STEREOTYPES

Pedro Bordalo

Katherine Coffman

Nicola Gennaioli

Andrei Shleifer*

November 21, 2015

Abstract

We present a model of stereotypes in which a decision maker assessing a group recalls only that group's most representative or distinctive types. Stereotypes highlight differences between groups, and are especially inaccurate (consisting of unlikely, extreme types) when groups are similar. Stereotypical thinking implies overreaction to information that generates or confirms a stereotype, and underreaction to information that contradicts it. Stereotypes can change if new information changes the group's most distinctive trait. We present experimental evidence on the role of representativeness in shaping subjects' mental representation of groups. We also evaluate the predictions of the model using large scale survey evidence on beliefs about political groups. In line with our predictions, beliefs are distorted in the direction of the groups' most representative types.

*Royal Holloway University of London, Ohio State University, Universitá Bocconi and IGIER, Harvard University. We are grateful to Nick Barberis, Roland Bénabou, Dan Benjamin, Tom Cunningham, Matthew Gentzkow, Emir Kamenica, Larry Katz, David Laibson, Sendhil Mullainathan, Josh Schwartzstein, Jesse Shapiro, Alp Simsek, Neil Thakral, and Maik Wehmeyer for extremely helpful comments. We thank the Initiative on Foundations of Human Behavior for support of this research. Corresponding Author: Andrei Shleifer, email: ashleifer@harvard.edu, tel: 6174955046, Littauer Center, 1805 Cambridge Street, Cambridge, MA 02138. JEL: D03, D83, D84.

1 Introduction

The Oxford English Dictionary defines a stereotype as a "widely held but fixed and oversimplified image or idea of a particular type of person or thing". Stereotypes are ubiquitous. Among other things, they cover racial groups ("Asians are good at math"), political groups ("Republicans are rich"), genders ("Women are bad at math"), demographic groups ("Florida residents are elderly"), and situations ("Tel-Aviv is dangerous"). As these and other examples illustrate, some stereotypes are roughly accurate ("the Dutch are tall"), while others much less so ("Irish are red-headed"; only 10% are). Moreover, stereotypes change: in the US, Jews were stereotyped as religious and uneducated at the beginning of the 20th century, and as high achievers at the beginning of the 21st (Madon et. al., 2001).

Social science has produced three broad approaches to stereotypes. The economic approach of Phelps (1972) and Arrow (1973) sees stereotypes as a manifestation of statistical discrimination: rational formation of beliefs about a group member in terms of the aggregate beliefs about that group. Statistical discrimination may impact actual group characteristics in equilibrium (Arrow 1973). For example, if employers hold adverse beliefs about the skills of black workers, blacks would underinvest in education, thereby fulfilling the adverse prior beliefs. Because in this theory stereotypes are based on rational expectations, it does not address a central problem that stereotypes are often inaccurate. The vast majority of Florida residents are not elderly, the vast majority of the Irish are not red-headed, and Tel-Aviv is really pretty safe.

The sociological approach to stereotyping pertains only to social groups. It views stereotypes as fundamentally incorrect and derogatory generalizations of group traits, reflective of the stereotyper's underlying prejudices (Adorno et al. 1950) or other internal motivations (Schneider 2004). Social groups that have been historically mistreated, such as racial and ethnic minorities, continue to suffer through bad stereotyping, perhaps because the groups in power want to perpetuate false beliefs about them (Steele 2010, Glaeser 2005). The stereotypes against blacks are thus rooted in the history of slavery and continuing discrimination. This approach might be relevant in some important instances, but it leaves a lot out. While some stereotypes are inaccurate, many are quite fair ("Dutch are tall," "Swedes are blond.") Moreover, many stereotypes are flattering to the group in question rather than pejorative ("Asians are good at math"). Finally, stereotypes change, so they are at least in part responsive to reality rather than entirely rooted in the past (Madon et. al., 2001)

The third approach to stereotypes – and the one we follow – is the "social cognition approach"(Schneider 2004). This approach gained ground in the 1980s and views social stereotypes as special cases of cognitive schemas or theories (Schneider, Hastorf, and Ellsworth 1979). These theories are intuitive generalizations that individuals routinely use in their everyday life, and entail savings on cognitive resources.[1] Hilton and Hippel (1996) stress that stereotypes are "mental representations of real differences between groups [. . .] allowing easier and more efficient processing of information. Stereotypes are selective, however, in that they are localized around group features that are the most distinctive, that provide the greatest differentiation between groups, and that show the least within-group variation." A related "kernel-of-truth hypothesis" holds that stereotypes are based on some empirical reality. As such, they are useful, but may entail exaggerations (Judd and Park 1993).

We show that this approach to stereotypes is intimately related to another idea from psychology: the use of heuristics in probability judgments (Kahneman and Tversky 1972). Just as heuristics simplify the assessment of complex probabilistic hypotheses, they also simplify the representation of heterogeneous groups. In this way, heuristics enable a quick and often reliable assessment of complex situations, but sometimes cause biases in judgments. Consider in particular the representativeness heurstic. Tversky and Kahneman (1983) write that "an attribute is representative of a class if it is very diagnostic; that is, the relative frequency of this attribute is much higher in that class than in the relevant reference class." Representativeness suggests that the reason people stereotype the Irish as red-headed is that red hair is more common among the Irish than among other groups, even though it is not that common in absolute terms. The reason people stereotype Republicans as wealthy is

[1] In the words of Lippmann (1922, pp.88-89), an early precursor of this approach: "There is economy in this. For the attempt to see all things freshly and in detail, rather than as types and generalities, is exhausting, and among busy affairs practically out of the question[. . .]. But modern life is hurried and multifarious, above all physical distance separates men who are often in vital contact with each other, such as employer and employee, official and voter. There is neither time nor opportunity for intimate acquaintance. Instead we notice a trait which marks a well-known type, and fill in the rest of the picture by means of the stereotypes we carry about in our heads."

that the wealthy are more common among Republicans than Democrats.[2] In both cases, the representation entails judgment errors: people overestimate the proportion of red-haired among the Irish, or of the wealthy among the Republicans. Representativeness thus generates stereotypes that differentiate groups along existing and highly diagnostic characteristics, exactly as Hilton, Hippel and Schneider define them.[3] While representativeness is not the only heuristic that shapes recall (availability, driven by recency or frequency of exposure, also plays a role), it is the key driving force of stereotypes which, in line with the social psychology perspective, are centered on *differences* among groups.[4]

In this paper, we systematically explore the connection between the representativeness heuristic and the social psychology view of stereotypes as intuitive generalizations. Formally, we assume that a type t is representative for group G if it is diagnostic of G relative to a comparison group $-G$, in that the diagnostic ratio $\Pr(G|t)/\Pr(-G|t)$ is high. Equivalently, a representative type for group G has a high likelihood ratio:

$$\frac{\Pr(t|G)}{\Pr(t|-G)}. \tag{1}$$

Due to limited working memory, the most representative types come to mind first and are overweighted in judgments. Predictions about G are then made under a distorted distribution, or stereotype, that overweights representative types. Our results obtain with minimal assumptions on such overweighting. We also describe a number of weighting specifications, and explore their properties.

The critical feature of our approach is that representativeness, and stereotypes, can only exist in context, that is, relative to a comparison group $-G$. This implies that, as the comparison group changes, so do representativeness, stereotypes, and assessments. In Section 2, as a motivation for our analysis, we present experimental evidence supportive of this key

[2] See www.nytimes.com/packages/pdf/politics/20041107_px_ELECTORATE.xls.

[3] Deaux and Kite (1985) stress that the features that distinguish a category from a comparison category are especially useful as identifying characteristics. According to Schneider (2004 p. 91), the stereotype for a category should have "membership diagnosticity": "all females have hearts (feature diagnosticity), but not all people who have hearts are female (membership diagnosticity). Similarly, membership diagnosticity can be nearly perfect, but feature diagnosticity may still be quite low; people who nurse babies are female, but far from all females are nursing at any given time[...] Hearts won't do the job for femaleness, but possession of a uterus works."

[4] See Section 3.2 and Appendix C for an in depth discussion of these issues.

prediction. We construct a group of mundane objects, G, and present it to participants next to a comparison group, $-G$. In our baseline condition, the comparison group is chosen so that no type is particularly representative of group G. In our treatment, we change the comparison group, $-G$, while leaving the target group, G, unchanged. The new comparison group gives rise to highly representative types within G. In line with the key prediction of our model, participants in the treatment condition shift their assessment of G toward the new representative types.

We next turn to the analysis of the model. To give a preview of some of our results, we find that representativeness often generates fairly accurate stereotypes but sometimes causes stereotypes to be inaccurate, particularly when groups have similar distributions that differ most in unlikely types. To illustrate this logic, consider the formation of the stereotype "Florida residents are elderly". The proportion of elderly people in Florida and in the overall US population is shown in the table below.[5]

age	$0-18$	$19-44$	$45-64$	65+
Florida	23.9%	31.6%	27.0%	17.3%
US	26.6%	33.4%	26.5%	13.5%

The table shows that the age distributions in Florida and in the rest of the US are very similar. Yet, someone over 65 is highly representative of a Florida resident, because this age bracket maximizes the likelihood ratio $\Pr(t|\text{Florida})/\Pr(t|\text{US})$.[6] When thinking about the age of Floridians, then, the "65+" type immediately comes to mind because in this age bracket Florida is most different from the rest of the US, in the precise sense of representativeness. Representativeness-based recall induces an observer to overweight the "65+" type in his assessment of the average age of Floridians.

Critically, though, this stereotype is inaccurate. Indeed, and perhaps surprisingly, only about 17% of Florida residents are elderly. The largest share of Florida residents, nearly as many as in the overall US population, are in the age bracket "19-44", which maximizes $\Pr(t|\text{Florida})$. Being elderly is not the most likely age bracket for Florida residents, but

[5] See http://quickfacts.census.gov/qfd/states/12000.html.

[6] In this problem, the likelihood ratio in (1) is $\Pr(t|\text{Florida})/\Pr(t|\text{rest of US})$, but it is easy to see that t maximizes $\Pr(t|\text{Florida})/\Pr(t|\text{rest of US})$ if and only if it maximizes $\Pr(t|\text{Florida})/\Pr(t|\text{US})$.

rather the age bracket that occurs with the highest relative frequency. A stereotype-based prediction that a Florida resident is elderly has very little validity.

Besides offering guidance on the circumstances in which stereotypes are more or less accurate, our model makes several predictions. Most importantly:

- Stereotypes amplify systematic differences between groups, even if these differences are in reality very small. When groups differ by a shift in means, stereotyping exaggerates differences in means, and when groups differ by a increase in variance, stereotyping exaggerates the differences in variances. In these cases (though not always), representativeness yields stereotypes that contain a "kernel of truth".

- Stereotypes are context dependent. The assessment of a given target group depends on the group to which it is compared. For instance, when comparing Irish to Scots, the stereotype of Irish may change from "red-haired" to "Catholic".

We bring these predictions to the data. In line with the social cognition approach to stereotypes, a significant body of psychological research on beliefs about gender, race, age and political groups finds that stereotypes broadly reflect reality but also display biases (for a review, see Jussim et al. 2015). Recent work highlights the fact that beliefs exaggerate group differences, particularly in the context of stereotypes about age and political groups (Chan et al 2012, Westfall et al 2015). This descriptive work provides the backdrop for our own empirical investigation, which specifically tests the two above predictions.

We use two data sets on political preferences, and beliefs about political preferences, in the U.S. Here, groups are political constituencies (Democrats and Republicans) and types are their positions on several issues. We first show that beliefs depart from the truth by exaggerating (mean) differences, as per the kernel of truth logic. We then show that distortions in beliefs can be accounted for by overweighting types that are representative of each political group, in light of the other group. This provides evidence of context dependence that complements that of Section 2.[7]

[7] We document a number of other properties of the model in the Online Appendix. Most importantly, we explore how stereotypical thinking distorts reaction to information. So long as stereotypes do not change, people under-react or even ignore information inconsistent with stereotypes. If enough contrary information is received, stereotypes change, leading to a drastic reevaluation of already available data. Representativeness-

Although we have argued that stereotypes, like heuristics, allow for quick and often useful assessments, they are not always benign. Some of the errors caused by inaccurate stereotypes are inconsequential. A driver being cut off on the road might form a quick gender or age stereotype of the aggressor, but then quickly drive on and forget about it. But stereotypical thinking can also have substantial consequences. One instance, discussed in Section 4.2, concerns the role of gender stereotypes in mathematics or occupational choice (Buser et al (2014)). Similarly, graduate admission officers scanning dozens of files might reject foreign candidates who bring to mind ethnic stereotypes and accept potentially less talented candidates with A's from Ivy League schools. We do not suggest that decision makers are uniformly bound to stereotypical thinking in all situations; rather that it requires substantial cost and deliberation to enrich one's mental representations, and even deliberation may not fully overcome the influence of stereotypes.

Since Kahneman and Tversky's (1972, 1973) work on heuristics and biases, several studies have formally modelled heuristics about probabilistic judgments and incorporated them into economic models. Work on the confirmation bias (Rabin and Schrag 1999) and on probabilistic extrapolation (Grether 1980, Barberis, Shleifer, and Vishny 1998, Rabin 2002, Rabin and Vayanos 2010, Benjamin, Rabin and Raymond 2011) assumes that the decision maker has an incorrect model in mind or incorrectly processes available data. Our approach is instead based on the central assumption that representative information comes foremost to mind when making judgments. The specific mental operation that lies at the heart of our model – namely, generating a prediction for the distribution of types in a group, based on data stored in memory – also captures base-rate neglect and confirmation bias. The underweighting or neglect of information in our model simplifies judgment problems in a way related to models of categorization (Mullainathan 2002, Fryer and Jackson 2008). In these models, however, decision makers use coarse categories organized according to likelihood, not representativeness. This approach generates imprecision but does not create a systematic bias for overestimating unlikely events, nor does it allow for a role of context in shaping assessments. In our empirical analysis of political beliefs, we explicitly compare the

based recall reconciles under-reaction with over-reaction to data, generating both confirmation bias and base-rate neglect.

predictions of representativeness-based recall to those of likelihood based models, and find that the evidence supports the former.

In modeling representativeness we follow the specification of Gennaioli and Shleifer (GS, 2010), but investigate a new set of questions. GS (2010) examine how representativeness distorts the assessed probabilities of alternative hypotheses, but not how the probability of a given hypothesis or group is distributed across its constituent elements. In the context of the current setting, GS (2010) ask how imperfect recall affects the assessed probability that a randomly drawn member from a universe Ω belongs to group G. The current paper, in contrast, asks which type t we expect to draw *once we know* that we are facing group G.[8] GS (2010) show how representativeness generates biased probabilistic assessments such as conjunction and disjunction fallacies. The current paper deals with perhaps a broader and more ubiquitous problem of stereotype formation, extensively studied by other social scientists but largely neglected by economists.

In the next section, we present suggestive experimental evidence on the role of representativeness in recall-based judgments. Section 3 describes our model. In Section 4 we examine the properties of stereotypes, including the forces that shape stereotype accuracy, and illustrate these properties with a number of examples. In Section 5 we show how the model can be brought to the data, and conduct an empirical analysis of political beliefs. Section 6 concludes. Appendix A contains the proofs.

We present a number of extensions and other results in the Online Appendix. Appendix B considers the case of unordered types, while Appendix C extends the model to account for the role of likelihood and availability in recall. In Appendix E we extend the analysis to the cases where types are continuous, and in Appendix F we describe how stereotypes can cause both under- and over-reaction to new information. Finally, in Appendix G and H we present the full details and analysis of our experiments and empirical analysis.

[8] Specifically, in GS (2010) the assessed probability that a certain hypothesis G is true is equal to:

$$\Pr(G) = \frac{\sum_{r \leq d} \pi_{t_r,G}}{\sum_{r \leq d} \pi_{t_r,G} + \sum_{r' \leq d} \pi_{t'_r,\Omega\backslash G}}$$

where t_r are the d most representative types for hypothesis G and $t_{r'}$ are the d most representative types for the alternative hypothesis $-G = \Omega \backslash G$.

2 Motivating Evidence on Group Assessment

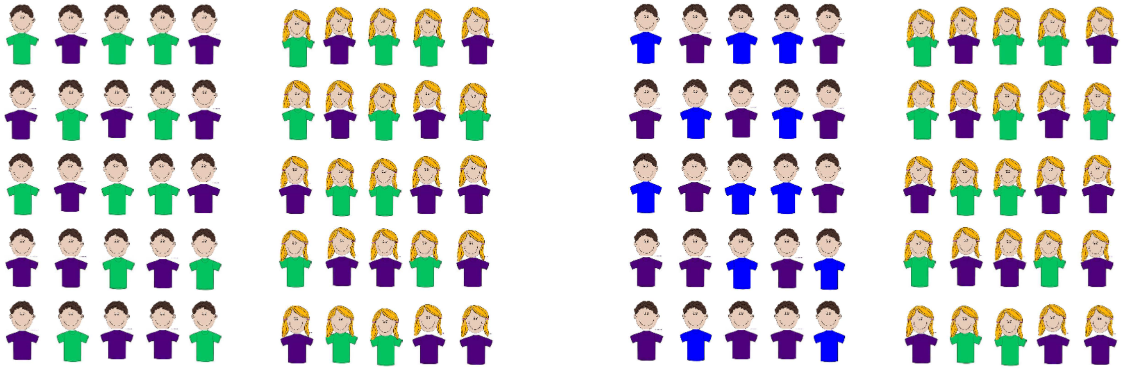
The assumption of representativeness-based recall implies that assessments of groups are made in contrast to, and emphasize differences with, comparison groups. Assessments are therefore context dependent, in the sense that judgements about a group depend on the features of the group it is compared to. We assess this prediction in a controlled laboratory environment. While field evidence on widely-held stereotypes is suggestive, the laboratory setting allows us to isolate the role of representativeness, abstracting from many other factors – historical, sociological, or otherwise – that may also play a role in stereotype formation. We construct our own groups of ordinary objects, creating a target group, G, and a comparison group, $-G$. We explore how participant impressions of G change as we vary the representativeness of different types within this target group simply by changing the comparison group.

We conducted several experiments, in the laboratory as well as on Amazon Mechanical Turk. Each involves a basic three-step design. First, participants are shown the target group and a randomly-assigned comparison group for 15 seconds. In this time frame, differences across the groups can be noticed but the groups' precise compositions cannot be memorized. The second step consists of a few filler questions, that briefly draw the participants' cognitive bandwidth away from their observation. Finally, participants are asked to recall the groups they saw, and assess them in various ways. Participants are incentivized to provide accurate answers.

We randomly assign participants to either the Control or the Representativeness condition. In the Control condition, G and $-G$ have nearly identical distributions, so that all types are equally representative for each group. In the Representativeness condition, G is unchanged, while the composition of the comparison group $-G$ is changed in such a way that a certain type becomes very representative for G. Context dependence implies that the assessment of G should now overweight this representative type, even though the distribution of G itself has not changed.

We ran six experiments of this form, with design changes focused on reducing participant confusion and removing confounds. Here, we describe the final, and most refined, versions of these experiments. In an attempt to provide a overview of the results while remaining concise, we also provide the results from pooled specifications that use all data collected. In Appendix G, we present additional details and report all experiments conducted. We also provide instructions and materials for each experiment and the full data set.

Consider first the experiment illustrated in Figure 1. A group of 25 cartoon girls is presented next to a group of 25 cartoon boys in t-shirts of different colors: blue, green, or purple. In the Control condition, Fig.1a, the groups have identical color distributions (13 purple, 12 green), so no color is representative of either group. The Representativeness condition, Fig.1b, compares the *same* group of girls with a different group of boys, for whom green shirts are replaced by blue shirts. Now only girls wear green and only boys wear blue. These colors, while still not the most frequent for either group, are now most representative. For each group, girls and boys, participants are asked to identify the modal color shirt worn by that group. Our key prediction is that when the less frequent color (the 12-shirt color) is representative for a group, participants will be more likely to believe it to be modal for that group. Note that the only factor that varies across treatments is the representativeness of the 12-color shirt. Thus, if we see differences across conditions, the causal role of representativeness-based recall in shaping group judgments is clear.[9]



(a) Control Condition

(b) Representativeness Condition

Figure 1: T-shirts Experiment

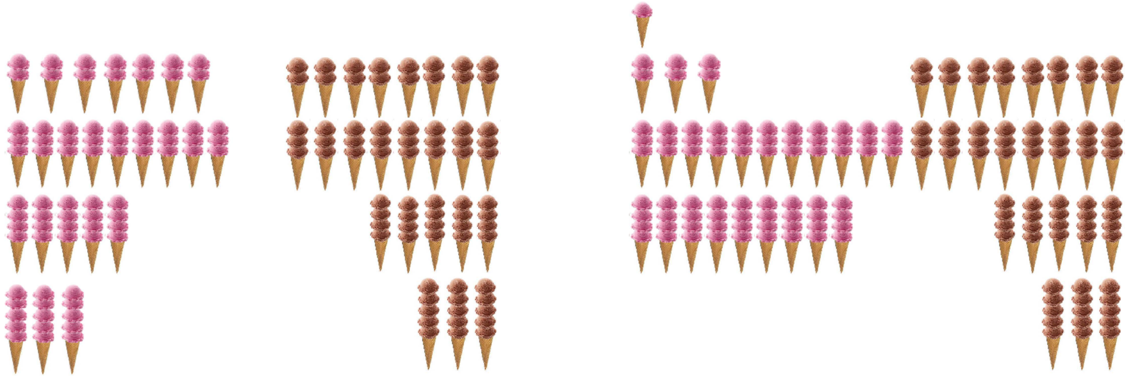
[9] Note that we vary which colors are used in which roles across participants – that is, some participants saw this particular color distribution, while others see, for example, green as the modal color, with purple as the diagnostic color for boys in the Rep. condition and blue as the diagnostic color for girls in the Rep. condition. We vary the colors across the roles to avoid confounding the characteristics of any particular color with its diagnosticity.

We collected data from 301 participants using this T-shirts design.[10] Consistent with the role of representativeness, participants assigned to the Representativeness condition are 10.5 percentage points more likely to recall the less frequent color (blue or green) as the modal color when it is representative of a group (35% of participants guess the less frequent color is modal in the Control condition, this proportion increases to 46% in the Representativeness condition, p=0.01, estimated from a probit regression reported in Appendix G). We also ask participants to estimate how many of each color T-shirts they saw in each group. In both treatments, the true difference in counts is one (13 purple shirts, 12 green or blue shirts). In the Control condition, participants on average believe they saw 0.54 more purple shirts than green or blue shirts, while in the Representativeness condition, participants believe they saw 0.72 *fewer* purple shirts than green or blue shirts (across treatment difference is significant with p=0.013 from two-tailed Fisher Pitman permutation test).

The next experiment, illustrated in Figure 2, shows that the effect survives even when the representative types are much less frequent, or in the tails. 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.2a, 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.2b, 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 collected data from 223 participants using this design. Participants are asked which flavor has more scoops on average. We use this binary question, rather than a more complex elicitation of their recollection of entire distributions, because it is simple, easy to incentivize, and should move with participants' impressions of the distributions. In reality, strawberry

[10] Throughout our analysis, we exclude any participant who participated in a previous version of the experiment and any participant who self-identified as color blind. In Appendix G, we show that our results are unchanged if we include these additional observations.



(a) Control Condition (b) Representativeness Condition

Figure 2: Ice cream cone Experiment

has a slightly higher average in both conditions (3.167 for strawberry versus 3.125 for chocolate). Yet, and consistent with context dependence, we see a shift in the prediction direction across condition. In the Control condition, 44% of participants incorrectly believe that the chocolate scoops have more scoops on average; in the Representativeness condition, this mistake is made by 54% of participants (p=0.16 estimated from a two-tailed test of proportions). We also ask participants (i) to estimate the average number of scoops in each group and (ii) to make a choice between the two lotteries induced by the distributions of cones. We find no significant differences across conditions for these questions.[11]

In total, we collected data for six experiments of this general structure, gathering evidence from more than 1,000 participants. As we describe in Appendix G, while there is substantial variation across experiments, when we pool all data collected we find significant aggregate treatment effects in line with a role of representativeness in judgment (see Appendix G).[12] Using a probit regression that pools all of the data for unordered type experiments similar to the T-shirts experiment (four versions, 741 participants), we find that participants are 9.3 percentage points more likely to guess that the less frequent type is modal when it is representative than when it is not (p=0.002). For ordered type experiments similar to ice cream cones (two versions, 402 participants), participants are 11.5 percentage points more

[11] We describe these two questions in more detail in Appendix G. We conjecture that participant risk preferences may have overwhelmed the role of representativeness-based recall in the lottery choice. We discuss this interpretation of this result, and provide supporting evidence, in Appendix G.

[12] We find effects in the predicted directions for all six designs, with significant effects for two. We discuss the extent to which our results are sensitive to the specifics of the design, in Appendix G.

likely to guess that the group of interest has a greater average than the comparison group when the right tail is representative (p=0.026).[13] Given our simple experimental setting with groups of mundane objects, we interpret our results – a significant and reasonably-sized impact on average beliefs – as an important proof of concept: the presence of representative types biases ex post assessment.

We next present the model of representativeness-based judgments that features the type of context dependence observed in our experiments.

3 A Model of Representativeness and Stereotypes

3.1 The Model

A decision maker (DM) faces a *prediction* problem, such as assessing the ability of a job candidate coming from a certain ethnic group, the future performance of a firm belonging to a certain sector, or future earnings based on own gender. Formally, the DM must assess the distribution of a categorical random variable T (ability, earnings) in a group G (social group, industrial sector), which is a proper subset of the entire population Ω. The random variable T takes values in a type space $\{t_1, \ldots, t_N\}$ that is naturally ordered, with $t_1 < \ldots < t_N$, and in many examples is assumed to be cardinal.[14] We denote by $\pi_{t,G}$ the true conditional probability $\Pr(T = t|G)$ of type t in group G and by π_t the true unconditional probability $\Pr(T = t)$ of type t in Ω.

The DM has stored in memory the full conditional distribution $(\pi_{t,G})_{t\in\{t_1,\ldots,t_N\}}$, but he retrieves from memory a simplified version of this distribution, which overweights the probability of those types that are most representative of G relative to an alternative group $-G$. Definitions 1 and 2 formalize this representativeness-based recall, following GS (2010).

[13] Results for the ordered types experiments, unlike the simpler T-shirts style design, were sensitive to the choice of platform, with consistently strong results on Amazon Mechanical Turk and weak or null results in laboratory samples. We discuss this in Appendix G.

[14] The model applies also to cases in which types: i) are not ordered, representing for instance occupations, ii) are multi-dimensional, capturing a bundle of attributes such as occupation and nationality, or iii) are continuous, such as when t follows a conditionally normal distribution. We consider these possibilities formally in Appendix B, D and Erespectively. Also, G may represent any category of interest, such as the historical performance of a firm or industry, actions available to a decision maker (T = set of payoffs, G = occupations), or categories in the natural world (T = ability to fly, G = birds).

Definition 1 *The representativeness of type t for group G is defined as the likelihood ratio:*

$$R(t,G) = \frac{\Pr(T=t|G)}{\Pr(T=t|-G)} = \frac{\pi_{t,G}}{\pi_{t,-G}}. \tag{2}$$

where $-G = \Omega \backslash G$.

In line with Tversky and Kahneman's (1983) definition, a type t is representative of G if it is relatively more likely to occur in G than in $-G$. The representative age of a Floridian is 65+ because people in this age bracket are more common in Florida as compared to the rest of the US. Definition 1 implies that DMs are attuned to log differences in probabilities. Representativeness depends on the percentage probability increase of a type between G and $-G$. This captures a form of diminishing sensitivity, whereby, for fixed probability difference, a type is more likely to be recalled if it is unlikely in the comparison group, namely when $\pi_{t,-G}$ is low.[15] Statistically, representative types are diagnostic of the target group G: the higher is $R(t,G)$, the more a Bayesian DM is confident that an observation t is drawn from G rather than from$-G$.

The ease of recall of highly representative types affects judgments because more easily recalled types are overweighted. We model distorted recall as follows. Denote by $\pi_G/\pi_{-G} \equiv (\pi_{t,G}/\pi_{t,-G})_{t\in T}$ the vector of representativeness of all types in G. We then have:

Definition 2 *The DM attaches to each type $t \in T$ in group G a distorted probability:*

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

where $h_t : R^T_+ \to \mathbb{R}_+$ is a weighting function having the following properties:

1) The weighting of type t weakly increases in the type's representativeness $\pi_{t,G}/\pi_{t,-G}$ and weakly decreases in the representativeness of other types $s \neq t$, $\partial h_t(\cdot)/\partial\left(\pi_{t,G}/\pi_{t,-G}\right) \geq 0$ and $\partial h_s(\cdot)/\partial\left(\pi_{s,G}/\pi_{s,-G}\right) \leq 0$. Thus, the distorted probability $\pi^{st}_{t,G}$ increases in $\pi_{t,G}/\pi_{t,-G}$.

[15] This feature of our definition of representativeness links to Weber's law of sensory perception, see Section 3.2. It also links to our previous work on salience, in which we postulated that log differences in payoffs determine the attention paid to lottery payoffs BGS (2012) and goods' attributes (BGS 2013). Equation (2) establishes the same principle for the domain of probabilities.

2) Distortions only arise when types are differentially representative. Formally, if $\pi_{t,G}/\pi_{t,-G} = 1$ for all $t \in T$, then $h_t(1) = 1$ for all $t \in T$.

We call the distribution $(\pi^{st}_{t,G})_{t \in T}$ the stereotype for G. The stereotype attaches a weakly higher probability to an objectively more likely type, but by property 1) it ceteris paribus overweighs the probability of more representative types. Property 2 implies that if all types are equally representative, the DM holds rational expectations about G.[16],[17]

Most of the results we explore in this paper hold for a general weighing function $h_t(\cdot)$. Specific functional forms capture added assumptions about the psychology of representativeness-based recall, and are useful in applications. We outline a few specifications and their properties.

- Rank-based stereotypes: the ranking of the representativeness of different types shapes distortions. Denote by $r(t) \in \{1, \ldots, T\}$ the representativeness ranking of type t. When $r(t) = 1$ type t is the most representative one (potentially with ties). We can specify two ways in which a type's representativeness ranking distorts its probability.
  - Rank-based truncation: the DM only recalls the $d \leq T$ most representative types and zero probability is attached to the remaining types.[18] This assumption is used in Gennaioli and Shleifer (2010). Denote by $I(r(t) \leq d)$ an indicator function taking value 1 if t belongs to the d most representative types. Then, the weighting function is $h_t = I(r(t) \leq d)$ so that:

$$\pi^{st}_{t,G} = \frac{\pi_{t,G} I(r(t) \leq d)}{\sum_{s \in T} \pi_{s,G} I(r(s) \leq d)}.$$

[16] In principle, if types are equally representative, the stereotypes could be biased in favor of more likely types. We abstract from this possibility to focus on the implications of representativeness-driven recall. In Appendix C we consider a model in which likelihood also distorts recall.

[17] To the extent that the distribution of types π_G is individual-specific, due to differences in information, experience, or even culture, our model allows for individual heterogeneity in stereotypes. Depending on the distributions, stereotypes may display less, or more, heterogeneity than the underlying beliefs.

[18] These neglected types are not viewed as impossible; they are just assigned zero probability in the DM's current thinking. This formulation allows us to model surprise and reactions to unforeseen contingencies, which have proved useful ingredients in modeling probabilistic judgments (GS 2010) as well as financial crises (Gennaioli, Shleifer, and Vishny, 2012).

  - Rank-based discounting: The DM discounts by a constant factor $\delta \in [0,1]$ the odds of type t relative to its immediate predecessor in the representativeness ranking. Lower δ implies stronger discounting of less representative types. Formally, the weighting function is $h_t = \delta^{r(t)-1}$ and the distortion function is:

$$\pi_{t,G}^{st} = \frac{\pi_{t,G}\delta^{r(t)-1}}{\sum_{s\in T}\pi_{s,G}\delta^{r(s)-1}}.$$

- Representativness based discounting: All else equal, the weight attached by the DM to type t increases continuously with its representativeness. One convenient formulation is $h_t = \left(\pi_{t,G}/\pi_{t,-G}\right)^{\theta}$ so that:

$$\pi_{t,G}^{st} = \frac{\pi_{t,G}\left(\pi_{t,G}/\pi_{t,-G}\right)^{\theta}}{\sum_{s\in T}\pi_{s,G}\left(\pi_{s,G}/\pi_{s,-G}\right)^{\theta}},$$

  where $\theta \geq 0$ captures the extent to which representativeness distorts beliefs. This formulation is particularly convenient when dealing with continuous distribution of the exponential class. Bordalo, Gennaioli and Shleifer (2015) show how continuous discounting allows to parsimoniously describe stereotypical beliefs as shifts in true distributions.

These functional forms all embody the central idea of our model, namely that the stereotype overweighs the probability of more representative types. Rank-based truncation captures a central manifestation of limited memory, namely forgetting unrepresentative types. Smoother discounting (based on ranking or on representativeness) may be more appropriate when the type space is small, such as in the T-shirts experiment of Section 2. Furthermore, smooth discounting can be more tractable in certain settings.

Section 4 characterizes the general properties of stereotypes and in particular their "kernel of truth" structure under the general weighting function of Definition 2. To bring the model to the data in Section 5, we derive linear approximations of stereotypical beliefs by assuming that the weighting function is differentiable with respect to a type's representativeness. This assumption excludes rank-based weighting but allows for many possibilities.

3.2 Discussion of Assumptions

In our model stereotypes are simplified mental representations of groups characterized by limited and selective recall of those groups' types. Recall is limited because not all types are equally strongly evoked from memory. Recall is selective because the types that are strongly evoked, and dominate the stereotype, are the most representative ones relative to a comparison group. In the next section, we show that representativeness-based recall generates the key features of stereotypes stressed in social psychology: namely, stereotypes often highlight (and exaggerate) real differences between groups and are selectively localised around the most distinctive features of the target group relative to other groups (Hilton and Hippel 1996), in line with the kernel of truth hypothesis.

Before moving to the formal analysis, we note some properties and limitations of our assumptions. First, we do not claim that representativeness is the only heuristic that shapes recall. Decision makers may for instance find it easier to recall types that are sufficiently likely. Another potentially important mechanism is availability, understood by Kahneman and Tversky (1972) as the "ease" with which information comes to mind. This may capture aspects such as recency and frequency of exposure, which might be independent of likelihood or representativeness.[19] In Appendix C we present a more general recall mechanism where the weight distortions are driven by a combination of representativeness and likelihood of types. This is equivalent to relaxing Property 3) of Definition 2. In this extension, stereotypes are less extreme in the sense that, controlling for representativeness, they tend to underweight unlikely tails. This model can offer a useful starting point to capture availability as well, even though a full model of availability is beyond the scope of this paper. Even in this more general setting, the influence of representativeness on recall is the driving force of stereotypes which, in line with the social psychology perspective, are based on underlying differences among groups. The experimental evidence in Section 2 also shows that representativeness alone can generate systematic biases in a lab environment that controls for likelihood, exposure, and other factors.

[19] For instance, in the aftermath of the 9/11 terrorist attacks, and the ensuing media coverage, a US respondent asked what Arabs are like might more easily recall terrorists than Bedouins, even when there are vastly more Bedouins than terrorists among Arabs, and even though all Bedouins are Arabs, so that Bedouins are more representative of Arabs than terrorists.

The second set of model-related issues concerns how to implement Definition 1 in applications, including specifying the set of types T and the comparison group $-G$ considered by the DM. Take first the specification of the group G and of the type space T. Often, the problem itself provides a natural specification of these features. This is the case in the empirically important class of "closed end" questions, such as those used in surveys, which provide respondents with a set of alternatives. This is also the case in our experiments of Section 2. More generally, the problem solved by the decision maker – such as evaluating the CV of a job applicant coming from a certain ethnic group – primes both a group and a set of types, such as the applicant's qualification or skill. When, as we assume here, types have a natural order (such as income, age, education), the granularity of T is also naturally given by the problem (income, age and years of schooling brackets). Where the set of types is not specified by the problem, decision makers spontaneously generate one.[20] Psychologists have sought for years to construct a theory of natural types (Rosch 1998). We do not make a contribution to this problem, but in many problems of interest in economics the set of types is naturally given.[21] Furthermore, we note that in our model details of the type space can be important under rank-based truncation, but they matter less under smooth discounting.

Consider next the role of the comparison group $-G$. This group captures the context in which a stereotype is formed and, again, is often implied by the problem: when $G =$ Floridians, $-G =$Rest of US population; when $G =$ Black Americans, $-G =$White Americans. A distinctive prediction of our model, confirmed by our experiments in Section 2, is that the stereotype for a given group G depends on the comparison group $-G$. Shih et al (1999) show that Asian-American women self-stereotype themselves as better or worse in math, with corresponding impact on performance, when their ethnicity or gender, respectively, is primed. When $-G$ is not pinned down by the problem itself, to derive testable predictions from representativeness, we set $-G = \Omega\backslash G$ where Ω is the natural population over which the unconditional distribution of types is measured.

[20] For example, suppose a person is asked to guess the typical occupation of a democratic voter in an "open ended" format (without being provided with a set of alternatives). Here the level of granularity at which types are defined is not obvious (e.g. teacher vs a university teacher vs a professor of comparative literature).

[21] Strictly speaking, granularity is also an issue when types are ordered. However, in contrast to non-ordered categories, in ordered categories distributions are typically smoother, so changing the bracketing has minor effects on estimates.

Finally, we offer some comments on whether such stereotypical thinking, despite the distortions it may cause, can be efficiency-enhancing given cognitive limitations. Although we do not present a formal analysis, our results point to some potentially relevant considerations.

The kernel of truth perspective indicates that the main benefit of representativeness-based stereotypes is that, in many cases, they allow people to more easily detect the correct direction of a signal. For example, when assessing a group associated with higher values of a certain type t, stereotypical thinkers immediately and correctly perceive the group's higher mean. When even small changes in means are payoff relevant, the kernel of truth contained in stereotypes helps make better choices. In this sense, representativeness can provide an efficient way to detect the direction of a signal using only a few types because it emphasizes contrast among groups or situations.

A useful analogy can be drawn between representativeness and the psychophysics of visual perception. Like representativeness, our perceptive apparatus emphasizes contrast. An object is perceived to be brighter if set against a darker background. The contrast principle in visual perception has been justified as an optimal way to identify brightness, color, size, distance, in the presence of multiplicative background noise (Kersten et al. 2004). For example, an object's apparent brightness is affected both by intrinsic shade but also by the overall luminosity of the environment. Contrasting the brightness of the object in question with that of a nearby object helps control for the ambient luminosity (Cunningham 2013).

The representativeness heuristic might be justified in a similar way. By emphasizing contrast, this heuristic allows individuals to easily detect differences across groups, even when those groups share many common factors. To see this, compare our results with those of a different model of limited memory, where recall of types is based on their likelihood. As the likelihood of extreme, tail types changes, increasing the average type of a group or rendering it more risky, likelihood-based stereotypes would not change, but would remain centered on the most likely, mean, type. The DM would thus fail to detect the fact that extreme events have now become more likely, which may have important consequences.

In sum, we see representativeness as the key assumption that generates stereotypes' most central feature: highlighting differences between groups, even if these differences are overwhelmed by their similarities, and even if the resulting beliefs are inaccurate.

4 Properties of Stereotypes

We now study stereotypical beliefs and their accuracy. To illustrate the role of representativeness, we first ask to what extent the most representative type is a good fit for the group, namely whether it is modal. Next, we assess the accuracy of the entire stereotypical distribution. To so so, we focus on a cardinal types and compare the stereotype's mean and variance to the true ones.

4.1 Likely vs Unlikely Exemplars

The most representative type for a group is the one that agents most easily recall and associate with the group itself, for instance a red haired Irishman or a 65+ year old Floridian. Social psychologists call this type the exemplar of the group. In the rank-based truncation model, the exemplar coincides with the stereotype when memory limits are so severe that only one type is recalled, namely $d = 1$.

The $d = 1$ stereotype is (relatively) accurate if the exemplar coincides with the group's modal type. Accuracy here means that a stereotypical distribution concentrated on a single type minimizes the distance $\sum_t (\pi^{st}_{t,G} - \pi_{t,G})^2$ to the true distribution if and only if the examplar is the modal type. In contrast, the $d = 1$ stereotype is relatively inaccurate if the exemplar is less likely. Equation (2) then yields the following characterization.

Proposition 1 *Consider the distributions $\pi_{t,G}$ and $\pi_{t,-G}$ for G and $-G$. If for any $t, t' \in T$:*

i) $\pi_{t,G} > \pi_{t',G}$ if and only if $\pi_{t,-G} > \pi_{t',-G}$, then the most representative type may coincide with the modal type for at most one group.

ii) $\pi_{t,G} > \pi_{t',G}$ if and only if $\pi_{t,-G} < \pi_{t',-G}$, then each group's most representative type is its modal type (and the types' likelihood and representativeness ranking coincide).

Case i) says that when groups have similar distributions, the most representative type is inaccurate for at least one group, potentially for both. Representativeness draws the DM's attention to group differences, leading him to neglect the fact that the groups are in fact quite similar, in concentrating around the same mode. This is the mechanism at play in the Florida example.

Case ii) says that the most representative type tends to be accurate, in the sense that it coincides with the modal type in both groups, when the distributions are very different. In this case, groups differ the most around their modes, so that representativeness and likelihood coincide. Thinking of Swedes as "blond haired" and Europeans as "dark haired" is accurate precisely because these are majority traits of the Swedish and European populations, respectively.[22]

In general, the relationship between representativeness and likelihood of types helps assess the accuracy of stereotypes. In the rank-based truncation model, conditional on any memory limit $d \leq T$, the stereotype minimizing the distance $\sum_t (\pi^{st}_{t,G} - \pi_{t,G})^2$ is obtained when the representativeness and likelihood rankings coincide. The least accurate stereotypes is instead obtained when the most representative types are the least likely ones. But even for smoother discounting functions, we show in section 4.2 that strong biases arise precisely when extreme and potentially unlikely types are representative. Ethnic stereotypes based on crime or terrorism exhibit precisely this error: neglect of the fact that by far the most common types in all groups are honest and peaceful. Indeed, in our laboratory findings, while the modal T-shirt color worn by both girls and boys in both conditions was purple, participants were more likely to recall green or blue as the modal color when this less frequent color was representative of the group.

4.2 Stereotypical Moments

We now characterize how the first two moments of a distribution are distorted by the process of stereotyping. The following results hold for any weighting function $h_t(\cdot)$ satisfying Definition 2. We consider two canonical cases for cardinal, ordered types that prove useful in illustrating the predictions of the model.

In the first case, groups G and $-G$ are such that the likelihood ratio $\pi_{t,G}/\pi_{t,-G}$ is monotonic in t. The monotone likelihood ratio property (MLRP) holds to a first approximation

[22] Even though such stereotypes are accurate conditional on $\mathbf{d} = \mathbf{1}$, there can still be significant exaggeration. For instance, voters in some U.S. states are perceived as "blue" or "red" because a majority of the population indeed votes Democrat or Republican. In reality, even in "blue" states, far from everyone votes Democrat. In the 2012 Presidential election, vote shares of either candidate ranged from 25% to 75% and over 35 states were in the 40% to 60% range.

in many empirical settings (see Figure 3) and is also assumed in many economic models, such as standard agency models.[23] If $\pi_{t,G}/\pi_{t,-G}$ is monotonically increasing (decreasing) in t, then group G is associated with higher (lower) values of t relative to the comparison group $-G$. Formally:

Proposition 2 *Suppose that MLRP holds. Then, for any weighting function $h_t(\cdot)$:*

i) If the likelihood ratio $\frac{\pi_{t,G}}{\pi_{t,-G}}$ is strictly increasing in t, then

$$\mathbb{E}^{st}(t|G) > \mathbb{E}(t|G) > \mathbb{E}(t\,|{-G}),$$

ii) If the likelihood ratio $\frac{\pi_{t,G}}{\pi_{t,-G}}$ is strictly decreasing in t, then

$$\mathbb{E}^{st}(t|G) < \mathbb{E}(t|G) < \mathbb{E}(t\,|{-G}).$$

Under MLRP, the most representative part of the distribution for G is the right tail if $\pi_{t,G}/\pi_{t,-G}$ increases in t or the left tail if $\pi_{t,G}/\pi_{t,-G}$ decreases in t. The representative tail is then overweighted while the other, non-representative tail is underweighted. As a consequence, the assessed mean $E^{st}(t|G)$ is too extreme in the direction of the representative tail. Critically, this exaggeration is directionally correct: the stereotype contains a kernel of truth. The DM overestimates the mean of G if this group has a higher mean than the comparison group, namely $E\left(t\,|G\right) > E\left(t\,|{-G}\right)$ and conversely if $E\left(t\,|G\right) < E\left(t\,|{-G}\right)$.

In line with the social cognition perspective, the stereotype contains a kernel of truth: it induces the agent to exaggerate the mean difference between G and $-G$, and in particular to inflate the association of G with its most representative types.[24] For instance, when judging

[23] Examples include the Binomial and the Poisson families of distributions with different parameters. The characterisation of distributions satisfying MLRP is easier in the case of continuous distributions, see Appendix E: two distributions $f(x)$, $f(x-\theta)$ that differ only in their mean satisfy MLRP if and only if the distribution $f(x)$ is log-concave. Examples include the Exponential and Normal distributions. To the extent that discrete distributions sufficiently approximate these distributions (as the Poisson distribution $Pois(\lambda)$ approximates the Normal distribution $N(\lambda, \lambda)$ for large λ), they will also satisfy MLRP.

[24] For a large class of distributions, the DM's assessment of the variance $\mathrm{Var}(t|G)$ is also dampened relative to the truth. In this case, stereotyping effectively leads to a form of overconfidence in which the DM both holds extreme views and overestimates the precision of his assessment. That extreme views and overconfidence (in the sense of over precision) go together has been documented in the setting of political ideology, among others (Ortoleva and Snowberg 2015). In our model, this occurs when one tail of the distribution is representative,

an asset manager who performs well, we tend to over-emphasize skill relative to luck because higher skill levels are relatively more associated with higher performance. This occurs even if for both skilled and unskilled managers high performance is mostly due to luck.

The second case in which we charaterize the stereotypical distributions, groups G and $-G$ have the same mean $E(t)$ but differ in their variance. We abstract from skewness and higher moments by considering distributions $(\pi_{t,G})_{t\in T}$ and $(\pi_{t,-G})_{t\in T}$ that share the same support and are both symmetric around the median/mean $E(t)$. In this case we have:

Proposition 3 *Suppose that in G more extreme types are relatively more frequent than in $-G$. Formally, the likelihood ratio $\frac{\pi_{t,G}}{\pi_{t,-G}}$ is U-shaped in t. Then, for any weighting function $h_t(\cdot)$ stereotypical beliefs satisfy:*

$$\begin{aligned} Var^{st}(t|G) &> Var(t|G) > Var(t|-G) > Var^{st}(t|-G), \\ \mathbb{E}^{st}(t|G) &= \mathbb{E}(t). \end{aligned}$$

When group G has a higher relative prevalence of extreme types, its representative types are located at both extremes of the distribution. The DM's beliefs about G are then formed by overweighting both tails while underweighting the unrepresentative middle. The overweighting of G's tails causes the assessment of its variance $Var^{st}(t|G)$ to be too high. For example, the skill distribution of immigrants to the US may be perceived as being bimodal, with immigrants being perceived as either unskilled or very skilled relative to the native population. The mean of the group, in contrast, is assessed correctly, because the stereotypical distribution remains symmetric around $E(t)$. As before, the stereotype contains a kernel of truth. It induces the agent to exaggerate the true differences between groups, namely the higher variance of G relative to its counterpart.

Finally, in many examples groups that differ in means also differ in variances to the extent that MLRP does not hold. In Appendix A, we generalise Proposition 3 to the case where distributions have different means (but are still symmetric around them). We show

so that the decision maker neglects types in the non-stereotypical tail. Formally, one can demonstrate that in this case the decision maker assesses the variance of types to be lower than the true value $\text{Var}(t|G)$ provided the tails of the distribution $\pi_{t,G}$ are not too heavy. The result is easier to formalise in the continuous case in terms of log-concave distributions (see Proposition 4).

that stereotypes exhibit a kernel of truth with respect to both mean differences and variance differences. That is, people inflate the mean of the group that has a truly higher mean and the variance of the group that has a truly higher variance (although stereotypes are no longer symmetric).

To summarise, the psychology of representativeness yields stereotypes that are consistent with the social cognition approach in which individuals assess groups by recalling and focusing on distinctive group traits. When there are systematic differences between groups, stereotypes get the direction right, but exaggerate differences.

4.3 Some Examples

We illustrate the predictions of the model of using a series of examples. We focus on kernel of truth about differences in group means because assessments of average group traits are more available than data on assessed variances. In Section 5, we map Proposition 1 into regression specifications that we run using two datasets on political beliefs.

Consider stereotypical beliefs about income distributions of Black and White households in the US. Figure 3, panel A, presents the true distributions $\pi_{t,B}$ and $\pi_{t,W}$ obtained from the US Census Bureau.[25] The types t are given by a coarsening of the income bins used by the Census. The panel also presents the representativeness $\pi_{t,W}/\pi_{t,B}$ of each bin t for the White household group (solid line). Two facts stand out: first, the distributions are broadly similar, overlapping over the entire income range and sharing a common modal type, namely middle income (the \$35-\$75k/year range); second, higher income bins are more representative of White households, as evidenced by the fact that the likelihood ratio is monotonically increasing in income t, as in Proposition 2, case i).

In panel B of Figure 3, we plot the stereotypical beliefs as predicted by our model, under truncation weighting when DMs recall only three types. Because higher income bins are more representative of White households, our model predicts that stereotypes about income distributions should be rather extreme: the stereotype of White households truncates away the non-representative left tail of lower income, while the stereotype of Black households truncates away the non-stereotypical right tail of higher income. The resulting assessment is

[25] See www.census.gov/prod/2013pubs/p60-245.pdf, Table A-1.

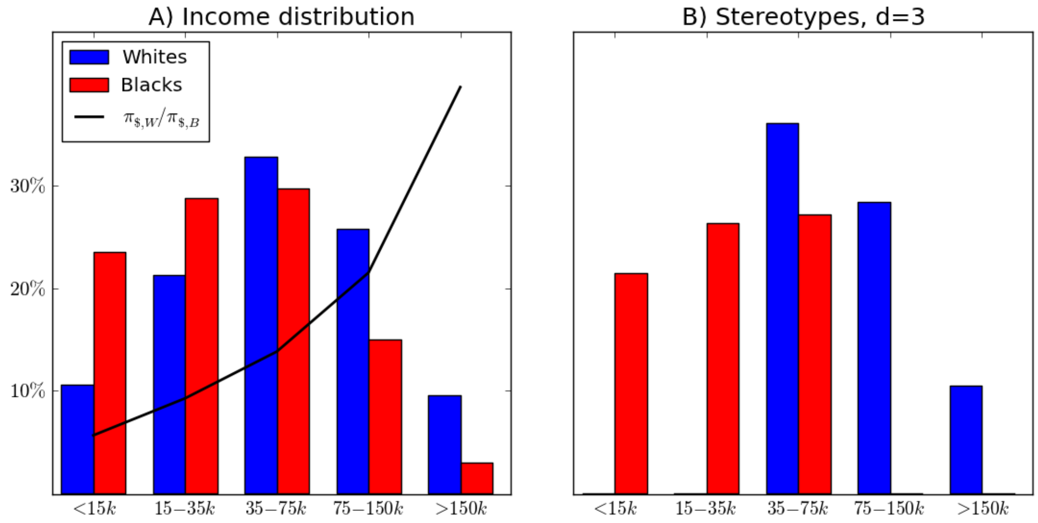


Figure 3: Income for Black and White households in the US: panel A) true distributions; panel B) stereotypical beliefs under truncation.

directionally correct: stereotypers estimate the mean income of blacks to be lower than that of whites, which is indeed the case in reality. The focus on tails however overestimates the mean income of White households and underestimates the mean income of Black households. In this example, they also underestimate the variance within each group.

Although we have not found data on beliefs about income distributions, survey respondents often describe a stereotypical black person as being poor (Devine 1989). Another piece of suggestive evidence comes from the standard finding in social psychology that subjects estimate poor Blacks to outnumber poor Whites (Gilens, 1996). In fact, because the White population in the US is over five times larger than the Black population, poor White households outnumber poor Black households by 2 to 1.[26] This evidence is consistent with the truncation – or at least, the dramatic underestimation – of the poor White household type, which is indeed unrepresentative in the data.

[26] Another piece of suggestive evidence comes from the General Social Survey (GSS). Respondents were asked what wealth level best characterises White and Black US households. In the subjective scale proposed by GSS, a score of 1 (respectively, 7) reflects a belief that almost everyone in the relevant group is rich (resp. poor). GSS respondents gave dramatically different answers for the different groups: two thirds of respondents believe that most Blacks are relatively poor (scores of 5 through 7), while 45% believe that most Whites are relatively well off (scores 1 through 3) and only 9% believe that most Whites are relatively poor. In reality, most blacks and most whites are in the middle class. See www3.norc.org/GSS+Website/Data+Analysis/.Data, questions WLTHBLKS and WLTHWHTS.

Consider another example, of potentially large economic consequence. A growing body of field and experimental evidence points to a widespread belief that women are worse than men at mathematics (Eccles, Jacobs, and Harold 1990, Guiso, Monte, Sapienza and Zingales 2008, Carrell, Page and West 2010). This belief persists despite the fact that, for decades, women have been gaining ground in average school grades, including mathematics, and have recently surpassed men in overall school performance (Goldin, Katz and Kuziemko 2006, Hyde et al, 2008). This belief, shared by both men and women (Reuben, Sapienza and Zingales 2014), may help account, in part, for the gender gap in the choices of high school tracks, of college degrees and of careers, with women disproportionately choosing humanities and health related areas (Weinberger 2005, Buser, Niederle and Oosterbeek 2014) and foregoing significant wage premiums to quantitative skills (Bertrand 2011).

Gender stereotypes in mathematics, particularly beliefs that exaggerate the extent of average differences, are consistent with the predictions of our model. The fact that men are over-representated at the very highest performance levels leads a stereotypical thinker to exaggerate the magnitude of mean differences. Figure 4 shows the score distributions from the mathematics section of 2012's Scholastic Aptitude Test (SAT), for both men and women.[27] The distributions are nearly identical, and average scores are only slightly higher for men (531 versus 499 out of 800). However, scores for men have a heavier right tail, with men twice as likely to have a perfect SAT math score than women.[28] In light of such data, the stereotypical male performance in mathematics is high, while the stereotypical female performance is poor. Predictions based on such stereotypes are inaccurate, exaggerating true differences. While differences in the right tail of the distribution are unlikely to be relevant in most decisions, stereotypical thinking driven by these differences has the potential to impact economically-important decisions, whether through self-stereotyping (i.e., choice of careers

[27] Standardized test performance measures not only innate ability but also effort and investment by third parties, Hyde et al, 2008. The mapping of test performance into inferences about innate ability is an issue not addressed by our model.

[28] Results are similar for the National Assessment of Educational Progress (NAEP), which are more representative of the overall population. For SAT scores see http://media.collegeboard.com/digitalServices/pdf/research/SAT-Percentile-Ranks-By-Gender-Ethnicity-2013.pdf. For NAEP scores for 17 year olds in mathematics, see http://nationsreportcard.gov/ltt_2012/age17m.aspx. See Hyde et al (2008), Fryer and Levitt (2010), and Pope and Sydnor (2010) for in-depth empirical analyses of the gender gap in mathematics.

or majors as in Buser, Niederle, Osterbeek (2014)) or through discrimination (i.e., hiring decisions as in Bohnet, van Geen, and Bazerman (2015)).

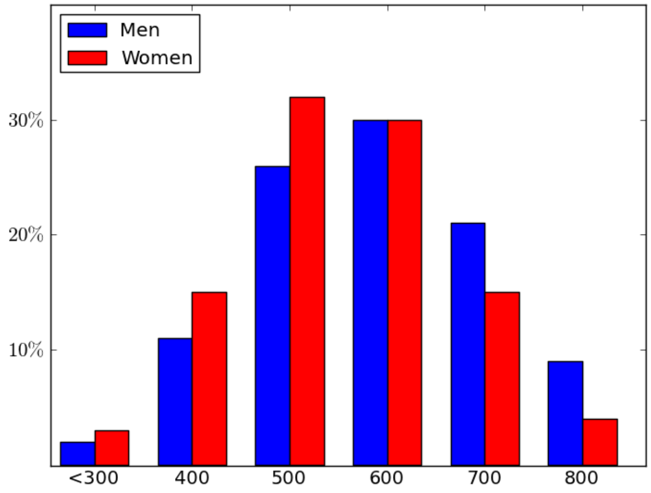


Figure 4: SAT scores by gender (2012)

The logic of extreme, yet directionally correct, stereotypes can also shed light on the well documented phenomenon of base rate neglect (Kahneman and Tversky, 1973). Indeed, Proposition 2 implies that the DM overreacts to information that assigns people to groups, precisely because such information generates extreme stereotypes.[29] Consider the classic example in which a medical test for a particular disease with a 5% prevalence has a 90% rate of true positives and a 5% rate of false positives. The test assigns each person to one of two groups, + (positive test) or − (negative test). The DM estimates the frequency of the sick type (s) and the healthy type (h) in each group. The test is informative: a positive result increases the relative likelihood of sickness, and a negative result increases the relative likelihood of health for any prior. Formally:

$$\frac{\Pr(+|s)}{\Pr(+|h)} > 1 > \frac{\Pr(-|s)}{\Pr(-|h)}. \tag{3}$$

This condition has clear implications: the representative person who tests positive is sick, while the representative person who tests negative is healthy. Following Proposition 2, the

[29] In Section F we explore in detail how stereotypical beliefs react to a different kind of information, namely information about the distribution of types when groups are *given*.

DM reacts to the test by moving his priors too far in the right direction, generating extreme stereotypes. He greatly boosts his assessment that a positively tested person is sick, but also that a negatively tested person is healthy. Because most people are healthy, the DMs assessment about the group that tested negative is fairly accurate but is severely biased for the group that tested positive. This analysis formalises Tversky and Kahneman's (1983) verbal account of base rate neglect.[30]

Extreme stereotypes may shed light on several other phenomena. When assessing the performance of firms in a hot sector of the economy, the investor recalls highly successful (and some moderately successful) firms in that sector. However, he neglects the possibility of failures, because failure is statistically non-diagnostic, and psychologically non-representative, of a growing sector – even if it is likely. This causes both excessive optimism (in that the expectation of growth is unreasonably high) and overconfidence (in that the variability in earnings growth considered possible is truncated). True, the hot sector may have better growth opportunities on average, but representativeness exaggerates this feature and induces the investor to neglect a significant risk of failure. Similarly, when assessing an employee's skill level, an employer attributes high performance to high skill, because high performance is the distinctive mark of a talented employee. Because he neglects the possibility that some talented employees perform poorly and that some non-talented ones perform well (perhaps due to stochasticity in the environment), the employer has too much faith in skill, and neglects the role of luck in accounting for the output.

[30] Our account is distinct from a mechanical underweighting of base-rates in Bayes rule, as in Grether (1980) and Bodoh-Creed, Benjamin and Rabin (2013). In those models, upon receiving the test results, the DM can update his beliefs in the wrong direction: he can be less confident that a person is healthy after a negative test than under his prior, which cannot happen in our model.

While this prediction of our model seems consistent with introspection, we are not aware of experimental evidence on this point. Griffin and Tversky (1992) present evidence consistent with pure neglect of base rates, but in a significantly different task, namely inferring the bias of a coin from a history of coin flips. Such experiments are hard to compare with the predictions of our model, because subjects are asked to generate distributions of different numbers of coin flips in their minds, which is a much more involved task than to recall types of a given distribution. Their assessments, then, might be wrong for other reasons. See Bodoh-Creed, Benjamin and Rabin (2013) for a detailed discussion.

5 Empirical Evidence on Political Stereotypes

We now bring our model to the data by using two data sets on political preferences, and beliefs about political preferences, in the U.S. We investigate three propositions. First, we test whether beliefs are correct or depart systematically from the truth. Second, we test if they depart from the truth by exaggerating (mean) differences, as per the kernel of truth logic. Third, we test if distortions in beliefs can be accounted for by the overweighting of highly representative types.

These tests allow us to assess the extent to which the data are consistent with theories of stereotypes. The statistical discrimination approach builds on the assumption that people form rational expectations of group traits. Comparing beliefs to the truth allows us to assess the validity of this assumption in our data. Our psychological model of stereotypes, in contrast, predicts that mean beliefs should exaggerate true differences among groups, in particular by putting too much weight on representative types. The second and third tests allow us to assess whether this prediction holds in our data. Finally, models of categorization (Mullainthan 2003, Fryer and Jackson 2008) predict that beliefs about average group traits can be distorted if they exaggerate the incidence of the most likely group type. We perform a test to discriminate among likelihood-based stereotypes and the representativeness-based stereotypes we propose.

To proceed, we first describe the data, then show how to test our model in this empirical setting, and finally, present the results.

5.1 The data

We have two data sets on political preferences and beliefs about political preferences. The first data set, from Graham et al (GSN, 2012), contains data from the Moral Foundations Questionnaire. Respondents (1,174 self-identified liberals and 500 self-identified conservatives) answer questions about their position on a subset of 45 issues: 20 moral relevance statements (e.g., “when you decide whether something is right or wrong, to what extent are the following considerations relevant to your thinking?") and 25 moral judgments (e.g., “indicate the extent to which you would agree or disagree"). For each issue, a randomly determined subset of participants states their own position, another subset states their belief on the position of a "typical liberal", and a third subset states their belief on the position of a “typical conservative". The data thus includes the distribution over positions for both liberal and conservatives, as well as the average believed typical position of liberals and of conservatives, on each of the 45 issues. Each position is elicited on 1 - 6 scale.

The second data set, from Westfall et al (WBCJ, 2015), contains data from more than 20,000 respondents to the American National Election Survey, between 1964 and 2012. The survey covers political issues of the day, such as the optimal amount of government spending and service provision (1984 through 2000), or the proper place of women in society (1972 through 1998). We focus on the 10 issues that ask participants to respond on a multi-point, 1 to 7, scale (rather than just indicate binary agreement or disagreement); each of these 10 issues is asked in multiple years. Participants are asked to provide their own position on the scale and their believed position of the Democratic and Republican party ("Where would you place the Democratic (Republican) party on this scale?"). The data includes, for each issue-year observation, the distribution of participant positions for both self-identified liberals and self-identified conservatives, as well as the distribution of believed typical positions of the Democratic and Republican Parties.

5.2 Empirical strategy

Our analysis focuses on beliefs about two groups, Conservatives and Liberals. The types are the possible positions for each issue (1, 2, . . . , 6, 7). For the GSN data, we interpret beliefs about the “typical" element of a group to coincide with the believed average position in that group. Similarly, for the WBCJ data we use the believed party positions as a proxy for believed mean of each group.[31] We then take as a benchmark the hypothesis that individuals hold accurate beliefs about each group, and in particular that believed mean position should equal true mean position, at least on average across subjects. The accurate beliefs hypothesis underlies the most common economic model of stereotyping, statistical discrimination.

[31] This assumption is consistent with the authors' interpretation of the data (GSN 2012, WBCJ 2015). Furthermore, to the extent that this assumption holds equally well for most issues within a data set, our focus on across-issue differences should allow us to test the predictions of our model even with an imperfect proxy for beliefs of mean positions.

To assess our representativeness-based model, we perform a regression exercise. The Appendix shows that under the assumption that the weighting function $h_t(\cdot)$ is differentiable, our model yields the following two regression specifications.

Corollary 1 *Define μ_G^T to be the true mean trait of group G and μ_G^B to be its believed mean trait, where $G \in \{conservative, liberal\}$. We then have:*

1) Kernel of truth regression. There is a coefficient $h > 0$ such that the following equation holds in our model as a first order approximation around the case of identical distributions $\pi_G / \pi_{-G} = 1$:

$$\mu_G^{st} = \mu_G^T (1 + h) - h \cdot \mu_{-G}^T. \tag{4}$$

2) Representativeness regression. Compute $\sum_{t \geq T-2} \pi_{t,cons} / \sum_{t \geq T-2} \pi_{t,lib}$ as the average representativeness of the three types above the median for conservatives. Our model then entails the following approximate equations:

$$\mu_{cons}^{st} = \mu_{cons}^T + \left(\bar{t}_{H,cons} - \bar{t}_{L,cons}\right) \left(\sum_{t \geq T-2} \pi_{t,cons}\right) \cdot \frac{\sum_{t \geq T-2} \pi_{t,cons}}{\sum_{t \geq T-2} \pi_{t,lib}}, \tag{5}$$

$$\mu_{lib}^{st} = \mu_{lib}^T + \left(\bar{t}_{L,lib} - \bar{t}_{H,lib}\right) \left(\sum_{t \geq T-2} \pi_{t,lib}\right) \cdot \frac{\sum_{t \geq T-2} \pi_{t,cons}}{\sum_{t \geq T-2} \pi_{t,lib}}, \tag{6}$$

where $\bar{t}_{H,G} = \sum_{t \geq T-2} t \cdot \pi_{t,G}$ is the average type among the largest three in G and $\bar{t}_{L,G} = \sum_{t \leq 3} t \cdot \pi_{t,G}$ is the average type among the smallest three, so that $\bar{t}_{H,G} > \bar{t}_{L,G}$ by construction.

The first regression allows us to test for the kernel of truth hypothesis, while the second set of regressions allows us to test for the role of representativeness.

Equation (4) says that respondents in our model inflate the average position of a group, say the conservatives, if and only if the group has a higher average position than the other group, namely the liberals. Formally, $\mu_{cons}^{st} > \mu_{cons}^T$ if and only if $\mu_{cons}^T > \mu_{lib}^T$. Because in our measurement scale higher types mean “more conservative", we expect: i) believed conservative average to be higher than the truth, and ii) the extent of overstatement to decrease in the average liberal position μ_{lib}^T. Conversely, we expect the average liberal position to be lower than the truth, the more so the higher the average conservative position μ_{cons}^T. The basis of these predictions is context dependence: information about the distribution of $-G$ is relevant for the beliefs about group G. This context dependence is inconsistent with rational expectations, in which only the group's own means should affect beliefs. To implement this regression, we test the hypothesis that the true mean μ_G^T is a significant predictor of the believed mean μ_G^{st} with a positive sign, while the other group's true mean μ_{-G}^T is a predictor of the believed mean with a negative sign.

Equations (5) and (6) say that repondents' assessment bias is shaped by representativeness. Consider Equation (5), which applies to conservatives. It says that participants inflate the average conservative position more, formally $\mu_{cons}^{st} - \mu_{cons}^T$ is higher, when the right tail is more representative for the conservatives. Similarly, Equation (6), says that subjects deflate the average liberal position more, formally $\mu_{cons}^{st} - \mu_{cons}^T$ is lower, when the right tail is more representative for the conservatives. To implement this regression, we test the hypothesis that the inflation in conservative positions is positively associated with the representativeness of the right tail for the coservatives, while the inflation in liberal positions is negatively associated with it. Once again, the representativeness of the right tail is computed using the true distribution of positions.

When interpreting Equations (5) and (6), we cannot rule out that in many cases the representative tail may also be the most likely one. As a consequence, these tests cannot perfectly distinguish a representativeness-based from a likelihood-based model of distorted beliefs. We perform two additional tests. First, we run versions of Equations (5) and (6) in which we control for the likelihood of tails. Second, we compute numerically the assessments under a representativeness-based model of stereotypes and those that would arise under a likelihood-based model of stereotype. We then assess which of these two is better able to match the data on beliefs.

5.3 Empirical Results

Before the main analysis, we illustrate the structure of the data and the nature of our predictions with two simple examples. We select two examples from the GSN data set, focusing on beliefs about the conservatives for these issues. In Example 1, participants are asked about their agreement with the statement, "It can never be right to kill a human being". In Example 2, participants are asked about the moral relevance of "whether or not someone cared for someone weak or vulnerable". As can be seen in Figure 5, in Example 1 the modal position (Strongly Disagree (1)) and most representative positions (Strongly Disagree (1)) coincide for conservatives. We can contrast this with Example 2 in Figure 5; here, the most representative types (Slightly Relevant (3), Not at all Relevant (1)) are not most likely for the conservative group. In the spirit of Proposition 1, we predict that beliefs will be distorted in the direction of the most representative types. Thus, we expect more exaggeration in Example 2 than in Example 1, since in Example 2 the most representative types (in the left tail) are far from the modal type, while in Example 1, they coincide. This is what we find: the conservative position is exaggerated by only 0.09 positions in Example 1 (true mean 2.99, believed mean 2.90), but by 1.06 positions in Example 2 (true mean 4.21, believed mean 3.15).

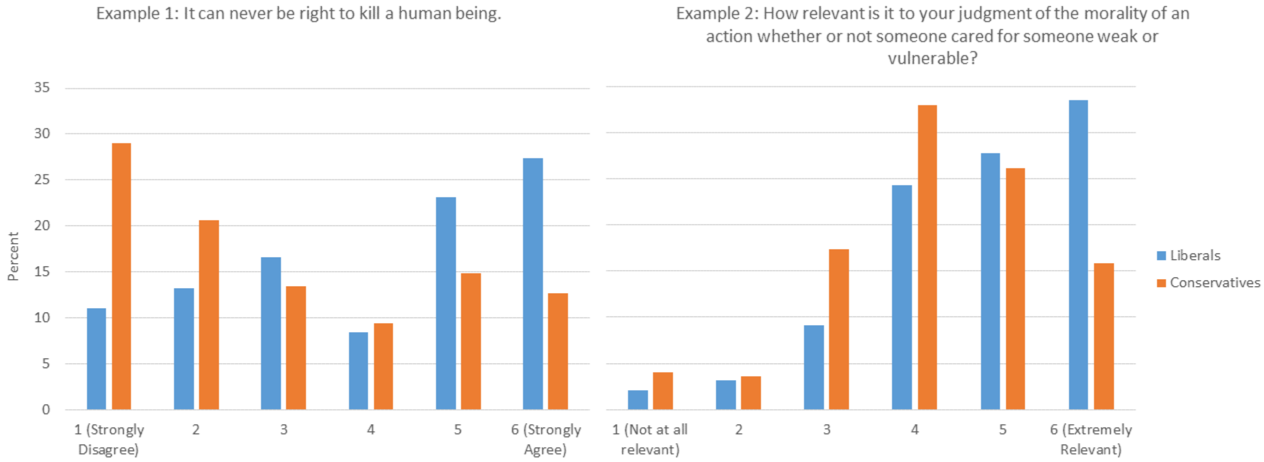


Figure 5: Two Examples

We now turn our attention to the full data sets. We treat each (issue, year) pair as an observation, and we cluster standard errors at the issue level. For the GSN data, we have 45 observations: 45 issues each measured in the same year. For the WBCJ data, we have 66 observations: 10 issues, each measured in multiple years.

Our first analysis simply documents that there is systematic exaggeration in both data sets. This is a primary focus of the original analysis in GSN (2012) and WBCJ (2015), and we report their interesting findings here. Figure 6 shows that the believed difference between typical conservative and typical liberal positions is larger than the true difference in mean positions, for 109 of the 111 observations. The data for both GSN (purple squares) and WBCJ (orange triangles) lie above the 45 degree line (dashed).[32] Average exaggeration is 0.62 positions on the scale (0.66 in the GSN data, 0.59 in the WBCJ data).[33]

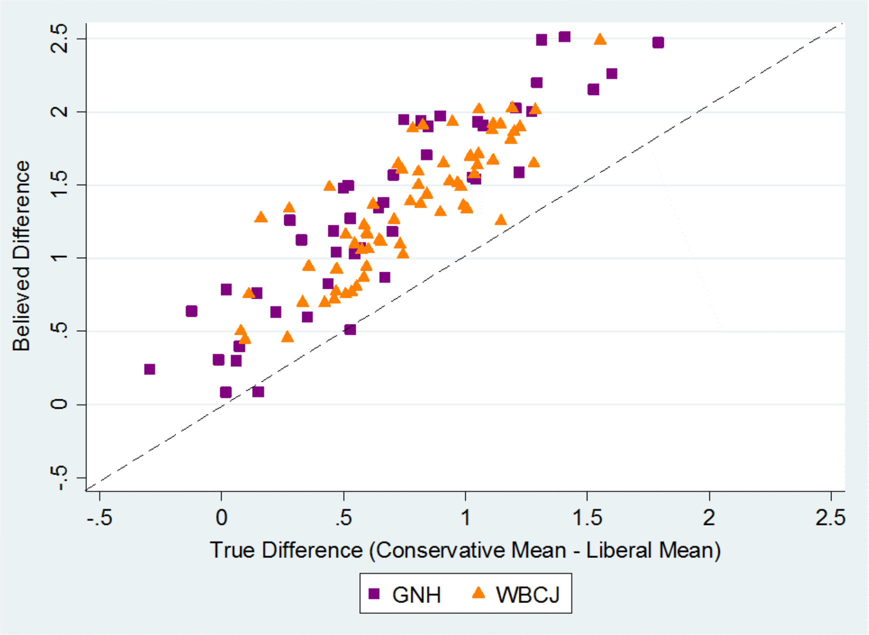


Figure 6: Exaggeration of Differences

The systematic and significant exaggeration of mean differences suggests that the benchmark model of accurate beliefs is missing something important. As we show next, this exaggeration reflects the fact that believed means are typically more extreme than true means. First, note that the kernel of truth regression (Equation 4) generates exaggeration of mean differences, just as documented in Figure 6:

[32] For convenience, we recode all issues so that the high end of the scale (6,7) represents the stereotypically more conservative position.

[33] A natural question to ask is how beliefs vary across liberals and conservatives. That is, do beliefs about a group G depend on membership in G versus $-G$. Our model does not speak to this issue. However, in an Appendix, we show that the results we document below hold for both beliefs held by conservatives and beliefs held by liberals.

Table 1: Information about -G Predicts Beliefs about G

OLS Predicting Believed Mean of Group G						
	G = Conservatives			G = Liberals		
	GSN	WBCJ	Pooled	GSN	WBCJ	Pooled
True Mean Conservatives	1.02**** (0.097)	0.98**** (0.133)	0.96**** (0.076)	-0.21**** (0.060)	-0.19 (0.116)	-0.25**** (0.060)
True Mean Liberals	-0.35*** (0.106)	-0.86**** (0.134)	-0.58**** (0.131)	0.987**** (0.066)	0.39*** (0.106)	0.73**** (0.135)
Constant	1.51*** (0.195)	3.35**** (0.269)	2.35**** (0.279)	0.69**** (0.122)	2.58**** (0.249)	1.56**** (0.270)
R-squared	0.83	0.53	0.66	0.92	0.32	0.68
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 WBCJ data set.

$$\mu_G^{st} - \mu_{-G}^{st} = (1 + 2h) \cdot \left[\mu_G^T - \mu_{-G}^T\right]$$

The results from regression (4) are shown in Table 1. While these results are strong evidence of context-dependence, and are consistent with our model (Equation 4), they do not pin down a role for representativeness of types. To provide further evidence of the role of stereotypical thinking, we develop a test that relates the magnitude of representativeness of tail types to the magnitude of belief distortions.

To map this prediction to the data, we implement the regressions in Equations (5), (6). We compute the average representativeness of tail positions (ARTP) following Corollary 1, $ARTP_{cons} = \sum_{t \geq T-2} \pi_{t,cons} / \sum_{t \geq T-2} \pi_{t,lib}$. We again test the hypothesis that $ARTP_{cons}$ is a significant predictor of μ_{cons}^{st} with a positive sign, and a predictor of μ_{lib}^{st} with a negative sign. Table 2 shows that, conditional on true mean, $ARTP_{cons}$ predicts believed mean for each group G as predicted. In Appendix H we repeat this analysis and obtain similar results controlling for the average likelihood on the tail positions. This rules out that what drives these effects is that representative tails are also likely tails, and indicates the additional effect

Table 2: Average Representativeness of Tail Positions Predicts Beliefs

	OLS Predicting Believed Mean of Group G					
	G = Conservatives			G = Liberals		
	GSN	WBCJ	Pooled	GSN	WBCJ	Pooled
True Mean	0.78****	0.24**	0.51****	0.72****	0.18***	0.41****
of G	(0.06)	(0.08)	(0.09)	(0.05)	(0.05)	(0.10)
$\mathbf{ARTP}_{cons}$	0.19**	0.55**	0.25***	-0.14**	-0.12*	-0.24****
	(0.07)	(0.22)	(0.08)	(0.06)	(0.07)	(0.05)
Constant	1.01****	2.60****	1.84****	0.93****	2.71****	2.01****
	(0.26)	(0.45)	(0.29)	(0.22)	(0.25)	(0.32)
R-squared	0.82	0.48	0.60	0.91	0.31	0.70
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 WBCJ data set.

of tails being representative.

Finally, we use a simple form of the model – the truncation model – to make predictions. We use the model to predict the believed mean of a group G at the observation level, when stereotypes include only the d most representative types. Table 3 presents the results for all possible values of d. Our benchmark is predicting the believed mean from the entire distribution, where $d = T$ (final column).

We follow the literature and present the mean squared prediction error (MSPE) as a measure of the magnitude of errors and the mean prediction error (MPE) as a measure of bias (e.g., a positive MPE means predicted means are systematically too high). We also present a simple counting metric: the fraction of observations for which the model underestimates the observed belief.

Table 3, Panel (a) summarizes the results for the GSN dataset (the same exercise is done for the WBCJ dataset in Appendix H, with similar results). The model with $d = 4$ or $d = 5$ produces smaller MSPE than the accurate beliefs benchmark ($d = 6$). 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

Table 3: Prediction Errors of Representativeness-Based Model for GSN Data

Representativeness Model: Truncation to d Most Representative Types						
	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 6$
(a) Predicting Believed Typical Mean of Conservatives in GSN 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 GSN 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

across over and underestimation for most of the truncation models; MPE is also smaller for $2 < d < 6$ than for $d = 6$. We do the same exercise for liberals in the GSN data in Panel (b). Again, the best representativeness-based truncation model, in this case $d = 5$, produces smaller MSPE and MPE than the accurate beliefs benchmark.

We can also compare our model to a model where beliefs are obtained by truncating to the d most likely types. Table 4 shows the MSPE, MPE, and rates of underestimation for the likelihood-based truncation model under different values of d. While the likelihood model produces smaller MSPE and MPE than the 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. And, while the best representativeness-based model is a 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 outperforms the best likelihood-based model and the accurate beliefs benchmark for all metrics. For liberals, the best representativeness-based model outperforms the best likelihood-based model and the accurate beliefs benchmark in terms of MSPE, with more mixed results for MPE.

Table 4: Prediction Errors of Likelihood-Based Model for GSN Data

Likelihood Model: Truncation to d Most Likely Types						
	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 6$
(a) Predicting Believed Typical Mean of Conservatives in GSN 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 GSN 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

6 Conclusion

We presented a model of stereotypical thinking, in which decision makers making predictions about a group recall only a limited range of the group's types or attributes from memory. Recall is limited but also selective: the recalled types are not the most likely ones given the DM's data, but rather the most representative ones, in the sense of being the most diagnostic types about the group relative to other groups. Representativeness implies that what is most distinctive of a group depends on what group it is compared to. We presented experimental evidence that confirms this context dependence in recall-based assessments of groups. Finally, we evaluated the predictions of the model using political data from existing large scale surveys. Again, we find context to be a key predictor of beliefs. Given the richness of the political data, we can go a step further and identify a role for representativeness in particular. As the representativeness of tail types increases, beliefs of a group are distorted in the direction of that tail. A truncation model where the decision-maker neglects least representative types in forming beliefs about a group fits the data better than the accurate beliefs benchmark.

Our approach provides a parsimonious and psychologically founded account of how DMs generate simplified representations of reality, from social groups to stock returns, and offers a unified account of disparate pieces of evidence relating to this type of uncertainty. First, the model captures the central fact that stereotypes highlight the greatest difference between groups, thus explaining why some stereotypes are very accurate, while others lack validity. Still, stereotypes often contain a “kernel of truth”, when they are based on systematic – even if small – differences between groups.

This same logic allows us to describe a number of heuristics and psychological biases, many of which arise in the context of prediction problems. Our model generates both base-rate neglect and confirmation bias (and makes novel predictions for when they occur). To our knowledge, ours is the first model to reconcile these two patterns of behavior, and in fact shows they both arise out of the assumption of representativeness-based recall.

Our model is centrally based on representativeness and it does not capture all the features of stereotypical thinking. However, it captures perhaps the central feature: when we think of a group, we focus on what is most distinctive about it, and neglect the rest.

References:

Adorno, Theodor, Else Frenkel-Brunswik, Daniel Levinson, and Nevitt Sanford. 1950. *The Authoritarian Personality.* New York, NY: Harper & Row.

Arrow, Kenneth. 1973. *The Theory of Discrimination.* In O. Ashenfelter and A. Rees, eds. Discrimination in Labor Markets. Princeton, N.J.: Princeton University Press: 3 – 33.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. "A Model of Investor Sentiment." *Journal of Financial Economics* 49 (3): 307 – 343.

Benjamin, Dan, Matthew Rabin, and Collin Raymond. 2015. "A Model of Non-Belief in the Law of Large Numbers." *Journal of the European Economic Association*, forthcoming.

Bertrand, Marianne. 2011. "New Perspectives on Gender" in O. Ashenfelter and D. Card eds, Handbook of Labor Economics, 4 (B): 1543 – 1590.

Bodoh-Creed, Aaron, Dan Benjamin and Matthew Rabin. 2013. "The Dynamics of Base-Rate Neglect." Mimeo Haas Business School.

Bohnet, Iris, Alexanda van Geen, and Max Bazerman. 2015. "When Performance Trumps Gender Bias: Joint Versus Separate Evaluation." *Management Science*, forthcoming.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012. ”Salience Theory of Choice under Risk." *Quarterly Journal of Economics* 127 (3): 1243 – 1285.

Bordalo, Pedro, Nicola Gennaioli and Andrei Shleifer. 2013. "Salience and Consumer Choice." *Journal of Political Economy* 121 (5): 803 – 843.

Bordalo, Pedro, Nicola Gennaioli and Andrei Shleifer. 2015. "Diagnostic Expectations and Credit Cycles." Harvard University mimeo.

Buser, Thomas, Muriel Niederle and Hessel Oosterbeek. 2014. "Gender, Competitiveness and Career Choices." *Quarterly Journal of Economics* 129 (3): 1409 – 1447.

Carrell, Scott, Marianne Page and James West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Quarterly Journal of Economics* 125 (3): 1101 – 1144.

Chan, Wayne, Robert McCrae, Filip De Fruyt, Lee Jussim, Corinna Lockenhoff, Marleen De Bolle, Paul Costa Jr., Angelina Sutin, Anu Realo, Juri Allik, Katsuharu Nakazato, EmŠlia FickovĞ, Marina Brunner-Sciarra, Nora Leibovich de Figueora, Vanina Schmidt, Chang-kyu Ahn, Hyun-nie Ahn, Maria Aguilar-Vafaie, Jerzy Siuta, Barbara Szmigielska, Thomas Cain, Jarret Crawford, Khairul Anwar Mastor, Jean-Pierre Rolland, Florence Nansubuga, Daniel R. Miramontez, Veronica Benet-MartŠnez, Jerome Rossier, Denis Bratko, Jamin Halberstadt, Mami Yamaguchi, Goran Knezevic, Thomas A. Martin, Mirona Gheorghiu, Peter B. Smith, Claduio Barbaranelli, Lei Wang, Jane Shakespeare-Finch, Margarida Lima, Waldemar Klinkosz, Andrzej Sekowski, Lidia Alcalay, Franco Simonetti, Tatyana Avdeyeva, Antonio Terracciano. 2012. "Stereotypes of Age Differences in Personality Traits: Universal and Accurate?" *Journal of Personality and Social Psychology* 103(6): 1050 – 1066.

Coren, Stanley, and Joel Miller. 1974. Size Contrast as a function of Figural Similarity." *Perception & Psychophysics* 16 (2): 355 – 357.

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

Cunningham, Tom. 2013. "Comparisons and Choice." Unpublished manuscript, Stockholm University.

Deaux, Kay, and Mary Kite. 1985. "Gender Stereotypes: Some Thoughts on the Cognitive Organization of Gender-related Information." *Academic Psychology Bulletin* 7: 123 – 144.

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

Devine, Patricia. 1989. "Stereotypes and Prejudice: Their Automatic and Controlled Components." *Journal of Personality and Social Psychology* 56 (1): 5 – 18.

Eccies, Jacquelynne, Janis Jacobs, Rena Harold. 1990. "Gender Role Stereotypes, Expectancy Effects, and Parents' Socialization of Gender Differences." *Journal of Social Issues* 46 (2): 183 – 201.

Fryer, Roland, and Matthew Jackson. 2008. "A Categorical Model of Cognition and Biased Decision-Making." *B.E. Journal of Theoretical Economics* 8(1).

Fryer, Roland and Steven Levitt. 2010. “An Empirical Analysis of the Gender Gap in Mathematics.” *American Economic Journal, Applied Economics* 2(2): 210 – 240.

Gennaioli, Nicola, and Andrei Shleifer. 2010. “What Comes to Mind.” *Quarterly Journal of Economics* 125 (4): 1399 – 1433.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny. 2012. “Neglected Risks, Financial Innovation, and Financial Fragility.” *Journal of Financial Economics* 104(3): 452 – 468.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny. 2015. “Neglected Risks: The Psychology of Financial Crises.” *American Economic Review, Papers & Proceedings* 105 (5): 310 – 14.

Glaeser, Edward and Yueran Ma. 2014. “The Supply of Gender Stereotypes and Discriminatory Beliefs." In *Human Capital in History: The American Record*, Leah Boustan, Carola Frydman, and Robert Margo, editors: 355 – 389.

Gilens, Martin. 1996. “Race and Poverty in America: Public Misperceptions and the American News Media." *Public Opinion Quarterly* 60 (4): 515 – 541.

Goldin, Claudia, Lawrence Katz and Ilyana Kuziemko. 2006. “The Homecoming of American College Women: The Reversal of the College Gender Gap.” *Journal of Economic Perspectives* 20 (4): 133 – 156.

Grether, David. 1980. “Bayes Rule as a Descriptive Model: The Representativeness Heuristic.” *Quarterly Journal of Economics* 95 (3): 537 – 557.

Griffin, Dale and Amos Tversky. 1992. “The Weighing of Evidence and the Determinants of Confidence.” *Cognitive Psychology* 24 (3): 411 – 435.

Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales. 2008. “Culture, Gender, and Math.” *Science* 320 (5880): 1164 – 1165.

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.

Hilton, James, and William Von Hippel. 1996. “Stereotypes.” *Annual Review of Psychology* 47 (1): 237 – 271.

Hyde, Janet, Sara Lindberg, Marcia Linn, Amy Ellis, and Caroline Williams. 2008. "Gender Similarities Characterize Math Performance." *Science* 321 (5888): 494 – 495.

Judd, Charles, and Bernardette Park. 1993. "Definition and Assessment of Accuracy in Social Stereotypes." *Psychological Review* 100: 109 – 128.

Jussim, Lee, Jarret T. Crawford, Stephanie Anglin, John R. Chambers, Sean T. Stevens, and Florette Cohen. 2015. "Stereotype Accuracy: One of the Largest and Most Replicable Effects in All of Social Psychology " in The Handbook of Prejudice, Stereotyping, and Discrimination, Todd Nelson, editor. Lawrence Erlbaum Publishing.

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.

Kahneman, Daniel, and Amos Tversky. 1972. "Subjective Probability: A Judgment of Representativeness." *Cognitive Psychology* 3 (3): 430 – 454.

Kahneman, Daniel, and Amos Tversky. 1973. "On the Psychology of Prediction." *Psychological Review* 80 (4): 237 – 251.

Kersten, Daniel, Pascal Mamassian, and Alan Yuille. 2004. "Object Perception as Bayesian Inference." *Annual Review of Psychology*, 55: 271 – 304.

Lippmann, Walter. 1922. *Public Opinion.* New York, NY: Harcourt.

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, Lee Jussim. 2001. "Ethnic and National Stereotypes: The Princeton Trilogy Revisited and Revised." *Personality and Social Psychological Bulletin* 27 (8): 996 – 1010.

Mullainathan, Sendhil. 2002. "Thinking through Categories.", Working Paper, Harvard University.

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.

Ortoleva, Pietro, and Erik Snowberg. 2015. “Overconfidence in Political Behavior.” *American Economic Review* 105 (2): 504 – 535.

Phelps, Edmund. 1972. “The Statistical Theory of Racism and Sexism.” *American Economic Review* 62 (4): 659 – 661.

Pope, Devin and Justin Sydnor. 2010. “Geographic Variation in the Gender Differences in Test Scores.” *Journal of Economic Perspectives* 24(2): 95 – 108.

Rabin, Matthew. 2002. "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics* 117 (3): 775 – 816.

Rabin, Matthew and Joel Schrag. 1999. “First Impressions Matter: A Model of Confirmatory Bias.” *Quarterly Journal of Economics* 114 (1): 37 – 82.

Rabin, Matthew and Dimitri Vayanos. 2010. "The Gambler’s and Hot-Hand Fallacies: Theory and Applications" *Review of Economic Studies* 77 (2): 730 – 778.

Reuben, Ernesto, Paola Sapienza and Luigi Zingales. 2014. “How Stereotypes Impair Women’s Careers in Science.” *Proceedings of the National Academy of Sciences* 111 (12): 4403 – 4408.

Rosch, Eleanor. 1973. “Natural Categories.” *Cognitive Psychology* 4 (3): 328 – 350.

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.

Schneider, David, Albert Hastorf, and Phoebe Ellsworth. 1979. *Person Perception* (2nd ed.). Reading, MA: Addison-Wesley.

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

Shih, Margaret, Todd Pittinsky, and Nalini Ambady. 1999. “Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance.” *Psychological Science* 10 (1): 80 – 83.

Steele, Claude. 2010. *Whistling Vivaldi: How Stereotypes Affect Us and What We Can Do.* New York, NY: W. W. Norton & Company.

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

Weinberger, Catherin. 2005. “Is the Science and Engineering Workforce Drawn from the Far Upper Tail of the Math Ability Distribution?” Working Paper, UCSB.

Westfall, Jacob, Leaf Van Boven, John Chambers, and Charles Judd. 2015. “Perceiving Political Polarization in the United States: Party Identity Strength and Attitude Extremity Exacerbate the Perceived Partisan Divide." *Psychological Science* 10 (2): 145 – 158.

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}. It then follows that 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 $t \in \{1, \ldots, N\}$ 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 average, $\mathbb{E}(t|G) < \mathbb{E}(t)$ (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 the stereotype consists of types 1 through d. Then, for any $d < N$, the stereotype truncates types $\{d+1, \ldots, N\}$ (the upper tail). It follows 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 for $d < N$ the stereotype consists of types $N - d + 1$ through N. By truncating the lower tail, it follows that $\mathbb{E}^{st}(t|G) > \mathbb{E}(t|G)$. ■

Proposition 3. Without loss of generality, let the set of types be $T = \{1, \ldots, N\} \subset \mathbb{N}$. We first outline a number of useful properties that follow from the assumption that the distribution $\pi_{t,G}$ is symmetric over T. First, we have $\pi_{t,G} = \pi_{N-t+1,G}$ and $(t - \mathbb{E}[t'^2 = (N - t + 1 - \mathbb{E}[t'^2$ for all $t \in T$. Second, we have that $\mathbb{E}[t|G] = \mathbb{E}[t]$.

Suppose now that the likelihood ratio $\pi_{t,G}/\pi_{t,-G}$ is symmetric and U-shaped in t (case i)). Then, for all t, types t and $N-t+1$ are equally representative (and are distinct except if $t=(N+1)/2$), with more extreme types being more representative. By assumption, the stereotype includes the d most representative types, with all types tied for d^{th} ranking included, namely $\{1, N\} \cup \{2, N-1\} \cup \ldots \left\{\frac{[2]}{2}, N-\frac{[d]}{2}+1\right\}$, where $[d]$ is the smallest even number equal or greater than d. This coincides with the set of types $t \in \{1, \ldots, N\}$ that satisfy $t \leq \frac{d+1}{2}$ or $t \geq N-\frac{d-1}{2}$. In particular, the stereotypical distribution is a symmetric truncation of the (symmetric) distribution $\pi_{t,G}$, and therefore satisfies $\mathbb{E}^{st}[t|G]=\mathbb{E}[t|G]$.

Denote by $T_G(d)$ the set of types in the stereotype of G, and by $\overline{\pi}_G=\sum_{t \in T_G(d)} \pi_{t,G}$ the total (true) probability included in this stereotype. For simplicity, consider the case where N is even. Then,

$$
\begin{aligned}
Var^{st}(t|G) &= 2\sum_{t=1}^{\frac{d+1}{2}} \left(t-\mathbb{E}^{st}[t|G]\right)^2 \cdot \frac{\pi_{t,G}}{\overline{\pi}_G} > 2\sum_{t=1}^{\frac{d+1}{2}} (t-\mathbb{E}[t])^2 \cdot \pi_{t,G} + 2\sum_{t=\frac{d+1}{2}}^{\frac{N}{2}} \left(t_{[d]/2}-\mathbb{E}[t]\right)^2 \cdot \pi_{t,G} \\
&> 2\sum_{t=1}^{\frac{N}{2}} (t-\mathbb{E}[t])^2 \cdot \pi_{t,G} = Var(t|G)
\end{aligned}
$$

where the sum $\sum_{t=a}^{b}$ is taken over the largest subset of integers contained within the interval $[a,b]$, and $[d]$ was defined above.

Suppose now the likelihood ratio $\pi_{t,G}/\pi_{t,-G}$ is symmetric and inverse-U-shaped in t (case ii)). Again, types t and type $N-t+1$ are equally representative, but more extreme types are now less representative. Therefore, the stereotypical support $T_G(d)$ consists of all the types $t \in \{1, \ldots, N\}$ that satisfy $\frac{N+1-d}{2} \leq t \leq \frac{N+1+d}{2}$. Again assuming, for simplicity, that N is even, we find:

$$
\begin{aligned}
Var^{st}(t|G) &= \sum_{t=\frac{N+1-d}{2}}^{\frac{N+1+d}{2}} \left(t-\mathbb{E}^{st}[t]\right)^2 \cdot \frac{\pi_{t,G}}{\overline{\pi}_G} \\
&< \sum_{t=\frac{N+1-d}{2}}^{\frac{N+1+d}{2}} \left(t_{\frac{N+1-[d]}{2}}-\mathbb{E}[t]\right)^2 \cdot \pi_{t,G} + 2\sum_{t=1}^{\frac{N+1-d}{2}} (t-\mathbb{E}[t])^2 \cdot \pi_{t,G} \\
&< 2\sum_{t=1}^{\frac{N}{2}} (t-\mathbb{E}[t])^2 \cdot \pi_{t,G} = Var(t|G)
\end{aligned}
$$

The same proof goes through for N odd. ■

Lemma 1. 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 9 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 9 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. ■

Proposition 6. We assume that the same number of observations are received at each stage of the learning process for both groups G and $-G$. This assumption is not restrictive, since only the relative frequency of observations matters. In particular, all probabilities remain unchanged if the sample size of one group is scaled up relative to the sample size of the other. Thus we can set $\sum_{t'} a_{t',G} = \sum_{t'} a_{t',-G} = a$ and $\sum_{t'} n_{t',G} = \sum_{t'} n_{t',-G} = n$.

Representativeness of a type t is now measured by the ratio

$$\frac{Pr(X = x|\alpha_S, n_S)}{Pr(X = x|\alpha_{-G}, n_{-G})} = \frac{\alpha_{t,G} + n_{t,G}}{\alpha_{t,-G} + n_{t,-G}}$$

where $\alpha_S = (\alpha_{t,S})_{t \in T}$ are the priors for group S and $n_S = (n_{t,S})_{t \in T}$ is the sample for group S.

Consider case i) where all observations occur in type t, so that $n_{t,G} = n$ and $n_{t',G} = 0$ for $t' \neq x$, and similarly for $-S$. Then the representativeness of types other than t does not change, while the representativeness of t is $(\alpha_{t,G} + n)/(\alpha_{t,-G} + n_{t,-G})$. This tends to one monotonically as n increases. Therefore, if $a_{t,G}/a_{t,-G} < 1$ then $(a_{t,G} + n)/(a_{t,-G} + n) < 1$ for all n: namely, if t is non-representative to begin with, then no amount of observations of t in population G (when accompanied by observations of t in population $-G$) will make

t representative for G.

Consider now case ii), where all observations in G occur in a non-representative type t while all observations in $-G$ occur in a representative (for G) type t'. In that case, the representativeness of t for group G increases as $(a_{t,G}+n)/(a_{t,-G})$, while the representativeness of t' for group G decreases as $(a_{t',G}+n)/(a_{t',-G}+n)$. The result follows. ∎

Proposition 7. Consider the case where a single observation of group G occurring in type t does not change the representativeness ranking of types – and thus the stereotype – for G.

If t is in the stereotype of G, then its estimated probability is $a_{t,G}/\sum_{t'=1}^{d} a_{t',G}$, which is boosted by a factor of $\sum_{t'=1}^{N} a_{t,G}/\sum_{t'=1}^{d} a_{t',G} > 1$, where d is the number of types in the stereotype. Suppose an observation occurs in type t. Its representativeness for G increases, and its assessed probability jumps to $(a_{t,G}+1)/(\sum_{t'=1}^{d} a_{t',G}+1)$. This corresponds to a larger increase of assessed probability than that made by a Bayesian whenever

$$\frac{a_{t,G}+1}{\sum_{t'=1}^{d} a_{t',G}+1} - \frac{a_{t,G}}{\sum_{t'=1}^{d} a_{t',G}} > \frac{a_{t,G}+1}{\sum_{t'=1}^{N} a_{t',G}+1} - \frac{a_{t,G}}{\sum_{t'=1}^{N} a_{t',G}}$$

namely when

$$\frac{a_{t,G}}{\sum_{t'=1}^{N} a_{t,G}} < \frac{\sum_{t'=1}^{d} a_{t,G}}{1+\sum_{t'=1}^{d} a_{t,G}+\sum_{t'=1}^{N} a_{t,G}} < \frac{1}{2}$$

The intuition is that the stereotype ignores some observations, it is as though the probability is being updated over a smaller sample size. Therefore, as long as the prior of t (in the stereotype) is not too large, the DM boosts it more than the Bayesian.

If t is not in the stereotype, then – given that the stereotype does not change – it does not become representative. Its assessed probability stays at zero, so the decision maker under-reacts to this observation relative to a Bayesian. ∎

Supplementary Material For Online Publication

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

[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.

$\Pr(\mathit{Republicans}|t)/\Pr(\mathit{Democrats}|t)$ but also $\Pr(t|\mathit{Republicans})$.

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 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, $t = \mathit{Evolution}$ maximizes $\Pr(\mathit{Democrats}|t)/\Pr(\mathit{Republicans}|t)$ but not $\Pr(t|\mathit{Democrats})$. 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 Likelihood, Availability, and Stereotypes

As we discussed in Section 3.2, 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.[35] When stereotyping the occupation of a democratic voter, people think about

[35] 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

"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.

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

$$R_k(t,G) = \frac{\pi_{t,G}}{\pi_{t,-G}+k} = \frac{1}{\frac{1}{R(t,G)} + k \cdot \frac{1}{\pi_{t,G}}} \tag{7}$$

where $k \geq 0$ and $R(t,G)$ is representativeness as defined in Definition 1. In Equation (7), 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.[36]

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 (7) 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 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 (7), and also in (2), may be broadly interpreted as capturing the availability, rather than just the frequency, of type t for group G. Formally, in the model of learning of Section F, 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.

of stimulus salience, associative activation, selective attention, specific training, and priming".

[36] 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.

In this approach, availability is related to neglect of the correlation structure of information (as discussed in Section 3.2, the psychology of availability is beyond the scope of this paper).

The concrete implications of Equation (7) 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 μ_G, μ_{-G} respectively, and variance σ. 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:

$$k \cdot (t_{E,G} - \mu_G) \cdot e^{\frac{\left(t_{E,G}-\mu_G\right)^2+2(\mu_G-\mu_{-G})\cdot\left(t_{E,G}-\frac{\mu_G+\mu_{-G}}{2}\right)}{2\sigma^2}} = \mu_G - \mu_{-G} \tag{8}$$

The left hand side of (8) 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 δ (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.[37]

D 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

[37] In the extended model given by (7), the parameters δ and k capture two natural types of bounds on recall: δ determines "how much" comes to mind (which might depend on effort), while k corresponds to the relative weight of likelihood in recall, which may vary across people.

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:

$$R_G(t_1, t_2) \equiv \frac{\pi_{(t_1,t_2),G}}{\pi_{(t_1,t_2),-G}} = \frac{\pi_{t_1,G}}{\pi_{t_1,-G}} \cdot \frac{\pi_{t_2,(G,t_1)}}{\pi_{t_2,(-G,t_1)}}. \tag{9}$$

where $\pi_{t_2,(G,t_1)} = \Pr(t_2|G, t_1)$. In light of Equation (9), 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.

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).[38]

[38] Here the stereotype allows for all possible values of t_2 because of the tie breaking assumption in Definition

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 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:

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

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

[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.

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.

E 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.

E.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 3 *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 in 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 δ is between 0 and 1, we are in an intermediate case.

Definition 4 *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.

E.2 The Normal Case

When $f(t|G)$ and $f(t|-G)$ are univariate normal, with means μ_G, μ_{-G} and variances σ_G, σ_{-G}, the stereotype of G is easy to characterize.

Proposition 4 *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 δ. Moreover, $E^{st}(t|G) > \mu_G > \mu_{-G} > E^{st}(t|-G)$.

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

Proof. Let ρ_{μ,σ^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 4). 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 7, 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.

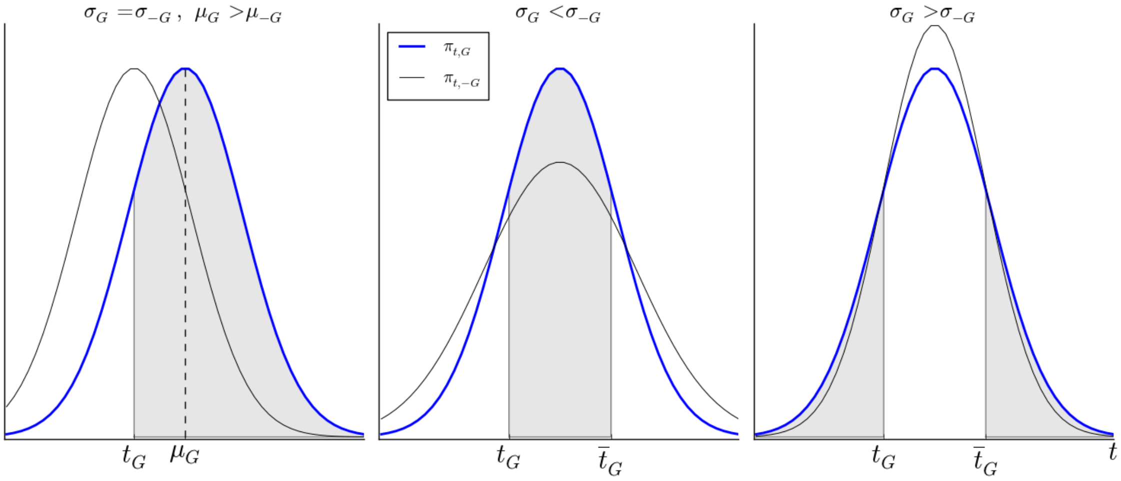


Figure 7: Stereotypes of a Normal distribution as a function of μ_{-G} and σ_{-G}.

Consider now case ii), where the variance of G is lower than that of $-G$, Figure 7, 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 μ_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 μ_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 7, 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 $t^K = (t_{G,k}, t_{-G,k})_{k=1,\dots,K}$.

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 ε_W is i.i.d. normally distributed with mean 0 and variance v, and θ_W is the group specific mean. Initially, groups are believed to be identical, in the sense that both θ_G and θ_{-G} are normally distributed with mean 0 and variance γ. After observing $(t_{G,1}, t_{-G,1})$, the distribution of θ_W is updated according to Bayesian learning. Updating continues as progressively more observations are learned. Thus, after observing the sample t^K, we have:

$$f(t|W, t^K) = \mathcal{N}\left(\frac{\gamma \cdot K}{v + \gamma \cdot K} \cdot \frac{\sum t_{W,k}}{K}; v \cdot \frac{v + \gamma \cdot (K+1)}{v + \gamma \cdot K}\right). \tag{10}$$

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 θ_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 4 implies:

Proposition 5 *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 4, 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, the process of 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. On the other hand, even small pieces of information can cause a strong over-reaction if they reverse the ranking between group averages.

F Stereotypes and Reaction to New Information

Stereotypes are hard to change, but they are far from immutable. For instance, stereotypes of immigrant populations change over time: in the early 20th century US, European Jews were stereotyped as religious and Asian immigrants were stereotyped as uneducated, yet both groups are stereotyped as high-achievers at the beginning of the 21st (Madon et. al., 2001). More recently, a rapid increase in the share of female doctors has coincided with shifting gender stereotypes in the medical profession. Medicine has historically been perceived as a stereotypical male profession, with women being viewed as less competent than their male counterparts (Decker 1986). However, this stereotype has faded, with specialties where women are more prevalent, such as pediatrics and dermatology, now being viewed as gender neutral (Couch and Sigler 2001). These patterns reflect at least in part changes in stereotypes in response to changes in reality. In fact, the experimental psychology literature documents that stereotypes change when individuals are faced with sufficiently pressing disconfirming information (Schneider 2004).

Our model can be naturally extended to investigate how stereotypes and beliefs change by the arrival of new information over time. To explore these dynamics, we suppose that at the outset, unlike in Section 3, the decision maker does not have perfect information about the categorical distribution $(\pi_{t,G})_{t=1,\dots,N}$ of the group G of interest, or about the distribution $(\pi_{t,-G})_{t=1,\dots,N}$ of the comparison group $-G$. Instead, the DM has priors over these distributions that are described by the Dirichlet distribution:

$$g\left[\pi_{t,W},\alpha_{t,W}\right]_{t=t_1,\dots,t_N} = \frac{\Gamma\left(\sum_t \alpha_{t,W}\right)}{\Pi_t \Gamma(\alpha_{t,W})} \cdot \Pi_t \pi_{t,W}^{\alpha_{t,W}-1} \ , \qquad \text{for } W = G, -G,$$

which are conveniently conjugate to the categorical distributions assumed so far. Parameters $\alpha_G = (\alpha_{t,G})_{t=t_1,\dots,t_N}$ and $\alpha_{-G} = (\alpha_{t,-G})_{t=t_1,\dots,t_N}$ pin down the prior expectations of a Bayesian agent:

$$\Pr(T = t|\alpha_W) = \mathbb{E}(\pi_{t,W}|\alpha_W) = \frac{\alpha_{t,W}}{\sum_u \alpha_{u,W}}, \qquad \text{for} \qquad W = G, -G. \tag{11}$$

In contrast to the Bayesian agent, the stereotype initially held by the DM depends on the probabilities in Equation (11) according to Definition 1. For simplicity, we set $\sum_t \alpha_{t,G} = \sum_t \alpha_{t,-G}$.

Suppose that a sample $n_W = (n_{1,W}, \dots, n_{N,W})$ is observed, where $n_{t,W}$ denotes the observation count in type t and let $\sum_t n_{t,W}$ be the total number of observations for group W. Then, a Bayesian's posterior probability of observing t is

$$\Pr(T = t|\alpha_W, n_W) = \mathbb{E}(\pi_{t,W}|\alpha_W, n_W) = \frac{\alpha_{t,W} + n_{t,W}}{\sum_u(\alpha_{u,W} + n_{u,W})}, \tag{12}$$

which is a weighted average of the prior probability of Equation (11) and the sample proportion $n_{t,W}/n_W$ of type t. As new observations arrive, the probability distribution in group

W, and thus stereotypes, are updated according to Equation (12).[39]

Consider how a DM influenced by representativeness updates beliefs. Proposition 6 describes how new information changes the set of types that come to mind, shedding light on when and how stereotypes change. Proposition 7 considers the effect of information on probability assessments on types that are already included in the stereotype.

Proposition 6 *Suppose that the DM observes the same number of realizations from both groups, formally* $\sum_u n_{u,G} = \sum_u n_{u,-G} = n$. *Then:*

i) If for both groups all observations occur on the same type t *that is initially non-representative for* G*, then this type does not become representative for* G*. Formally, if* $n_{t,G} = n_{t,-G} = n$ *for a type* t *such that* $\alpha_{t,G}/\alpha_{t,-G} < 1$*, then* $\Pr(T = t|\alpha_W, n_G)/\Pr(T = t|\alpha_W, n_{-G}) < 1$ *for all* n.

ii) If all observations for G *occur in a non representative type for* G*, while those for* $-G$ *occur in a type that is representative for* G*, then for a sufficiently large number of observations the stereotype for* G *changes. Formally, if* $n_{t,G} = n$ *for a type* t *such that* $\alpha_{t,G}/\alpha_{t,-G} < 1$*, while* $n_{t',-G} = n$ *for a type* t' *such that* $\alpha_{t',G}/\alpha_{t',-G} > 1$*, then for* n *sufficiently large* $\Pr(T = t'|\alpha_W, n_G)/\Pr(T = t'|\alpha_W, n_{-G}) < 1 < \Pr(T = t|\alpha_W, n_G)/\Pr(T = t|\alpha_W, n_{-G})$.

The stereotype for a group does not necessarily change if the new observations are contrary to the initial stereotype. The stereotype is modified only if the new information for G and $-G$ is sufficiently different, as in case ii). Only when a disproportionate number of non-stereotypical observations occur for group G do these previously neglected types become sufficiently more likely in, and thus representative of, G.

Proposition 6 describes how much contrary data is needed in order to change a stereotype. For example, a process of economic development can significantly improve the livelihoods of all groups in a population, but it does not dispel the negative stereotype of a group that still includes a disproportional share of low income households. Instead, if a (subset of a)

[39] While we assume for simplicity that updating is Bayesian, the representativeness mechanism that links priors to stereotypes can naturally be coupled with a non-Bayesian updating process. Psychologists have documented a tendency to search for information that confirms one's beliefs (Lord, Ross and Lepper 1979, Nickerson 1998). Schwartzstein (2014) proposes a model of biased learning in which information is used to update beliefs only about dimensions that are attended to.

negatively stereotyped group outperforms the overall population, then its stereotype can change. Attitudes towards certain immigrant groups (Jews, Asians) have changed only as a subset of them overtook the overall population in terms of socioeconomic status. Attitudes towards women in the medical profession have changed only after a dramatic catch up in the number of female doctors, particularly in specialties in which female doctors are as frequent as male doctors (Noori and Weseley 2011).

We now consider how the initial stereotype for group G (formally, the priors over G and $-G$) affects the way in which the DM processes new information about G. We assume the support of the stereotype for G is fixed, and explore how the DM reacts to further information about G (note that, once the support of G's stereotype is determined, information about $-G$ plays no role in determining the stereotypical distribution).

Proposition 7 *Let $d > 1$. Suppose that one observation about type t is received in group G (formally, $n = n_{t,G} = 1$). Then:*

i) If t belongs to the stereotype of G and its probability is sufficiently low, the DM over-reacts (relative to the Bayesian) in revising upward his assessment of t's probability. Formally, there is a threshold $\nu \in (0, 1/2)$ such that the DM's assessed probability of t increases by more than under Bayesian updating if and only if $\alpha_{t,G}/\sum_u a_{u,G} < \nu$.

ii) If t does not belong to the stereotype of G, the DM does not update its probability at all, so he under-reacts relative to the Bayesian DM.

Proposition 7 indicates that stereotypes can both over and under-react to information. In case i), the DM strongly over-reacts to information confirming the stereotype. Intuitively, because the DM neglects non-representative types, he does not fully account the current observation may be due to sampling variability. As a consequence, his beliefs overreact when a type he does attend to is confirmed by the data. If criminal activity is part of a group's stereotype, the DM over-reacts to seeing a criminal from that group and his judgments become even more biased against the group. If a growth company generates surprisingly positive earnings, investors further upgrade their belief that the stock is a good investment, because they neglect the possibility that an extreme observation may be due to noise.

At the same time, case ii) shows that the DM under-reacts (relative to a Bayesian) to information inconsistent with the stereotype. This is because insofar as the stereotype is unaffected, the probability of a non-stereotypical type is not upgraded, as the type remains neglected in the assessment of the group. This is the main departure of our model from the kernel of truth logic: the agent discards some news because these are not strong enough to overturn his prior belief.

Upon observing a highly successful member of a group stereotyped as low socioeconomic status, DMs code the occurrence as an "anomaly" and continue to believe that the group at large should be viewed through the lens of the negative stereotype. People can espouse racist views and yet be friendly with individual members of the group they disregard (Schneider 2004). However, as shown in Proposition 7, non-stereotypical information is often ineffective at changing beliefs even if it swamps the few instances underlying the stereotype.[40]

Proposition 7 implies that the DM exhibits a type of confirmation bias (Lord, Ross and Lepper 1979, Nickerson, 1998). Faced with two observations of different types from group G (formally, $n_{t,G} = n_{t',G} = 1$ and $n = 2$), such that t belongs to the stereotype of G but t' does not, the DM over-reacts to information consistent with the stereotype and ignores information inconsistent with it.[41] In this way, our approach provides a unified mechanism that gives rise to both base-rate neglect and confirmation bias: base-rate neglect arises when representative types are unlikely, while confirmation bias arises when new information does not change representativeness and allows stereotypes to persist. In the context of representativeness-based predictions, these biases are two sides of the same coin. The approach can also unify several other biases, such as overconfidence but also – under

[40] Propositions 6 and 7 formalise features of psychological models of stereotype change, such as the conversion and sub-typing models (Rothbart 1981, Weber and Crocker 1983), in which disconfirming evidence is treated as "exceptions to the rule" up to the point when it becomes sufficiently pressing, and engenders a stereotype change. In the experimental psychology literature on stereotype change, types are usually unordered and multidimensional (e.g., ethnicities are defined in terms of socio-economic status, proneness to violence, musical tastes, etc), see Hewstone et al (2000) and references therein. This literature finds that stereotypes fail to change when some extreme disconfirming evidence is observed (i.e. group elements that violate the stereotype along every dimension). In contrast, when sufficiently many group members are observed that violate the stereotype only along a given dimension, but are otherwise representative, those observations may become representative of the group as a whole and change the stereotype. Consistent with Proposition 6, stereotype change in these experiments seems driven by changes in (relative) frequency, and not by the shock of observing extreme exceptions.

[41] Lord, Ross and Lepper (1979) suggest that confirmation bias arises in response to information that provides ambiguous support to an underlying hypothesis. Here, the information provided received presents ambiguous support for the stereotype.

appropriate extensions – polarisation effects.[42]

G Experiments

G.1 Analysis of All Unordered Types Experiments

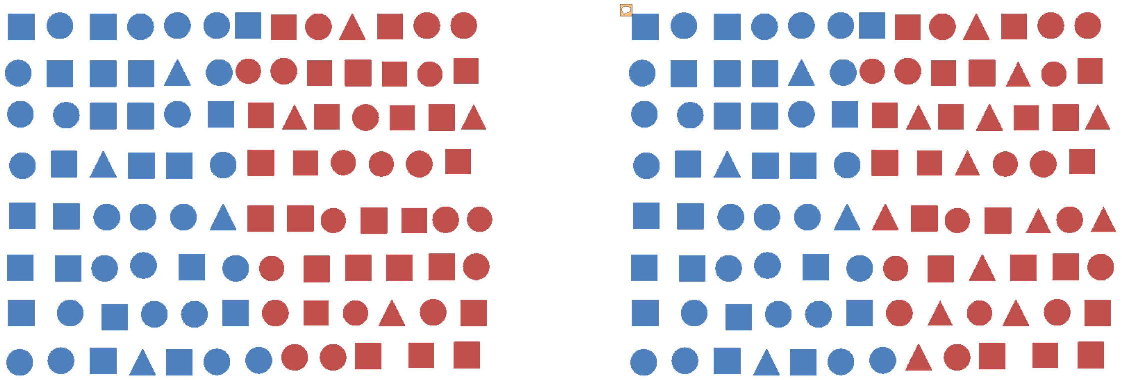
We conducted four experiments on unordered types. The final experiment, using cartoon characters in T-Shirts, were reported in the main text. Here we discuss the other experiments and their results.

G.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 8.

This design is not as clean as the T-shirts design presented in the paper. Most impor-

[42] Polarization arises as a consequence of confirmation bias when DMs have heterogeneous priors. Proposition 6 then implies that a given set of observations can lead different DMs with different stereotypes to each reinforce their own stereotype, and thus update in opposite directions.



(a) Control Condition (b) Representativeness Condition

Figure 8: 50 Shapes Experiment

tantly, 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.

This shapes experiment was conducted on MTurk in November 2014 with 217 participants.[43] 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.

[43] 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.)

G.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 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.[44]

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).[45] 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

[44] 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.

[45] 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.

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).

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.[46]

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

[46] 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.

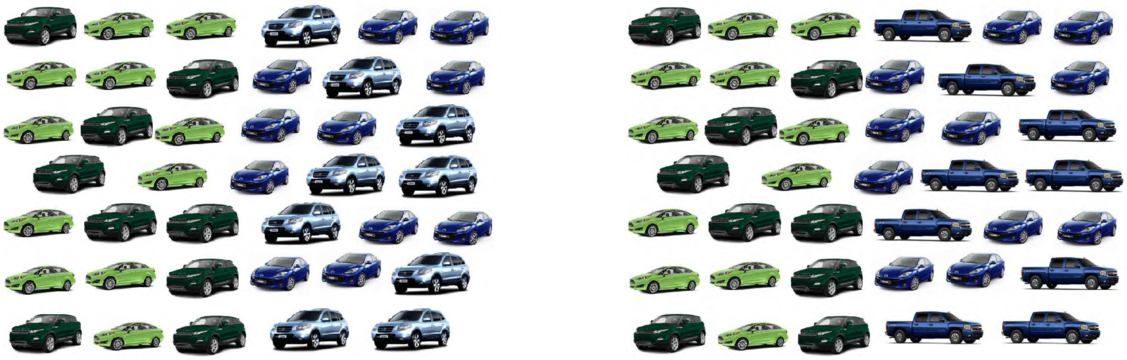
of our prediction, though it is not significant.

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 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. /textbfIn 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.

G.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 9.



(a) Control Condition

(b) Representativeness Condition

Figure 9: Vehicles Experiment

We conducted this experiment with 57 participants on MTurk in January 2015.[47] 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

[47] 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.

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 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.

G.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.[48] 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

[48] 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.

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 subsample, lab or MTurk.

G.1.5 Summary of Unordered Types Experiments

Table 5 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).

G.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

Table 5: Summary of All Unordered Types Experiments

Experiment	Brief Description	# of Participants	Percentage Point Increase in Probability of Believing Less Common Type is Modal when it is Representative	p-value
1	(Lots of) Shapes on MTurk	217	5.6 pp	0.17
2	(Fewer) Shapes Pooled	166	1.4 pp	0.84
	MTurk Only	100	5.6 pp	0.49
	Lab Only	66	-13.8 pp	0.28
3	Vehicles on MTurk	57	17.4 pp	0.09
4	T-Shirts Pooled	301	10.5 pp	0.014
	MTurk Only	105	7.6 pp	0.25
	Lab Only	196	14.4 pp	0.007
Pooled		741	9.3 pp	0.002

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 6: 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

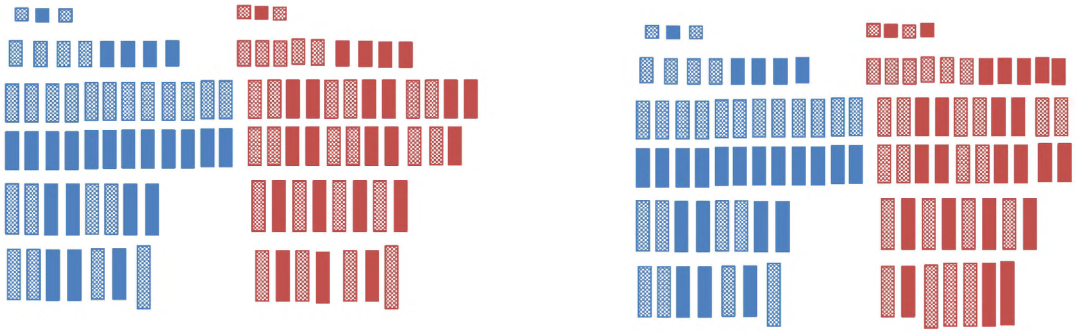
set of results.

G.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 6 displays the distribution, and Figure 10 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 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.

We chose to arrange the rectangles by height for participants so that it might be easier



(a) Control Condition

(b) Representativeness Condition

Figure 10: Rectangles Experiment

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.[49] 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.[50] Participants were randomly-assigned to view either the Control rectangles or the Rep. rectangles for 15 seconds. Then, they completed simple addition 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 par-

[49] 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.

[50] 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.

ticipants 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.

G.2.2 Ordered Types Experiment 2: Ice Cream

In the main text, we reported the results of the ice cream cones experiment. Here, we discuss those results in more detail. 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.

We ran this new simplified ice cream cone design on MTurk in January 2015 with 65 participants.[51] 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 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

[51] 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.

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. While an estimated treatment effect of zero would be consistent with noisy data or confused participants, a directional effect in the wrong way 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.[52] 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

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

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

G.2.3 Summary of Ordered Types Experiments

Table 7 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 11.7 percentage points more likely to believe that the shorter group is taller on average when it has a representative tall tail. If we include, in addition, all color blind participants, this estimate is 11.9 percentage points (p=0.022); and, if we include all observations, including participants who have participated in multiple versions of the experiment, this estimate is 10.4 percentage points (p=0.040). 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, -3.5 pp (p=0.51). This difference in treatment effect across platform is significant (p=0.002).

Table 7: Summary of All Ordered Types Experiments

Experiment	Brief Description	# of Participants	Percentage Point Increase in Probability of Believing Shorter Group is Taller on Average in Rep. Condition	p-value
1	Rectangles	179	7.3 pp	0.32
	Rectangles on MTurk	113	19.1 pp	0.04
	Rectangles in Lab	66	-11.0 pp	0.38
2	Ice Cream	223	9.2 pp	0.17
	No Decoy Ice Cream in Lab	56	-12.0 pp	0.37
	Ice Cream on MTurk	65	30.7 pp	0.02
	Ice Cream with Decoy in Lab	101	3.5 pp	0.73
Pooled		402	11.7 pp	0.03

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.

H Empirical Analysis: Further Results

H.1 Truncation model on WBCJ dataset

H.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 10. In the GSN data, Liberals hold more exaggerated beliefs about both Conservatives and Liberals than Conservatives do. The pattern is different in the WBCJ data. Conservatives in the WBCJ data have exaggerated beliefs about Liberals, but not about Conservatives. Liberals in the WBCJ 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

Table 8: Prediction Errors of Representativeness-Based Model for WBCJ Data

Representativeness Model: Truncation to d 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 WBCJ 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 WBCJ 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 9: Prediction Errors of Likelihood-Based Model for WBCJ Data

Likelihood Model: Truncation to d 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 WBCJ 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 WBCJ 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

Table 10: 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
GSN	0.35	0.71	0.03	0.21
WBCJ	-0.11	0.18	0.78	0.36

and the other group.

Next, we test Specification I, asking whether we observe the same context-dependence for beliefs held by either group. In Table 11, 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 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 12, we explore Specification II, asking whether ARTP 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 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.

Table 11: 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	1.13****	0.92****	-0.36****	-0.23****
Conservatives	(0.076)	(0.087)	(0.073)	(0.065)
True Mean	-0.55****	-0.65****	0.68****	0.81****
Liberals	(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 WBCJ data set.

Table 12: 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	0.71****	0.42****	0.36****	0.58****
of G	(0.09)	(0.10)	(0.10)	(0.10)
$ARTP_G$	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 WBCJ data set. One liberal observation is missing from the GSN data as there is no mass on stereotypical liberal positions for either group for one issue.