) prediction that changes in the distribution typically change also the support of the stereotype by triggering the DM to recall or forget some states, even when the states' relative representativeness does not change.

## D.2 The Normal Case

When $f(t|G)$ and $f(t|-G)$ are univariate normal, with means $\mu_G$, $\mu_{-G}$ and variances $\sigma_G$, $\sigma_{-G}$, the stereotype of $G$ is easy to characterize.

**Proposition 5** *In the normal case, the stereotype works as follows:*

*i) Suppose $\sigma_G = \sigma_{-G} = \sigma$. Then, if $\mu_G > \mu_{-G}$ the stereotype for $G$ is $\overline{T}_G = [t_G, +\infty)$, where $t_G$ decreases with $\delta$. Moreover, $\mathbb{E}^{st}(t|G) > \mu_G > \mu_{-G} > \mathbb{E}^{st}(t|-G)$.*

*If instead $\mu_G < \mu_{-G}$, the stereotype for $G$ is $\overline{T}_G = (-\infty, t_G]$, where $t_G$ now increases with $\delta$. Moreover, $\mathbb{E}^{st}(t|G) < \mu_G < \mu_{-G} < \mathbb{E}^{st}(t|-G)$. In both cases, $Var^{st}(t|G) < Var(t|G)$ and $Var^{st}(t|-G) < Var(t|-G)$.*

*ii) Suppose that $\sigma_G < \sigma_{-G}$. Then, the stereotype for $G$ is $\overline{T}_G = [\underline{t}_G, \bar{t}_G]$ where $\underline{t}_G$ decreases and $\bar{t}_G$ increases with $\delta$. Moreover, $Var^{st}(t|G) < Var(t|G)$.*

*iii) Suppose that $\sigma_G > \sigma_{-G}$. Then, the stereotype for $G$ is $\overline{T}_G = (-\infty, \underline{t}_G] \cup [\bar{t}_G, +\infty)$ where $\underline{t}_G$ increases and $\bar{t}_S$ decreases with $\delta$. Moreover, $Var^{st}(t|G) > Var(t|G)$.*

**Proof.** Let $\rho_{\mu,\sigma^2}$ denote the probability density of $N(\mu, \sigma^2)$, namely $\rho(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$. The exemplar $\hat{t}_G$ of $G \equiv N(\mu_G, \sigma_G^2)$ relative to $-G \equiv N(\mu_{-G}, \sigma_{-G}^2)$ satisfies $\hat{t}_E = \text{argmax}_t \frac{\rho_{\mu_G,\sigma_G^2}}{\rho_{\mu_{-G},\sigma_{-G}^2}}$ where

$$\frac{\rho_{\mu_G,\sigma_G^2}}{\rho_{\mu_{-G},\sigma_{-G}^2}} = \frac{\sigma_{-G}}{\sigma_G} \cdot \exp\left\{-t^2\left(\frac{1}{2\sigma_G^2} - \frac{1}{2\sigma_{-G}^2}\right) + t\left(\frac{\mu_G}{\sigma_G^2} - \frac{\mu_{-G}}{\sigma_{-G}^2}\right) - \left(\frac{\mu_G^2}{2\sigma_G^2} - \frac{\mu_{-G}^2}{2\sigma_{-G}^2}\right)\right\}$$

When $\sigma_G < \sigma_{-G}$, the function above has a single maximum in $t$, namely that which maximizes the parabola in the exponent, $\hat{t}_E = \frac{\frac{\mu_G}{\sigma_G^2} - \frac{\mu_{-G}}{\sigma_{-G}^2}}{\frac{1}{\sigma_G^2} - \frac{1}{\sigma_{-G}^2}}$ from which the result follows.

When $\sigma_G > \sigma_{-G}$, the function above is grows without bounds with $|t|$, so that $\hat{t}_G \in \{-\infty, +\infty\}$.

When $\sigma_G = \sigma_{-G} = \sigma$, the exemplar $\hat{t}_G$ of $G \equiv N(\mu_G, \sigma^2)$ relative to $-G \equiv N(\mu_{-G}, \sigma^2)$ satisfies

$$\hat{t}_G = \text{argmax}_t e^{-\frac{\mu_G^2 - \mu_{-G}^2}{2\sigma^2}} \cdot e^{\frac{t}{2\sigma^2}(\mu_G - \mu_{-G})}$$

so that $\hat{t}_G = -\infty$ if $\mu_G < \mu_{-G}$ and $\hat{t}_G = +\infty$ otherwise. If $\mu_G < \mu_{-G}$ all values of $t$ are equally representative. ■

When the two distributions have the same variance, the stereotype is formed by truncating from the original distribution the least representative tail (as in Section 3). In fact, when the mean in $G$ is above the mean in $-G$, the likelihood ratio is monotone increasing and the exemplar for $G$ is $+\infty$; otherwise it is $-\infty$. In both cases, the exemplar is inaccurate because it relies on a highly representative but very low probability realization.

Figure A1, left panel, represents the distribution considered by the DM for the high mean group when traits are normally distributed with the same variance across groups.

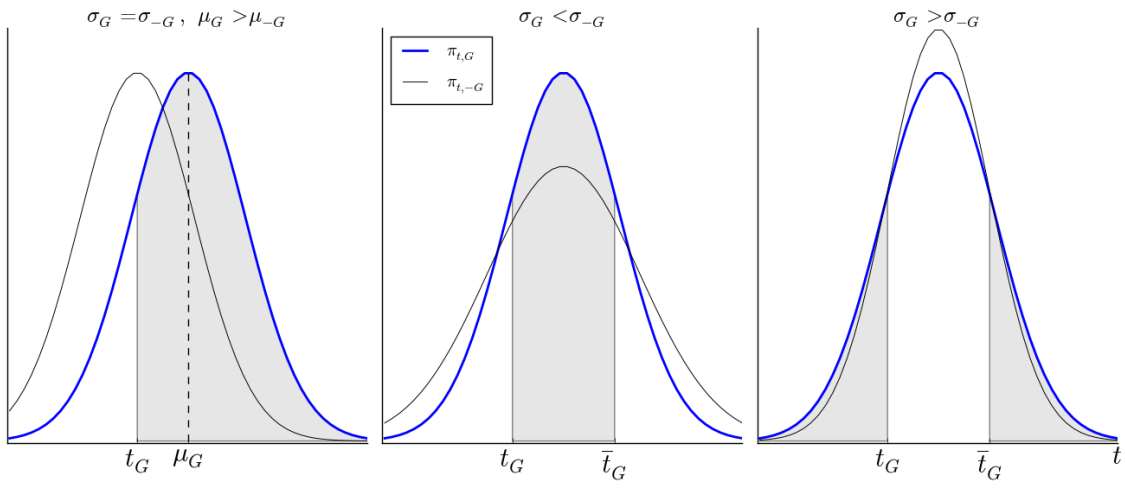



Figure A1: Stereotypes of a Normal distribution as a function of $\mu_{-G}$ and $\sigma_{-G}$.

Consider now case ii), where the variance of $G$ is lower than that of $-G$, Figure A1, middle panel. The stereotype consists of an interval around an intermediate exemplar, denoted by $\hat{t}_G$. When the distribution in $G$ is more concentrated than that in $-G$, the exemplar is accurate

and captures a relatively frequent, intermediate event. It is however somewhat distorted, because $\hat{t}_G$ lies below the group's true mean $\mu_G$ if and only if $\mu_G < \mu_{-G}$. Interestingly, when the mean in the two groups is the same, the low variability group is represented by its correct mean, namely $\mu_G$. Again, because the distinctive feature of group $G$ is being more "average" than group $-G$, its stereotype neglects extreme elements and decreases within group variation.

Finally, consider case iii). Now the variance in $G$ is higher than that in $-G$, Figure A1, right panel. As a consequence, both tails are exemplars and the stereotype includes both tails, truncating away an intermediate section of the distribution. This representation increases perceived volatility and thus captures the distinctive trait of $G$ relative to $-G$, which is precisely its higher variability. Stereotyping now induces the DM to recall group $G$'s most extreme elements and to perceive $G$ as more variable than it really is. This is a testable prediction of our model that stands in contrast with the previous cases, and with the common description that stereotypes reduce within-group variability (Hilton and Von Hippel 1996). However, it is consistent with the more basic intuition that stereotyping highlights the most distinctive features of group $G$, in this case its extreme elements. As an illustration of this mechanism, when thinking about stock returns, investors may think of positive scenarios where returns are high, or negative scenarios where returns are low, but neglect average returns, which are more typical of safer asset classes.

Consider now dynamic updating in this normal case. The DM receives information about the distributions $f(t|G)$ and $f(t|-G)$ over time. In each period $k$, a sample $(t_{G,k}, t_{-G,k})$ of outcomes is observed, drawn from the two groups. The history of observations up to period $K$ is denoted by the vector $t^K = (t_{G,k}, t_{-G,k})_{k=1,\ldots,K}$.

Based on $t^K$, and thus on the conditional distributions $f(t|W, t^K)$ for $W = G, -G$, the DM updates stereotypes and beliefs. In one tractable case, the $k = 0$ initial distribution $f(t|W)$ is also normal for $W = G, -G$. Formally, suppose that $t_W = \theta_W + \varepsilon_W$ where $\varepsilon_W$ is i.i.d. normally distributed with mean 0 and variance $v$, and $\theta_W$ is the group specific mean. Initially, groups are believed to be identical, in the sense that both $\theta_G$ and $\theta_{-G}$ are normally distributed with mean 0 and variance $\gamma$. After observing $(t_{G,1}, t_{-G,1})$, the distribution of $\theta_W$ is updated according to Bayesian learning. Updating continues as progressively more

observations are learned. Thus, after observing the sample $t^K$, we have:

$$f(t|W,t^K) = \mathcal{N}\left(\frac{\gamma \cdot K}{v+\gamma \cdot K} \cdot \frac{\sum t_{W,k}}{K}; v \cdot \frac{v+\gamma \cdot (K+1)}{v+\gamma \cdot K}\right). \tag{9}$$

The posterior mean for group $W$ is an increasing function of the sample mean $\sum t_{W,k}/K$ for the same group. The variance of the posterior declines in sample size $K$, because the building of progressively more observations reduces the variance of $\theta_W$, in turn reducing the variability of outcomes. However, and importantly, because the same number of observations is received for each group, both groups have the same variance in all periods.

Consider now how learning affects stereotypes. Proposition 5 implies:

**Proposition 6** *At time $K$, the stereotype for group $G$ is equal to $[t_G, +\infty)$ if $\sum t_{G,k} > \sum t_{-G,k}$ and to $(-\infty, t_G]$ if $\sum t_{G,k} < \sum t_{-G,k}$. As a result:*

*i) Gradual improvement of the performance of group $G$ does not improve that group's exemplar (and only marginally affects its stereotype) provided $\sum t_{G,k}$ stays below $\sum t_{-G,k}$. In particular, common improvements in the performance of $G$ and $-G$ (which leave $\sum t_{G,k} - \sum t_{-G,k}$ constant) leave stereotypes unaffected.*

*ii) Small improvements in the relative performance of $G$ that switch the sign of $\sum t_{G,k} - \sum t_{-G,k}$ have a drastic effect on stereotypes.*

**Proof.** Since the variances of the sample populations $G$ and $-G$ are equal, the stereotypes are fully determined by the sample means. From Proposition 5, if $\sum_t t_{G,k} > \sum_t t_{-G,k}$, then the sample mean of $G$ is higher than that of $-G$, so that its exemplar is $\hat{t}_G = +\infty$. If instead $\sum_t t_{G,k} < \sum_t t_{-G,k}$, the exemplar of $G$ is $\hat{t}_G = -\infty$. Cases i) and ii) follow directly from this. ■

Even in the normal case, stereotyping suffers from both under- and over-reaction to information. If new information does not change the ranking between group averages, exemplars do not change and stereotypes only respond marginally. Thus, even if a group gradually increases its average, its stereotype may remain very low. In contrast, even minor information can cause a strong over-reaction if it reverses the ranking between group averages.

# E Likelihood, Availability, and Stereotypes

As we discussed in Section 2.2, our formulation of representativeness-based stereotypes leads in some instances to extreme predictions and, importantly, neglects other factors that influence what features come to mind when thinking about a group, such as likelihood and availability.[36] When stereotyping the occupation of a democratic voter, people think about "professor" rather than a "comparative literature professor." While the latter is probably more representative, the former is more likely and thus comes to mind more easily.

In this section we show that our model can be easily adapted to account for some effects of likelihood on recall. When we do so, our predictions become less extreme, in the sense that stereotypes become centered around relatively more likely or available types, but the distortions of stereotypes still follow the logic of representativeness, as in our main analysis. This extension can also capture the effects of a crude measure of availability on recall. For simplicity, we focus on a rank-based trunction specification.

Suppose that the ease of recall of a type $t$ for group $G$ is given by:

$$R_k(t,G) = \frac{\pi_{t,G}}{\pi_{t,-G} + k} = \frac{1}{\frac{1}{R(t,G)} + k \cdot \frac{1}{\pi_{t,G}}} \tag{10}$$

where $k \geq 0$ and $R(t,G)$ is representativeness as defined in Definition 1. In Equation (10), the ease of recalling type $t$ increases when that type is more representative, namely when $R(t,G)$ is higher, but also when type $t$ is more likely in group $G$, namely when $\pi_{t,G}$ is higher. The value of $k$ modulates the relative strength of these two effects: for small $k$, representativeness drives ease of recall, while for large $k$ likelihood drives recall.[37]

In this new formulation, the stereotype is formed as in Definition 2 except that now what comes to mind are the $d$ types that are easiest to recall. When representative types are also likely, recall based on Equation (10) does not change the stereotype for group $G$. When instead representativeness and likelihood differ for group $G$, recall driven by $R_k(t,G)$ may

[36] According to Kahneman and Frederick (2005) "the question of why thoughts become accessible – why particular ideas come to mind at particular times – has a long history in psychology and encompasses notions of stimulus salience, associative activation, selective attention, specific training, and priming".

[37] When $k = 0$, we are in a pure representativeness model. As $k$ increases, likelihood becomes progressively more important in shaping recall relative to representativeness. As $k \to \infty$, only likelihood matters for shaping recall and stereotypes.

yield a different stereotype than a pure representativeness model.

To see how the model can capture some features of availability, note than the term $\pi_{t,G}$ in (10), and also in (2), may be broadly interpreted as capturing the availability, rather than just the frequency, of type $t$ for group $G$. Formally, we would assume that the estimate of $\pi_{t,G}$ is determined by the share of observations from $G$ that are of type $t$, even if these observations are not independent. Thus, as the same episodes of terrorism are mentioned repeatedly in the news, their ease of recall is inflated. In this approach, availability is related to neglect of the correlation structure of information (as discussed in Section 2.2, the psychology of availability is beyond the scope of this paper).

The concrete implications of Equation (10) are best seen in the case where the type space is continuous, and more specifically when $t$ is normally distributed in groups $G$ and $-G$, with means $\mu_G, \mu_{-G}$ respectively, and variance $\sigma$. In this case, the easiest to recall type $t$ for group $G$ is given by:

$$t_{E,G} = \text{argmin}_t e^{\frac{(t-\mu_G)^2-(t-\mu_G)^2}{2\sigma^2}} + k \cdot e^{\frac{(t-\mu_G)^2}{2\sigma^2}}$$

When $\mu_G > \mu_{-G}$, the easiest to recall type $t_{E,G}$ satisfies:

$$k \cdot (t_{E,G} - \mu_G) \cdot e^{\frac{\left(t_{E,G}-\mu_G\right)^2+2(\mu_G-\mu_{-G})\cdot\left(t_{E,G}-\frac{\mu_G+\mu_{-G}}{2}\right)}{2\sigma^2}} = \mu_G - \mu_{-G} \tag{11}$$

The left hand side of (11) is increasing in $t_{E,G}$, which implies that $t_{E,G}$ is a strictly increasing function of $k$ satisfying $\lim_{k\to\infty} t_{E,G}(k) = \mu_G$ and $\lim_{k\to 0} t_{E,G}(k) = \infty$. In words, the group $G$ with higher mean is stereotyped with an inflated assessment that goes in the direction of the most representative type $t = \infty$. The extent of this inflation increases as $k$ gets smaller. The stereotype for group $G$ in this case is an interval around the easiest to recall type that captures a total probability mass of $\delta$ (truncating both tails, but especially the left one). Moreover, as in the case $k = 0$, the stereotype has a lower variance than the true distribution. A corresponding result is obtained if group $G$ has a lower mean than $-G$.

This analysis implies that the basic insights that stereotypes emphasise differences, and lead to base rate neglect, carry through to this case.[38]

[38]In the extended model given by (10), the parameters $\delta$ and $k$ capture two natural types of bounds on recall: $\delta$ determines "how much" comes to mind (which might depend on effort), while $k$ corresponds to the

# F Experiments

## F.1 Analysis of All Unordered Types Experiments

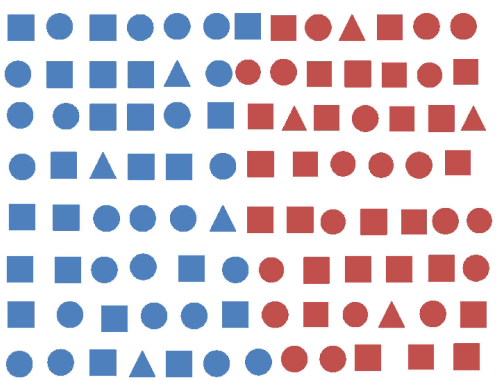
We conducted four experiments on unordered types. The final experiment, using cartoon characters in T-Shirts, were reported in the main text. Here we discuss the other experiments and their results.

### F.1.1 Unordered Types Experiment 1: (Lots of) Triangles, Squares, and Circles

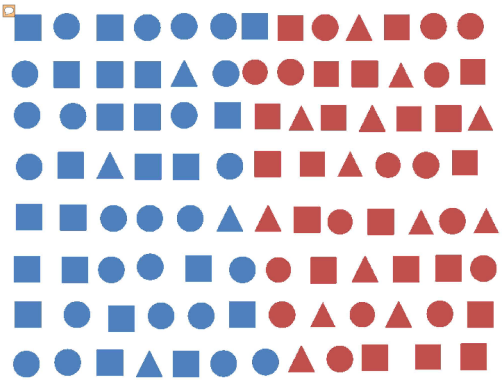
The first unordered types experiment used groups of 50 shapes each. The groups were characterized by color (red shapes or blue shapes) and the types were shapes (triangles, squares, and circles). In both conditions, the blue group contained 22 squares, 24 circles, and 4 triangles. In the Control condition, this blue group was presented next to a similar red group that contained 26 squares, 20 circles, and 4 triangles. Note that in the Control condition, within each group, the most representative type and the modal type coincide: among the blue shapes, circles are both most representative and modal, and among the red shapes, squares are both most representative and modal. In the Representativeness (Rep.) condition, we drive a wedge between modality and representativeness by changing the distribution of red shapes presented next to the blue group. In the Rep. condition, the red group contains 21 squares, 16 circles, and 13 triangles. While circles are still most representative and modal among the blue group, in the red group the modal shape is a square while the most representative shape is a triangle. Our prediction is that participants will be more likely to guess that the triangle is modal among the red shapes in the Rep. condition than in the Control condition. The images as they appeared to participants are reproduced in Figure A2.

This design is not as clean as the T-shirts design presented in the paper. Most importantly, if we do see an increase in the fraction of participants that believe triangles are modal in the Rep. condition, we cannot rule out that this is simply driven by the fact that there are more red triangles in the Rep. condition than in the Control condition. We present the results below with that caveat.

relative weight of likelihood in recall, which may vary across people.



(c) Control Condition



(d) Representativeness Condition

Figure A2: 50 Shapes Experiment

This shapes experiment was conducted on MTurk in November 2014 with 217 participants.[39] Participants viewed the shapes for 15 seconds and then completed 10 simple addition problems (computing sums of two-digit numbers) before answering a series of questions about the shapes they saw. They were asked to guess what the most common shape among each group was and to estimate the frequency of each shape in each group. They received $0.30 for completing the HIT and an additional $1 if they answered one of the randomly-selected questions about the shapes correctly.

We find that 7% of participants in the Control treatment and 13% of participants in the Rep. treatment believe that the triangle is the modal red shape. The direction matches our prediction but the effect is not significant at conventional levels (p=0.17).

After conducting this experiment, we altered the design to eliminate the potential confound. In all designs going forward, we hold fixed the number of objects of the type of interest across the Control and Rep. treatment and simply alter the comparison group to change whether or not the type is diagnostic.

### F.1.2 Unordered Types Experiment 2: (Fewer) Triangles, Squares, and Circles

The next iteration improved on the original shapes design in a few important ways. First, we cut down the number of shapes, reducing the groups from 50 shapes each to 25 shapes

[39]This count excludes 3 participants who self-identified as color blind. Neither the point estimates or p-values reported below are changed if those participants are included in the analysis. The HIT was posted once for 200 participants and we had 220 complete the experiment on Qualtrics via the link (some fail to submit the payment code to MTurk for payment, allowing us to overshoot our target.)

each. Second, we changed the distributions such that the number of red triangles was held constant across condition, but the number of blue triangles varied to change how diagnostic the triangles were for the red group. In both conditions, the red group contained 6 squares, 10 circles, and 9 triangles. In the Control condition, this group was presented next to a blue group that contained 9 squares, 8 circles, and 8 triangles. In the Rep. condition, the red group was presented next to a different blue group that contained 11 squares, 12 circles, and 2 triangles. Thus, while the number of red triangles is the same across conditions, triangles are much more representative of the red group in the Rep. condition than in the Control condition. We predict this shift in representativeness of the red triangles will lead to an increase in the proportion of participants who guess that triangles are modal in the red group and an increase in the estimated frequency of red triangles.

We ran this experiment both on MTurk and at the Stanford Experimental Economics Laboratory in January 2015. The MTurk protocol was very similar to Experiment 1, the previous shapes experiment. Participants viewed the objects for 15 seconds, answered 10 simple addition questions, then answered a series of questions about the shapes. Participants were paid \$0.30 for completing the HIT and an additional \$1 if they answered a randomly-selected question about the shapes correctly. We collected data from 100 participants, 50 in each condition.[40]

In the Control condition, 18% of participants believed triangles were modal in the red group; in the Rep. condition, this grows to 24% (p=0.46 from two-tailed test of proportions).[41] Participants in the Rep. condition estimate that there are 9.98 red triangles on average, while participants in the Control condition estimate that there are 9.39 red triangles on average (two-tailed t-test, p=0.65). But, this difference is largely driven by one participant who provided an unusually large estimate of red triangles in the Rep. condition (50). If we exclude this participant, the data on estimated frequencies is not directionally consistent with our hypothesis, with the average estimate of red triangles being 9.39 in the Control condition being and 9.16 in the Rep. Condition (two-tailed t-test, p=0.82).

[40]This count excludes 1 participant who self-identified as color blind. Including this participant does not impact the results presented below. We posted the HIT once for 100 participants.

[41]Using a probit regression that controls for demographics (gender and year of birth) also estimates approximately a 6 percentage point increase in the fraction of participants that believe the triangles are modal in the red group.

The protocol in the Stanford laboratory was more complicated, with several potentially important changes. First, instead of arranging the shapes on a page for participants, we provided participants with an envelope that contained cutouts of each of the 50 total shapes for their condition. Participants were given 1-minute to open the envelope and view the contents. Second, in the laboratory, we had participants complete both an ordered and an unordered types experiment, back-to-back, in a randomly-assigned order. Third, after viewing the objects in the envelope and completing the math problems, participants were asked to describe their envelope, in writing, to another participant in the lab. This was incentivized as "advice". Take a participant who had been given an envelope labeled "A" (i.e. was assigned to the Control condition). We told this participant that later in the experiment, we were going to ask another participant in the lab, who had been given a different envelope, a question about envelope "A". This participant would receive the advice, but not the envelope. If the participant answered the question about envelope "A" correctly, both the advice giver and the other participant would receive additional payment. Thus, participants were incentivized to write down information about the shapes in their envelope that would be accurate and useful. Thus, we likely encouraged some careful reflection on their envelope before the participant had answered any of our other questions of interest about the shapes. We ran four laboratory sessions, with 66 total participants.[42]

The Stanford laboratory results do not support our hypotheses. In the Control condition, 33% of participants believe triangles are the modal red shape; in the Rep. condition, 27% of participants believe triangles are the modal red shape (p=0.59 from two-tailed test of proportions). This result does not depend on whether participants completed this unordered types experiment first or second. Participants also estimate -0.61 fewer red triangles in the Rep. condition than in the Control condition. This difference goes in the opposite direction of our prediction, though it is not significant.

The results for this design are the weakest among our unordered types experiments. While we do not have conclusive evidence on what drives these effects, we do have a hypothesis that seems consistent with the data. It may be the case that participant judgments were swayed

[42]Our ex ante plan was to run four sessions, though we had thought this would yield closer to 100 participants. After four sessions, we stopped and attempted to improve the design as described below.

by the total number of each type, pooled across groups. Consider the triangle questions. We expect that the 9:2 red triangle to blue triangle ratio in the Rep. condition, relative to the 9:8 red triangle to blue triangle ratio in the Control condition, will lead participants to estimate a larger share of red triangles. But, it is also true that they see 11 total triangles in the Rep. condition, but 17 total triangles in the Control condition. In the laboratory experiment, unlike on MTurk, the shapes are not arranged by group for participants; they are loose in an envelope. If the distinction between the groups is not natural at the moment when they are forming their impressions of the envelope they saw, the fact that there are fewer total triangles may carry more weight than the representativeness of the triangle within each group. This force may push them toward estimating that there were fewer red triangles in the Rep. condition.

After these results, we sought to improve the experiment. In particular, we moved to simpler distributions, where only two types of objects appeared within a given group. This amplified the extent of diagnosticity, as certain types now appear in only one of the two groups. We also shifted from using shapes to more familiar objects, thinking that this might make "groups" a more natural concept. We also switched back to displaying the objects in a fixed arrangment for participants, so we could arrange the objects into obvious groups.

#### F.1.3 Unordered Types Experiment 3: Cars, Trucks, and SUVs

Next, we ran a version of the experiment that used groups of vehicles. The groups were defined by color, with a group of blue vehicles and a group of green vehicles. The types were defined by type of vehicle: pick-up truck, sedan, or SUV. Each group had 20 vehicles. The distributions were similar to the T-shirt design. The green group of vehicles consisted of 9 SUVs and 11 sedans. In the Control condition, this group was displayed next to a group of blue vehicles with the same distribution, 9 SUVs and 11 sedans. Thus, in the Control condition, there is no vehicle type that is diagnostic of a group. In the Rep. condition, the green group was displayed next to a blue group with 9 trucks and 11 sedans. As with the T-shirts design, this creates a tension between the modal type and the diagnostic type in each group. In the green group, the sedan is modal but the SUV is diagnostic; in the blue group, the sedan is modal but the truck is diagnostic. Thus, we predict that participants

in the Rep. condition will be more likely to guess that the "9 vehicle" type is modal for a group, because the 9-vehicle type is diagnostic in this condition. The images, exactly as they appeared to participants, are reproduced in Figure A3.

(a) Control Condition (b) Representativeness Condition

Figure A3: Vehicles Experiment

We conducted this experiment with 57 participants on MTurk in January 2015.[43] The protocol was very similar to the T-shirts experiment reported in the main text. Participants were given 15 seconds to review the objects, seeing the green group next to a randomly-chosen comparison group, either the Control blue group or the Rep. blue group. Then, participants were asked what the most common type of vehicle was for each group and were asked to estimate the frequency of different types of vehicles for each group. Participants received \$0.30 for completing the HIT and an additional \$2 in incentive pay if they answered a randomly-selected question correctly.

Our results support our hypothesis. In the Control condition, when the 9-vehicle type is not diagnostic, participants guess that the 9-vehicle type is modal in 22% of cases. In the Rep. condition, when the 9-vehicle type is diagnostic of each group, participants guess that this 9-vehicle type is modal in 40% of cases (significantly different than the Control condition using a two-tailed test of proportions with $p = 0.042$). Because we have two observations per individual (her guess of the most common blue vehicle type and her guess of the most common green vehicle type), it is useful to run a probit regression that allows us to cluster

[43]The HIT was posted once, for 150 participants, to be randomized in equal proportions into this experiment and the ice cream ordered types experiment. We collected data from 76 participants for this experiment, but 19 had participated in a previous version of the experiment, leaving us with 57 participants. The results below are directionally stronger if we include those repeat participants.

observations at the individual level. When we predict the probability of guessing the 9-vehicle type is modal from a participant's randomly-assigned treatment, her demographic information (gender and year of birth), and a dummy for whether the guess was for the blue or green vehicles, we estimate that participants in the Rep. condition are 17.4 percentage points more likely to guess the 9-vehicle type is modal (p=0.09).

We can also look at estimated frequencies of different types across condition. In this experiment, we only asked participants to estimate the number of green sedans and SUVs and blue sedans and pickups (the types that appeared in the Rep. conditions). Thus, because we are missing estimates of the blue SUVs, we cannot do quite the same analysis presented for the T-shirts design, where we compared the estimate of the modal type and the 9-vehicle type for each group across conditions. But, we can do this analysis for the green group, asking how the estimated difference in number of sedans and SUVs varies across conditions. We predict that participants will estimate a greater gap between green sedans and SUVs in the Control than in the Rep. condition. Using an OLS regression, we predict the estimated difference between the number of green sedans and green SUVs from a participant's randomly-assigned condition and her demographic information. We find that the effect is small but directionally supportive of our hypothesis, with participants in the Rep. condition estimating the difference in sedans and SUVs to be about 0.5 counts smaller than participants in the Control condition (p=0.68).

We moved from using the vehicles to using the cartoon characters wearing T-shirts in an attempt to simplify the objects. The pictures of vehicles are highly detailed, providing many features that could capture participants' attention during their brief 15-second viewing. Furthermore, recognizing the same type across group was not straightforward – i.e. the green sedan and blue sedan have many differences in addition to color. We wanted to move to a format where fewer features would vary, and where recognizing the same type across group would be simpler. This led us to the T-shirts design.

#### F.1.4 Unordered Types Experiment 4: T-Shirts

The T-Shirts design was reported in the main text. We ran this experiment in the laboratory and on MTurk. On MTurk, participants received $0.30 for completing the experiment and an

additional $1 if they answered the randomly-selected question correctly. Data was collected in February 2015. The laboratory sessions were conducted at the Ohio State Experimental Economics Laboratory in March 2015. Participants dropped into the lab for approximately five minutes, receiving a $5 show-up fee and up to $5 more in incentive pay. Inthe lab, we added two questions on risk preferences between the viewing of the objects and the questions about the T-shirt people in order to better obscure our focus.

We had 301 total participants, 196 in the laboratory and 105 on MTurk.[44] We have two observations for each individual: her guess of the most common color shirt among the girls and her guess of the most common color shirt among the boys. Our main hypothesis is confirmed in the pooled data (including guesses about both girls and boys): participants in the Control condition believe the 12-shirt color is modal in 35% of cases, while this mistake is made in 46% of cases in the Rep. condition (p=0.01 from two-tailed test of proportions). Using a probit regression that clusters observations at the individual level, we estimate that when the 12-shirt color is diagnostic of a group, a participant is 10.5 percentage points more likely to believe it is the modal color (p=0.01). This effect is significant when we restrict attention to the sample from the laboratory (14.4 percentage points, p=0.007) and directional in the smaller MTurk sample (7.6 percentage points, p = 0.26).

We also analyze the difference in estimated counts of the modal color shirt and the counts of the 12-shirt color shirt the participant saw (we subtract estimated counts of the 12-shirt color from estimated counts of the modal color for each participant for each group). We find that, on average, participants in the Control condition estimate having seen 0.54 more modal color shirts than 12-shirt color shirts, while participants in the Rep. condition estimate having seen 0.72 fewer modal color shirts than 12-shirt color shirts (this across treatment difference is significant with p = 0.013 using a two-tailed Fisher Pitman permutation test). Using an OLS regression, we find that when the 12-shirt color is representative, participants estimate the difference in counts between the true modal color and the 12-shirt color to be 1.39 counts smaller (p=0.006). The results are similar and significant within either

[44]We recruited 150 participants for the MTurk experiment, but 45 who completed our HIT had already completed a previous version of the experiment and are excluded from our analysis. The target for the laboratory sample was 200 participants over three days of drop-in sessions. We had 202 participate, but we exclude 6 laboratory participants who self-reported color blindness. The results are very similar if all of these participants (both repeat participants for MTurk and color blind) are included.

subsample, lab or MTurk.

### F.1.5 Summary of Unordered Types Experiments

Table A1 summarizes the results from the four unordered types designs. For each experiment, we run a probit regression predicting the probability that the participant believed a less common type was the modal type from whether or not the type was representative. For the vehicle and T-shirts experiments, we have two observations per individual and we cluster the standard errors at the individual level. For the shapes experiments, we have one observation per individual. We report the marginal effect of assignment to the Rep. condition (where the less common type was representative) on the probability of guessing that the less common type was modal. The last row reports the same coefficient, but from a probit regression that uses all of the data from the unordered types experiments. We include a dummy for each particular experiment and cluster observations at the individual level. We find a directional effect consistent with our hypothesis in five of the six samples – all but the Stanford laboratory sample for Experiment 2 (the 25 Shapes design). When we pool all data, we estimate that a participant is 9.3 percentage points more likely to believe the less common type is modal when it is representative than when it is not (p=0.002). If we include, in addition, all color blind participants, this estimate is 9.0 percentage points (p=0.002); and, if we include all observations, including all observations from participants who have participated in previous versions of the experiments, this estimate is 8.3 percentage points (p=0.003).

We perform a similar analysis using the data on estimated frequencies. The ideal analysis would look at how the magnitude of the difference in the estimated frequency of the modal type less the estimated frequency of the less common type changes across condition. We can do this calculation in Experiments 2 - 4. In Experiment 1, the true frequency of the less common type varies across condition, so this analysis is not useful. In Experiments 2 - 4, the true frequency of both the modal type and the less common type are held constant across treatment. Therefore, we can explore how this difference varies based upon whether the less common type is representative. The prediction is that the difference in estimated frequencies should decrease when the less common type is representative, as participants in

the representativeness condition will estimate fewer counts of the modal type (as it is now less representative for the group) and more counts of the less common type (as it is now more representative for the group).[45] When we pool the data from Experiments 2 - 4, we estimate that the difference in estimates of the modal type and the less common type decrease by approximately 1.06 counts in the representativeness condition (p=0.010). If we include, in addition, color blind participants, this estimate is a decrease of 1.09 counts (p=0.008), and if we include all observations, including those from participants who have participated in multiple versions of the experiment, the estimate is a decrease of 1.19 counts (p=0.002).

## F.2 Analysis of All Ordered Types Experiments

We conducted two experiments on ordered types. The final version, using ice cream cones, was reported in the main text. Here, we report the other experiment and discuss the complete set of results.

### F.2.1 Ordered Types Experiment 1: Rectangles

Our first design for the ordered types experiment used groups of rectangles of varying heights. We created a group of blue rectangles, each of which were 1-unit wide and 1, 2, 3, 4, or 5 units tall. In the Control condition, this group was presented next to a group of red rectangles of the same width with a very similar distribution over heights. In the Rep. condition, the blue group was presented next to a red group of rectangles with the same width, but with a distribution over heights that created a representative tall type for the red group. Table A2 displays the distribution, and Figure A4 presents the images, exactly as they appeared to participants.

In the Control condition, no type is very representative of either group, and the small difference in the distributions occurs at types close to the mean. In the Rep. condition, on

[45]Note that in Experiment 2, we did not ask participants for their estimates of counts of the modal type. Therefore, we simply analyze the change in the difference 0 minus the estimated counts of the less common type for that experiment. We have one observation for each individual (0 - estimate of red triangles). For Experiment 3, we also have one observation per individual (estimate of green sedans - estimate of green SUVs). For Experiment 4, we have two observations per individual (estimate of modal color - estimate of 12-shirt color for both boys and girls). We cluster at the individual level, giving us 824 observations for 523 individuals.

Table A1: Summary of All Unordered Types Experiments

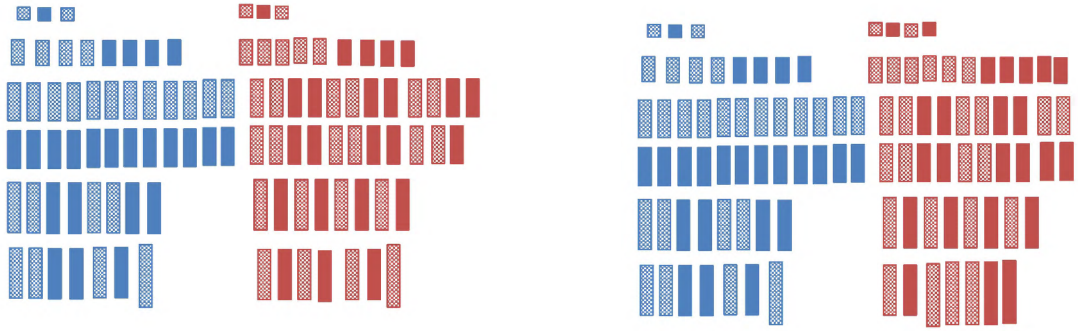
| Experiment | # of Subjs. | Percentage Point Increase in Prob. Guess Less Common Type is Modal when it is Representative | p-value | Change in Diff. in Estimated Frequencies $\Delta$(Modal - Less Common) | p-value |
|---|---|---|---|---|---|
| 1: 50 Shapes on MTurk | 217 | 5.6 pp | 0.17 | N/A | |
| 2: 25 Shapes Pooled | 166 | 1.4 pp | 0.84 | -0.12 | 0.88 |
| MTurk Only | 100 | 5.6 pp | 0.49 | -0.50 | 0.70 |
| Lab Only | 66 | -13.8 pp | 0.28 | 0.61 | 0.34 |
| 3: Vehicles on MTurk | 57 | 17.4 pp | 0.09 | -0.45 | 0.68 |
| 4: T-Shirts Pooled | 301 | 10.5 pp | 0.014 | -1.39 | 0.006 |
| MTurk Only | 105 | 7.6 pp | 0.25 | -2.19 | 0.012 |
| Lab Only | 196 | 14.4 pp | 0.007 | -1.16 | 0.056 |
| **Pooled** | **741** | **9.3 pp** | **0.002** | **-1.07** | **0.010** |

Notes: Std. errors are clustered at the individual level. We report the marginal effect of the coefficient on treatment from a probit regression predicting the probability of the error. Each specification includes all demographic variables collected for that experiment. The pooled specification includes only treatment and gender, as this is the only demographic variable that was collected across all experiments.

Table A2: Distributions for Ordered Types Experiment 1

| Height in Units (Types) | Counts for Blue Group | Counts for Control Red Group | Counts for Rep. Red Group |
|---|---|---|---|
| 1 | 3 | 3 | 4 |
| 2 | 8 | 9 | 11 |
| 3 | 24 | 23 | 20 |
| 4 | 14 | 14 | 10 |
| 5 | 1 | 1 | 5 |
| Total Counts | 50 | 50 | 50 |
| Mean Height | 3.04 | 3.02 | 3.02 |

the other hand, we create a highly representative type for the red group, as there are five 5-unit tall rectangles in the Rep. red group and only one 5-unit tall rectangle in the blue group. Importantly, across both conditions, the means of the two groups are held constant, with the blue group always having a mean height of 3.04 units and the red group having a mean height of 3.02 units. The prediction is that participants will be more likely to guess that the red rectangles are taller on average in the Rep. Condition than in the Control condition, because of the representative tall type among the Rep. red group.



(a) Control Condition (b) Representativeness Condition

Figure A4: Rectangles Experiment

We chose to arrange the rectangles by height for participants so that it might be easier to digest and make sense of the groups in a short period of time. We also varied the fill of the rectangles, with half of each group's rectangles having a solid fill and half displaying a checkered fill.[46] Our fear was that if only the heights varied by shape, participants might anticipate that we were particularly interested in their impressions of the heights of the rectangles. So, we chose to vary the fill as well to create another plausible dimension of interest.

The first experiment using this rectangles design was conducted on MTurk in November 2014 with 113 participants.[47] Participants were randomly-assigned to view either the Control rectangles or the Rep. rectangles for 15 seconds. Then, they completed simple addition

[46]The fill was performed such that approximately half of each type within each group received each fill. That way, the representativeness patterns we sought to induce in the distributions were preserved within each fill.

[47]The HIT was posted once for 100 participants, and 114 completed the experiment via the link to Qualtrics. We exclude one participant who self-identified as color blind. The point estimates and p-values reported below are unchanged if this participant is included.

problems for approximately 3 minutes, computing sums of two-digit numbers. Finally, participants were asked questions about the shapes they saw, including which color rectangles were taller on average, which group of rectangles they would prefer to choose from if they were going to earn \$0.50 per unit height of a randomly-drawn rectangle, and the average height of each group of rectangles. We also asked about the fill of the rectangles they saw, so not all questions would focus on height. Participants received \$0.30 and up to an additional dollar in incentive pay based upon their answers to the questions about the shapes.

The results are consistent with our hypotheses. In the Control condition, 40% of participants believed the red group was taller on average, while in the Rep. condition, 60% of participants believed the red group was taller on average (p=0.03 from two-tailed test of proportions). When we look at which group of shapes participants preferred to bet on, the results are weaker but still directionally supportive: 45% of participants in the Control and 58% of participants in the Rep. group prefer to choose from the red shapes when they will be paid based upon the height of a randomly-drawn rectangle (p=0.16 from two-tailed test of proportions).

There is no difference in estimated average height of the red shapes across condition (3.28 in the Control versus 3.29 in the Rep. condition, p=0.95). If we look at the estimated average height of the blue shapes – recall that the blue shapes are identical across condition – we see that participants in the Rep. condition believe they are slightly smaller on average, though this difference is not significant (3.30 in the Control versus 3.22 in the Rep. condition, p = 0.59).

We took this design into the laboratory in January 2015 at the Stanford Experimental Economics Laboratory. There were a few potentially important changes to the protocol in the laboratory. For one, we had participants complete both an ordered and an unordered types experiment, back-to-back, in a randomized order. Note that this is the same sample for whom we reported results for the unordered types Experiment 2 above. Instead of participants viewing the objects on a computer screen, we passed out envelopes that contained a printed handout of either the Control or the Rep. shapes. After viewing the handout in the envelope and completing the math problems, participants were asked to describe the handout they had seen, in writing, to another participant in the lab. This was incentivized as "advice",

implemented as described in the previous Stanford laboratory description for unordered types Experiment 2. We ran four laboratory sessions, with 66 total participants.

The results from the laboratory were inconsistent with our hypotheses. We find that 46% of participants in the Control condition believed the red shapes were taller on average, while only 36% of participants in the Rep. condition made this error (p=0.45 from two-tailed test of proportions). When we look at the choices about which group participants preferred to bet on, the results are even more striking. Nearly 67% of participants in the Control condition prefer to choose from the red shapes, while only 27% of participants in the Rep. condition prefer to choose from the red shapes (p=0.001 from two-tailed test of proportions). Looking at the data on estimated average heights across condition, there are no significant differences. Directionally, participants estimate both the blue and the red shapes to be taller on average in the Rep. condition than in the Control condition.

There are a few issues with the rectangles design that we sought to address in later experiments. First, it may have been tricky for participants to recognize and process heights of rectangles. We tried to describe the types in terms of "units" of height, but this likely felt a bit confusing to participants. Therefore, we wanted to move to an ordered space that had more obviously distinct types. That is why we shifted to using "scoops" of ice cream, where the difference between 1, 2, 3, 4, or 5 "units" would be more easily recognizable and familiar. Second, there may have been too many shapes on the page for participants to make sense of in a 15-second viewing period. Looking at the advice participants wrote in the laboratory sessions is very informative. Many participants accurately recalled and described the first row of rectangles (featuring three 1-unit tall blue rectangles and four 1-unit tall red rectangles in the Rep. condition, and three 1-unit red and blue rectangles in the Control condition), but no advice sheet even attempted to describe the final row. It may be that with only 15 seconds, participants only have time to focus on part of the page, and the top of the page may be a likely place to start. This type of behavior would hurt us substantially: if participants are mostly focused on the top of the page, they will miss out on the representative tall types we generated. Even worse, in the first row, there are more short red shapes than short blue shapes in the Rep. condition but not in the Control condition. This could lead to participants thinking, contrary to our prediction, that the red

group is shorter on average in the Rep. condition. If the first row or two is what participants mainly recall, it could also explain why so many participants prefer to bet on blue in the Rep. condition, as they remember there were more of the worst possible payoff shapes among the red group. We decided to cut down the number of objects in order to give participants a better chance to view the group as a whole during a short window. And, perhaps more importantly, we altered the distributions so that the group with the representative tall type would not also have comparatively more of the shortest possible type.

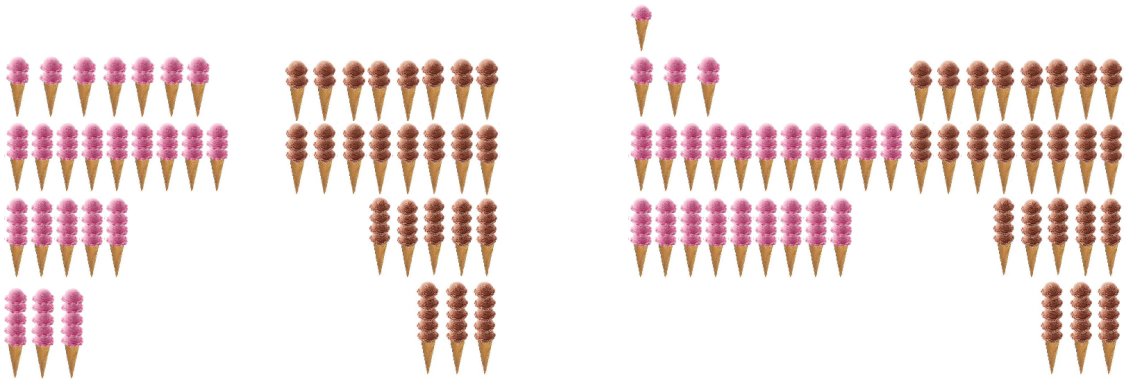
### F.2.2 Ordered Types Experiment 2: Ice Cream

After running the rectangles experiments, we sought to simplify the protocol as much as possible. We did this by reducing the number of objects, but also by eliminating the math problems from between the viewing of the objects and the answering of our questions of interest. This led to the ice cream cone design, illustrated in Figure A5.

Groups are sets of 24 ice cream cones: group membership is defined by ice cream flavor (chocolate vs strawberry), and types are the number of ice cream scoops, ranging from 1 to 5. In the Control condition, Fig.A5a, distributions are very similar, with most cones having intermediate numbers (2 or 3) of scoops. Here, no type is particularly representative of either group. In the representativeness condition, Fig.A5b, the same chocolate cones are presented next to a different group of strawberry cones. In the Representativeness condition, strawberry cones have the same average number of scoops as do the Control condition strawberry cones, but, importantly, they do not contain any 5-scoop cones. This makes the right tail, 5-scoop cones very representative for the chocolate group. Similarly, in this condition only the strawberry group has a cone with 1 scoop, making the left tail very representative for that group.

We ran this ice cream cone design on MTurk in January 2015 with 65 participants.[48] When asked which flavor had more scoops on average, 34% of the Control condition guesses chocolate and 67% of the Rep. condition guesses chocolate (p=0.009). We ask participants

[48]We posted the HIT once for 150 participants, with participants randomized in equal proportions into either this experiment or the vehicles experiment described above. Eighty-four MTurk participants completed this experiment, but 19 of these individuals had participated in a previous version of this experiment and therefore are excluded from this analysis. The results reported are unchanged if these participants are included.



(a) Control Condition (b) Representativeness Condition

Figure A5: Ice cream cone Experiment

a related question using choices over lotteries. They are told that we are going to randomly choose one of the ice cream cones they saw, with the participant earning \$0.50 for every scoop the randomly-chosen cone has. They are asked to choose which flavor we draw from. The proportion that chooses the chocolate lottery grows from 37% in the Control to 57% in the Rep. condition (p=0.11). Finally, we explore the participants' estimates of the average number of scoops on both the chocolate and strawberry cones. In the Control condition, participants believe the strawberry cones have on average 2.85 scoops and the chocolate cones have on average 2.82 scoops. In the Rep. condition, participants believe the strawberry cones have 2.82 scoops on average and the chocolate cones have 2.71 scoops on average. None of these differences, either across condition or flavor, are significant. Overall, the fraction of participants who provide greater estimates of the average number of chocolate scoops than the average number of strawberry scoops is larger in the Rep. conditions than in the Control conditions (60% versus 40%, p = 0.11 from two-tailed test of proportions).

After running this experiment on MTurk, we sought to bring this design into the laboratory. The first ice cream laboratory protocol used the same ice cream cone images with participants directed to answer our questions of interest immediately following the viewing of the objects. In this way, it was likely quite clear to participants what our goal as researchers was: to test their recall of the images they saw. While participants on MTurk are often asked simple attention checks or to report basic objective information (who is this a picture of, transcribe this audio clip, answer this survey question), participants in the laboratory are

likely less familiar with this type of design. It is possible that they were skeptical or wary of being tricked – i.e. why am I being asked what seems to be an obvious question?

We had 56 participants complete this experiment at the Ohio State Experimental Economics Laboratory in March 2015 before we stopped to evaluate what was going on. The results from this experiment look similar to the laboratory data from Stanford. In the Control condition 45% of participants believe the chocolate cones had more scoops on average, while only 36% of participants in the Rep. condition make this error. The effect goes in the opposite direction of our prediction, but is not significant (p =0.48 from two-tailed test of proportions). Similarly, the proportion of people who prefer to bet on the chocolate cones falls from 48% in the Control to 32% in the Rep. condition (p=0.21). There is no difference in estimated average number of scoops of either flavor across condition.

Having seen these results, we brainstormed why participants in the Rep. condition in the lab would be less likely to believe the chocolate cones are taller on average. A directional effect opposite the predicted direction suggests something else at work – something that is not at work on MTurk, where both the rectangle and the ice cream design produced results that support our hypotheses. We conjecture that participants in the laboratory are more skeptical of being tricked, perhaps because they are not usually asked something so simple in a typical economics experiment. It may also be that our ice cream design was "too good" – that is, from a quick look at the objects, the chocolate cones quite strikingly appear to have more scoops on average, that participants are worried that this is actually a trick question. We do not have direct data on this issue, but we did change the design in an attempt to address this problem head on.

We used the same distributions of ice cream cones, but added a new, small section on risk preferences between the viewing of the objects and the questions about the objects. This creates a plausible alternative research question – we could be interested in how viewing a particular arrangement of ice cream cones impacts a participant's risk preferences. We paid participants for this risk preference section, and we framed the questions about the ice cream cones as more of an attention check than an item of interest. We also added a question to the very end of the experiment asking participants what they believed the experiment was trying to test. In this new design, no participant correctly identified our focus on number

of scoops. Our interpretation of the data is that the introduction of this "decoy" encourages less skepticism on the part of participants, and perhaps helps us more successfully elicit their quick, gut reactions to the objects, much the way we were able to do on MTurk. This is speculative, but it does seem consistent with the data we have collected.

We had 101 new participants from the Ohio State Experimental Economics Laboratory complete the updated ice cream protocol.[49] When asked which flavor had more scoops on average, 51% guess chocolate in the Control and 56% guess chocolate in the Rep. condition (p=0.61). There are no significant differences in estimates of average scoops across flavor or condition. The Rep. treatment produces an insignificant decrease in the proportion that prefer the chocolate lottery (45% to 38%, p=0.47).

A natural question to ask is why results for the choices over lotteries would be weaker than the results for which flavor had more scoops on average (or which shapes were taller in the rectangles experiment). While an individual who believes that chocolate cones have more scoops on average should believe there is also a greater expected value from the chocolate cone lottery, it does not guarantee that the chocolate cone lottery is the expected utility maximizing choice: risk preferences may also play a role. Therefore, indicating that chocolate cones have more scoops on average does not guarantee that a reasonable participant will also choose the chocolate cone lottery. To shed light on this issue, we asked a different set of participants from the same laboratory population about their hypothetical preferences over these lotteries. We presented the three lotteries (chocolate cones, Control strawberry cones, and Rep. strawberry cones) side-by-side, described as abstract gambles (there was no mention of ice cream and no visual representation of the lotteries). They were then asked to rank the attractiveness of these gambles from most to least attractive. In a sample of 196 participants, 22% prefer the chocolate cones lottery to the lottery induced by the Rep. strawberry, while 39% prefer that same chocolate cones lottery to the lottery induced by the Control strawberry cones. This suggests that risk preferences were likely working against us finding an effect in support of our hypothesis, as this data would predict a 17 percentage point decrease in the proportion choosing chocolate under the Rep. condition. In light of this baseline, the fact that we see only a 10 percentage point decrease in the lab and a 20

[49] Our ex ante target was 100 participants over two days of drop-in sessions.

percentage point *increase* on MTurk suggests that the presence of diagnostic types is shifting choices in line with our hypothesis.

### F.2.3 Summary of Ordered Types Experiments

Table A3 summarizes the results from the two ordered types experiments. For each experiment, we run a probit regression predicting the probability that the participant guessed the shorter group was taller on average from her treatment assignment. We report the marginal effect of assignment to the Rep. condition (where the tallest possible type is the most representative type in the shorter group) on the probability of guessing that the shorter group is taller on average. The last row reports the same coefficient, but from a probit regression that uses all of the data from the ordered types experiments. We include a dummy for each particular sample and cluster observations at the individual level. We find a directional effect consistent with our hypothesis in three of the five samples. When we pool all data, we estimate that a participant is 9.3 percentage points more likely to believe that the shorter group is taller on average when it has a representative tall tail (p=0.062). If we include, in addition, all color blind participants, this estimate is 9.6 percentage points (p=0.055); and, if we include all observations, including participants who have participated in multiple versions of the experiment, this estimate is 8.4 percentage points (p=0.083). Note that for ordered types experiments, there is a significant difference between the laboratory studies and the MTurk studies. Using only the MTurk example, we estimate that a participant is *25 pp* more likely to guess that the shorter group is taller when it has a representative tall tail (p=0.001); the estimate for the laboratory sample is directionally negative, -3.1 pp (p=0.65). This difference in treatment effect across platform is significant (p=0.006).

We also provide the estimated treatment effect on the probability of choosing the shorter group to bet on for each experiment and the pooled estimate. There is no support for this prediction in the data (see discussion on confound of risk preferences above).

Table A3: Summary of All Ordered Types Experiments

| Experiment | # of Subjs. | Percentage Point Increase in Prob. Believed Shorter Group is Taller in Rep. Condition | p-value | Percentage Point Increase in Betting on Shorter Group in Rep. Condition | p-value |
|---|---|---|---|---|---|
| 1: Rectangles Pooled | 179 | 7.3 pp | 0.32 | -6.8 pp | 0.37 |
| MTurk Only | 113 | 19.1 pp | 0.04 | 12.8 pp | 0.18 |
| Lab Only | 66 | -11.0 pp | 0.38 | -39.3 pp | 0.001 |
| 2: Ice Cream Pooled | 223 | 9.2 pp | 0.17 | -1.7 pp | 0.80 |
| Lab, No Decoy | 56 | -12.0 pp | 0.37 | -14.1 pp | 0.28 |
| MTurk | 65 | 30.7 pp | 0.01 | 18.8 pp | 0.13 |
| Lab, Decoy | 101 | 3.5 pp | 0.73 | -7.5 pp | 0.45 |
| **Pooled** | **402** | **9.3 pp** | **0.062** | **-3.8 pp** | **0.45** |

Notes: Std. errors are clustered at the individual level. We report the marginal effect of the coefficient on treatment from a probit regression predicting the probability of the error. Each specification includes all demographic variables collected for that experiment. The pooled specification includes only treatment and gender, as this is the only demographic variable that was collected across all experiments.

# G Empirical Analysis: Further Results

We repeat the analysis in the main text, implementing the regressions in Equations (6, 7). But, we add an additional control: the average likelihood of tail positions. This is the average frequency of the three types above the median for conservatives and the average frequency of the three types below the median for liberals. We again test the hypothesis that $R_H^{cons}$ is a significant predictor of $\mathbb{E}^{st}(t|cons)$ with a positive sign, and a predictor of $\mathbb{E}^{st}(t|lib)$ with a negative sign. Table A4 shows that, conditional on true mean and our measure of likelihood of tail positions, $R_H^{cons}$ predicts believed mean for each group $G$ as predicted.

## G.1 Model Predictions

In this section, we apply the rank-based truncation model to make predictions for mean beliefs about liberals and conservatives in each data set. For each target group for each data set, we use the rank-based truncation model to make a prediction for the believed

Table A4: $R_H^{cons}$ Predicts Beliefs Controlling for Likelihood

| OLS Predicting Believed Mean of Group G | | | | | | |
|---|---|---|---|---|---|---|
| | G = Conservatives | | | G = Liberals | | |
| | GNH | ANES | Pooled | GNH | ANES | Pooled |
| True Mean of $G$ | 0.15 | 0.65 | -0.10 | 0.92**** | -0.13 | -0.30** |
| | (0.36) | (0.46) | (0.43) | (1.40) | (0.31) | (0.14) |
| $R_H^{cons}$ | 0.15* | 0.55** | 0.24*** | -0.12* | -0.17* | -0.27**** |
| | (0.07) | (0.23) | (0.07) | (0.06) | (0.08) | (0.03) |
| $ALTP_G$ | 6.68* | -5.60 | 7.41 | 2.26 | -5.13 | -9.77**** |
| | (3.74) | (6.39) | (5.10) | (2.87) | (5.05) | (1.88) |
| Constant | 2.17*** | 1.72 | 2.75**** | -0.17 | 4.73** | 6.25**** |
| | (0.70) | (1.00) | (0.67) | (1.40) | (2.03) | (0.82) |
| R-squared | 0.83 | 0.50 | 0.63 | 0.91 | 0.35 | 0.77 |
| Obs. (Clusters) | 45 (45) | 66 (10) | 111 (55) | 45 (45) | 66 (10) | 111 (55) |

Notes: Std. errors in parentheses, clustered at the issue level. *, **, ***, and **** denote significance at the 10% level, 5%, 1%, and 0.1% level, respectively. In pooled specifications, we include a dummy variable indicating whether the observation came from ANES data set.

mean position of the group. We do this for each value of $d$, truncating to only the most representative type ($d = 1$), then to the two most representative types ($d = 2$), and so on. We then compare the predictions of the model to the observed data. We report three measures of error: mean squared prediction error (MSPE), mean prediction error (MPE), and the fraction of observations for which the model underestimates the true belief. To compute MSPE for a given group and a given value of $d$, we subtract the model prediction for each observation from the observed data, square this difference, then take the mean of these squared differences. The MPE is computed similarly, but the differences are not squared. MSPE is a standard way of evaluating the magnitude of prediction errors, while MPE and our rate of underestimation speak to bias in the predictions.

Table A5, Panel (a) summarizes the results for the GNH dataset for conservatives. The model with $d = 4$ or $d = 5$ produces smaller MSPE than the accurate beliefs benchmark ($d = 6$). Using a signed rank test that compares the distribution of squared errors generated by our model to the distribution of squared errors generated by the accurate beliefs benchmark, we can reject the null of no difference between the models for $d = 4$ ($p<0.05$) and $d = 5$ ($p<0.001$). Furthermore, in both cases, the errors are less systematic. While 41 of the 45 true means are less than the observed beliefs (indicating consistent underestimation), errors are more evenly distributed across over and underestimation for most of the truncation models. We do the same exercise for liberals in the GNH data in Panel (b). In this case, $d = 5$ directionally outperforms the accurate beliefs benchmark, but no version of our model produces significantly smaller squared errors than the accurate beliefs benchmark.

Table A6 shows the MSPE, MPE, and rates of underestimation for the likelihood-based truncation model under different values of $d$. The likelihood-based truncation model truncates to the $d$ most likely types, rather than the $d$ most representative types. While the likelihood model produces smaller MSPE and MPE than the representativeness-based stereotype model for small values of $d$, for each group, the best representativeness-based model produces smaller MSPE errors than the best likelihood-based model. In terms of statistical significance, the likelihood model only significantly outperforms the representativeness-based model for conservatives for $d = 1$ ($p<0.001$), while the representativeness-based model is superior for conservatives for $d = 4$ ($p<0.01$) and $d = 5$ ($p<0.001$). The likelihood model

Table A5: Prediction Errors of Representativeness-Based Model for GSN Data

| Representativeness Model: Truncation to Most Representative Types | | | | | | |
|---|---|---|---|---|---|---|
| | $d=1$ | $d=2$ | $d=3$ | $d=4$ | $d=5$ | $d=6$ |
| (a) Predicting Believed Typical Mean of Conservatives in GNH Data | | | | | | |
| Mean Squared Prediction Error | 3.02 | 1.05 | 0.54 | 0.30 | 0.27 | 0.48 |
| Mean Prediction Error | -1.36 | -0.62 | -0.28 | 0.040 | 0.33 | 0.56 |
| Rate of Underestimation | 3/45 | 7/45 | 13/45 | 25/45 | 35/45 | 41/45 |
| N | 45 | 45 | 45 | 45 | 45 | 45 |
| (b) Predicting Believed Typical Mean of Liberals in GNH Data | | | | | | |
| Mean Squared Prediction Error | 2.47 | 1.42 | 0.68 | 0.20 | 0.073 | 0.083 |
| Mean Prediction Error | 0.94 | 0.79 | 0.57 | 0.27 | 0.071 | -0.093 |
| Rate of Underestimation | 39/45 | 41/45 | 42/45 | 42/45 | 26/45 | 17/45 |
| N | 45 | 45 | 45 | 45 | 45 | 45 |

performs better for liberals (significantly outperforming representativeness-based truncation for $d = 1, 2, 3$ ($p<0.0001$), but significantly underperforming the representativeness-based model for $d = 5$ ($p<0.10$).

While the best representativeness-based model is a directionally better predictor of observed beliefs than the accurate beliefs benchmark for both liberals and conservatives in terms of MSPE and MPE, the best likelihood-based model never beats the accurate beliefs benchmark in terms of MSPE. For conservatives, the best representativeness-based model ($d = 5$) outperforms the best likelihood-based model ($d = 5$) and the accurate beliefs benchmark for all metrics ($p<0.001$ for comparing squared errors of best rep. and likelihood, $p<0.001$ for comparing squared errors of best rep. and accurate beliefs benchmark). For liberals, the best representativeness-based model directionally outperforms the best likelihood-based model and the accurate beliefs benchmark in terms of MSPE, with more mixed results for MPE.

Figure A6 summarizes our evaluation of the two models for the GNH data, using the mean squared prediction error (MSPE) as a measure of the magnitude of errors.

Tables A7 and A8 repeat this exercise for the ANES data.

Table A6: Prediction Errors of Likelihood-Based Model for GSN Data

| Likelihood Model: Truncation to Most Likely Types | | | | | | |
|---|---|---|---|---|---|---|
| | $d=1$ | $d=2$ | $d=3$ | $d=4$ | $d=5$ | $d=6$ |
| (a) Predicting Believed Typical Mean of Conservatives in GNH Data | | | | | | |
| Mean Squared Prediction Error | 1.28 | 1.09 | 0.85 | 0.62 | 0.56 | 0.48 |
| Mean Prediction Error | 0.58 | 0.58 | 0.53 | 0.54 | 0.57 | 0.56 |
| Rate of Underestimation | 35/45 | 35/45 | 31/45 | 38/45 | 39/45 | 41/45 |
| N | 45 | 45 | 45 | 45 | 45 | 45 |
| (b) Predicting Believed Typical Mean of Liberals in GNH Data | | | | | | |
| Mean Squared Prediction Error | 1.17 | 0.61 | 0.31 | 0.18 | 0.12 | 0.083 |
| Mean Prediction Error | 0.52 | 0.37 | 0.21 | 0.077 | -0.018 | -0.093 |
| Rate of Underestimation | 32/45 | 32/45 | 32/45 | 29/45 | 24/45 | 17/45 |
| N | 45 | 45 | 45 | 45 | 45 | 45 |

Table A7: Prediction Errors of Representativeness-Based Model for ANES Data

| Representativeness Model: Truncation to Most Representative Types | | | | | | | |
|---|---|---|---|---|---|---|---|
| | $d=1$ | $d=2$ | $d=3$ | $d=4$ | $d=5$ | $d=6$ | $d=7$ |
| (a) Predicting Believed Typical Mean of Conservatives in ANES Data | | | | | | | |
| Mean Squared Prediction Error | 2.02 | 1.84 | 1.30 | 0.76 | 0.57 | 0.52 | 0.46 |
| Mean Prediction Error | -1.25 | -1.14 | -0.89 | -0.59 | -0.40 | -0.21 | 0.058 |
| Rate of Underestimation | 3/66 | 6/66 | 7/66 | 8/66 | 15/66 | 23/66 | 38/66 |
| N | 66 | 66 | 66 | 66 | 66 | 66 | 66 |
| (b) Predicting Believed Typical Mean of Liberals in ANES Data | | | | | | | |
| Mean Squared Prediction Error | 4.81 | 2.95 | 1.39 | 0.38 | 0.28 | 0.43 | 0.63 |
| Mean Prediction Error | 1.76 | 1.65 | 1.07 | 0.44 | 0.013 | -0.30 | -0.53 |
| Rate of Underestimation | 63/66 | 65/66 | 64/66 | 55/66 | 31/66 | 22/66 | 13/66 |
| N | 66 | 66 | 66 | 66 | 66 | 66 | 66 |

Table A8: Prediction Errors of Likelihood-Based Model for ANES Data

| Likelihood Model: Truncation to Most Likely Types | | | | | | | |
|---|---|---|---|---|---|---|---|
| | $d=1$ | $d=2$ | $d=3$ | $d=4$ | $d=5$ | $d=6$ | $d=7$ |
| (a) Predicting Believed Typical Mean of Conservatives in ANES Data | | | | | | | |
| Mean Squared Prediction Error | 2.82 | 1.45 | 1.14 | 0.97 | 0.70 | 0.53 | 0.46 |
| Mean Prediction Error | -0.068 | -0.22 | -0.20 | -0.20 | -0.073 | 0.002 | 0.058 |
| Rate of Underestimation | 38/66 | 29/66 | 23/66 | 25/66 | 26/66 | 31/66 | 38/66 |
| N | 66 | 66 | 66 | 66 | 66 | 66 | 66 |
| (b) Predicting Believed Typical Mean of Liberals in ANES Data | | | | | | | |
| Mean Squared Prediction Error | 2.67 | 1.35 | 0.94 | 0.88 | 0.78 | 0.71 | 0.63 |
| Mean Prediction Error | -0.13 | -0.30 | -0.34 | -0.28 | -0.37 | -0.46 | -0.53 |
| Rate of Underestimation | 20/66 | 26/66 | 23/66 | 28/66 | 23/66 | 19/66 | 13/66 |
| N | 66 | 66 | 66 | 66 | 66 | 66 | 66 |

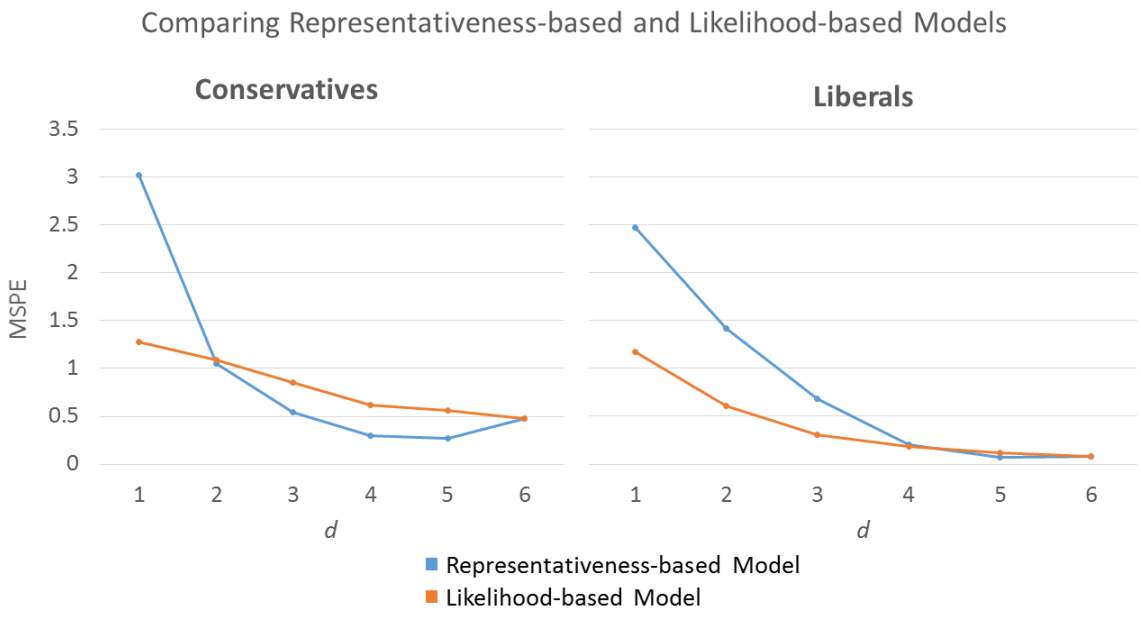



Figure A6: Comparing the Representativeness-Based and Likelihood-Based Models

## G.2 Beliefs of Conservatives and Liberals

In this section, we show that the predictions of our model hold both for beliefs held by Conservatives and beliefs held by Liberals. First, we document exaggeration in Table A9. In the GNH data, Liberals hold more exaggerated beliefs about both Conservatives and Liberals than Conservatives do. The pattern is different in the ANES data. Conservatives in the ANES data have exaggerated beliefs about Liberals, but not about Conservatives. Liberals in the ANES data have exaggerated beliefs about both Liberals and Conservatives, with more exaggerated beliefs about Liberals than Conservatives. Given the differences across the two data sets, it is hard to draw general conclusions about whether beliefs are more exaggerated when predicting positions of the other group. In most cases, for both Liberals and Conservatives, reported beliefs are more extreme than the truth for both their own group and the other group.

Next, we test the prediction of Equation 5, asking whether we observe the same context-dependence for beliefs held by either group. In Table A10, we predict the believed mean of a group G from the true mean of the group G and the true mean of -G. Our model predicts that information about -G will be predictive of believed mean of G. The key here

Table A9: Information about -G Predicts Beliefs about G, Conservatives versus Liberals

| Exaggeration of Beliefs about G | | | | |
|---|---|---|---|---|
| | G = Conservatives | | G = Liberals | |
| | Held by Conservatives | Held by Liberals | Held by Conservatives | Held by Liberals |
| GNH | 0.35 | 0.71 | 0.03 | 0.21 |
| ANES | -0.11 | 0.18 | 0.78 | 0.36 |

Table A10: Information about -G Predicts Beliefs about G, Conservatives versus Liberals

| OLS Predicting Believed Mean of G in Pooled Data | | | | |
|---|---|---|---|---|
| | G = Conservatives | | G = Liberals | |
| | Held by Conservatives | Held by Liberals | Held by Conservatives | Held by Liberals |
| True Mean Conservatives | 1.13**** | 0.92**** | -0.36**** | -0.23**** |
| | (0.076) | (0.087) | (0.073) | (0.065) |
| True Mean Liberals | -0.55**** | -0.65**** | 0.68**** | 0.81**** |
| | (0.118) | (0.149) | (0.140) | (0.147) |
| Constant | 1.46**** | 2.87**** | 2.14**** | 1.16**** |
| | (0.279) | (0.305) | (0.260) | (0.282) |
| R-squared | 0.77 | 0.57 | 0.46 | 0.76 |
| Obs. (Clusters) | 111 (55) | 111 (55) | 111 (55) | 111 (55) |

Notes: Std. errors in parentheses, clustered at the issue level. *, **, ***, and **** denote significance at the 10% level, 5%, 1%, and 0.1% level, respectively. In pooled specifications, we include a dummy variable indicating whether the observation came from ANES data set.

is whether this prediction holds independent of whether we are considering beliefs about G held by Conservatives or Liberals. Thus, we present two specifications side-by-side, one predicting beliefs held by Conservatives about a group G, and one predicting beliefs held by Liberals of that same group G. We see quite similar results when we explore beliefs held by Conservatives and beliefs held by Liberals. In particular, both sets of beliefs demonstrate the same strong evidence for context-dependence that we documented in the main text.

In Table A11, we test the predictions of Equations 6 and 7, asking whether $R_H^{cons}$ also has predictive power for beliefs held by Conservatives and Liberals. Again, we see that the results do not strongly depend on who holds the beliefs. In predicting the Conservatives' belief of the mean Conservative position or the Liberals' belief of the mean Conservative

Table A11: Average Representativeness of Tail Positions Predicts Beliefs, Conservatives versus Liberals

| OLS Predicting Believed Mean of G in Pooled Data | | | | |
|---|---|---|---|---|
| | G = Conservatives | | G = Liberals | |
| | Held by Conservatives | Held by Liberals | Held by Conservatives | Held by Liberals |
| True Mean of $G$ | 0.71**** | 0.42**** | 0.36**** | 0.58**** |
| | (0.09) | (0.10) | (0.10) | (0.10) |
| ${R_H^{cons}}_G$ | 0.20** | 0.30**** | -0.12 | -0.07 |
| | (0.09) | (0.09) | (0.10) | (0.07) |
| Constant | 1.03**** | 2.24**** | 1.99**** | 1.10**** |
| | (0.30) | (0.32) | (0.24) | (0.27) |
| R-squared | 0.72 | 0.50 | 0.32 | 0.75 |
| Obs. (Clusters) | 111 (55) | 111 (55) | 110 (54) | 110 (54) |

Notes: Std. errors in parentheses, clustered at the issue level. *, **, ***, and **** denote significance at the 10% level, 5%, 1%, and 0.1% level, respectively. In pooled specifications, we include a dummy variable indicating whether the observation came from ANES data set. One liberal observation is missing from the GNH data as there is no mass on stereotypical liberal positions for either group for one issue.

position, the average representativeness of tail positions has predictive power. Similarly, in predicting the Conservatives' belief of the mean Liberal position or the Liberals' belief of the mean Liberal position, the average representativeness of Liberal tail positions has a negative, but insignificant on beliefs.

**References for Online Appendix:**


Couch, James, and Jennifer Sigler. 2001. "Gender Perception of Professional Occupations." *Psychological Reports* 88 (3): 693 – 698.

Decker, Wayne. 1986. "Occupation and Impressions: Stereotypes of Males and Females in Three Professions." *Social Behavior and Personality* 14 (1): 69–75.

Hewstone, Miles, Manfred Hassebrauck, Andrea Wirth, and Michaela Waenke. 2000. "Pattern of Disconfirming Information and Processing Instructions as Determinants of Stereotype Change." *British Journal of Social Psychology* 39: 399 – 411.

Kahneman, Daniel, and Shane Frederick. 2005. "A Model of Heuristic Judgment," in *The Cambridge Handbook of Thinking and Reasoning*, Keith Holyoak and Robert Morrison, eds. Cambridge, UK: Cambridge University Press.

Lord, Charles, Lee Ross, and Mark Lepper. 1979. "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology* 37 (11): 2098 – 2109.

Madon, Stephanie, Max Guyll, Kathy Aboufadel, Eulices Montiel, Alison Smith, Polly Palumbo, and Lee Jussim. 2001. "Ethnic and National Stereotypes: The Princeton Trilogy Revisited and Revised." *Personality and Social Psychological Bulletin* 27 (8): 996 – 1010.

Nickerson, Raymond. 1998. "Confirmation Bias: a Ubiquitous Phenomenon in Many Guises." *Review of General Psychology* 2 (2): 175 – 220.

Noori, Kamyar, and Allyson Weseley. 2011. "Beyond Credentials: The Effect of Physician Sex and Specialty on How Physicians Are Perceived." *Current Psychology* 30: 275 – 283.

Rothbart, Myron. 1981. "Memory Processes and Social Beliefs." In Cognitive Processes in Stereotyping and Intergroup Behavior, ed. DL Hamilton, Hillsdale, NJ: Erlbaum: 145 – 81.

Schneider, David. 2004. *The Psychology of Stereotyping.* New York, NY: The Guilford Press.

Schwartzstein, Joshua. 2014. "Selective Attention and Learning." *Journal of the European Economic Association* 12 (6): 1423 – 1452.

Weber, Renée, and Jennifer Crocker. 1983. "Cognitive Processes in the Revision of Stereotypic Beliefs." *Journal of Personality and Social Psychology* 45 (5): 961 – 977.