

Explaining Differences in Free Riding Beliefs

A Review and Preliminary Theory

Charlotte Cavallé

(Ford School of Public Policy, University of Michigan)

People who oppose generous social benefits for the poor and the unemployed often believe that recipients are free riding, i.e., abusing society's generosity by failing to act to improve their plight. People who support more generous social benefits express the opposite concern: the moral wrong is on the side of society, which does too little to help people who cannot be blamed for their economic conditions. What explains these differences in welfare attitudes and free riding beliefs? To answer this question, this paper focuses on the well-documented, yet puzzling, correlation between 1) free riding beliefs on the one hand and 2) liberal-authoritarian values on the other. Underpinning this correlation, I hypothesize, is a disagreement over how to best address social dilemmas, i.e. how to maximize pro-social behavior and minimize free riding. I provide preliminary evidence for this line of inquiry and conclude by discussing implications for future research.

Research shows that policy preferences correlate with empirical beliefs about important features of the world. For example, individuals who oppose income redistribution are more likely to believe that differences in market income reflect differences in effort and talent. Alongside perceptions of economic fairness, researchers have documented the existence of a second category of beliefs also important for social policy preferences, namely beliefs about the prevalence of free riding among net beneficiaries of redistributive social policies. These latter beliefs co-vary with support for policies that benefit those who cannot (temporarily or permanently) provide for themselves: the higher the perceived prevalence of free riding, the lower the support for such transfers.

As shown by Cavaille (2021), beliefs about the prevalence of free riding have little to do with one's economic conditions. Instead, they correlate with what researchers call "liberal-authoritarian values" (LAVs), an umbrella term used to describe individual differences in preferred forms of governance broadly defined. People with more liberal values support forms of governance "with broader participation of members and more subjects left to personal autonomy" while people with more authoritarian values support forms of governance "with less participation and a broader realm of subjects covered by binding codes of conduct" (Kitschelt and Rehm 2014). Items commonly used to construct LAV indices include items asking about stiffer sentences, the death penalty and child rearing practices (Hetherington and Weiler 2005). Differences in LAVs also predict differences in attitudes toward non-economic issues such as gay marriage or the decriminalization of marijuana .

Why are answers to questions about sentencing, and discipline in school informative of answers to questions about the deservingness of able-bodied beneficiaries of social solidarity? Why does an emphasis on personal autonomy predict both support for gay rights and the belief that welfare recipients are deserving of benefits? For Häusermann and Kriesi (2015), the correlation between welfare attitudes and free riding beliefs on the one hand and LAVs on the other suggests a blurring of the boundaries between economic and non-economic attitudes. In contrast, I propose and test a theoretical framework in which this correlation is to be expected, shedding a new light on the mechanisms underpinning the formation of welfare attitudes and free riding beliefs.

To understand differences in free riding beliefs, researchers currently draw on three non-competing lines of inquiry. One, most prominent in social psychology, emphasizes the role of ideological difference in poverty attribution, namely the tendency for conservatives to prefer dispositional ex-

planations and for liberals to prefer situational explanations for social problems (Morgan, Mullen and Skitka 2010). Another emphasizes the importance of racial stereotypes and anti-immigrant sentiment (Alesina and Stantcheva 2020). A third builds on research in sociology on boundary formation and stigmatization to argue that status threat shapes perceptions of the poor and the unemployed (Lamont and Molnar 2002; Sennett and Cobb 1993). Of these three lines of reasoning, none can account for the correlation between free riding beliefs on the one hand and LAVs on the other. Furthermore, as I show, these theories hinge on empirical patterns that are themselves an artifact of this robust correlation.

Underpinning this correlation between free riding beliefs and LAVs, I argue, is a more fundamental disagreement over how to best address social dilemmas, i.e. how to maximize pro-social behavior and minimize free riding. A social dilemma is a situation in which everyone would be better off cooperating, yet the dominant strategy is to defect.¹ The evolutionary psychologist Jonathan Haidt argues that people fundamentally differ in how they intuitively reason about social dilemmas, cooperation and free riding. He identifies two ideal-typical approaches. One emphasizes the need to “suppress selfishness” by “strengthening groups and institutions” and “binding individuals into roles and duties in order to constrain their imperfect natures” (Graham, Haidt and Nosek 2009: 1030). The other approach focuses on encouraging prosocial behavior by “protecting individuals directly (...) and by teaching individuals to respect the rights of other individuals” (Turiel 1983; Shweder et al. 1997). Haidt’s argument suggests individual-level differences with regards to approaches to the monitoring of free riding and moral hazard concerns. These differences, I argue, underpin the correlation between free riding beliefs and LAVs. First, people with a binding worldview are more likely to assume that “bad intentions” are more prevalent than “good ones” and to favor policing and education strategies that limit individual autonomy, explaining their more authoritarian response patterns to questions about order and hierarchy. Second, they are also more likely to worry about moral hazard and to overweight evidence of free riding, resulting in comparatively more pessimistic empirical beliefs regarding the prevalence of free riding. One reason, explored in detail in this paper, is the existence of individual differences in what I call *error sensitiv-*

¹ Indeed, if most people are not cooperating, then any form of pro-social behavior is mostly wasted effort: better to behave selfishly like everyone else. However, if everyone else is cooperating, then it is still more advantageous to behave selfishly and let others do the hard work while still reaping the benefits of collective action and cooperation.

ity: people with more authoritarian values prefer to minimize instances in which someone gets an undeserved reward even if it means increasing instances in which someone receives an undeserved punishment, while people with more liberal values have the opposite preference.

Using observational and experimental survey data collected in Great Britain, I find tentative evidence for this argument. First, I provide preliminary evidence that people differ in terms of error sensitivity. Second, I experimentally manipulate the perceived cost of free riding (an undeserved reward) by priming respondents to think about the chronic under-funding of the National Health Service. I then ask respondents who they think should be denied priority access to the NHS, if anyone. I find that, under scarcity, people are more likely to exclude not only illegal immigrants and non citizens, but also smokers and people who have not regularly paid taxes because of weak labor market attachment. The treatment effect is larger for people who find undeserved rewards worse than undeserved punishments.

This paper makes two contributions to research on welfare attitudes and free riding beliefs. First, it provides a critical review of the existing literature on the formation of free riding beliefs, drawing attention to the overlooked, yet substantively important, correlation between free riding beliefs and LAVs. Second, it builds on this correlation to propose a novel line of inquiry for investigating how free riding beliefs form. A full test of the argument is beyond the scope of this paper. Still, while tentative, the evidence provided in this paper is promising: the framework has more explanatory power than existing theories and generates new predictions such as the relationship between LAVs and the willingness to exclude smokers or low contributors in conditions of scarcity. In the conclusion, I discuss next steps for future research including how to expand the framework to include group boundaries and its relationship to the study of ideological dimensions in Western electoral democracies.

1 Welfare Attitudes and Free Riding Beliefs: Existing Research

Research shows that attitudes toward programs that redistribute to the worse off are conditional on the belief that beneficiaries are “deserving.” Deserving recipients are those who are in need for not fault of their own: however hard they try, the conditions they face make it difficult for them

to improve their economic conditions. Undeserving recipients are those who free ride on shared resources by failing to act to become self-reliant (Gilens 1999; Petersen et al. 2012; Sniderman, Tetlock and Brody 1993).² Deservingness evaluations can be thought of as a classification exercise: to what extent can someone be classified as a free rider abusing society's generosity or as a cooperator simply down on their luck? Research shows that most people share the same understanding of what behaviors classify a given individual as deserving or not.³ Furthermore, they have a natural inclination, and even feel the duty, to help the "deserving" and punish the "undeserving" (Fong, Bowles and Gintis 2006).

How does the moral impulse to help the deserving cooperator and punish the undeserving free rider manifests itself in the realm of redistributive policies? If people were third party judges jointly presiding over the cases of specific individuals whose behavior and inner travails are known to all involved, we would expect some modicum of agreement: judges that rule by applying the same criteria and principles using the same background information tend to come to similar conclusions. In practice, however, people never get to directly decide who should receive benefits and who should not: all they can do is form an opinion on the status quo. In this context, they only get to apply these principles and criteria to an imagined and socially constructed "modal recipient." One implication is that agreement over what *ought to be* –help deserving cooperators and punish undeserving free riders– coexists with disagreement over what *is* –i.e. the extent to which the status quo rewards free riders or not–. People who dislike the status quo and support less generous social benefits for the poor and the unemployed tend to believe that most recipients are free riding, i.e., abusing society's generosity by failing to act to improve their plight. In contrast, people satisfied with the status quo or people supportive of more generous benefits are not concerned about free riding. To them, "being on welfare" is not a choice or evidence of a moral failure: if there is a moral wrong, it is on the side of society, which does too little to help people who cannot be blamed for their economic conditions.

² Deservingness evaluations are often studied alongside attitudes toward means tested program but they matter more broadly. For example, support for universal access to public healthcare based on need, not past payroll contributions, is also shaped by the perception that the sick are not responsible for their plight (Jensen and Petersen 2014) and that people who contribute the least are "trying their best" given the constraints they face.

³ When given the same detailed information about a *specific* individual's effort level and the constraints he or she faces, people tend to reach the *same* conclusion regarding this person's deservingness and feel compelled to do "something" about it, i.e. (Petersen et al. 2012).

Why do people reach such different conclusions? In the next section, I review the different ways researchers across the social sciences have approached this question, with an emphasis on theories that help explain individual-level differences. As a result, I will overlook research documenting the impact of contextual factors such as the unemployment rate or media discourse on aggregate free riding beliefs (Gilens 1999; Uunk and van Oorschot 2019), focusing instead on theories helping explain why despite similar unemployment or media contexts people nevertheless form different assessments of the prevalence of free riding among the poor and the unemployed.

The Ideo-Attribution Effect

Psychologists have long noted that people differ in how they interpret causality when observing the actions of others and that these differences overlap with ideology broadly defined. This is especially striking when it comes to issues such as crime, wealth concentration, poverty or even AIDS infection: while “conservatives” explain negative outcomes by “referencing dispositional causes, such as people’s lack of will power, personal discipline, self-reliance, or diminished moral standards,” “liberals” explain the same outcomes “by appealing to unjust social practices and structures” (Skitka and Washburn 2016: 78). Dominant explanations of what scholars have called the *ideo-attribution effect* point to stable individual differences in reasoning styles that run deep into people’s personalities and even brain structures (Jost et al. 2003; Amodio et al. 2007). Simply put, people who tend to perceive the causes of behavior as residing mostly within people are more attracted to conservative political ideas and dispositional explanations of social problems. People who tend to interpret behavior as situational and constrained are more attracted to liberal political ideas and structural explanations of social problems.

While plausible, this line of reasoning suffers from important limitations. First, as convincingly demonstrated by Morgan, Mullen and Skitka (2010), conservatives are quick to emphasize external constraints and underplay dispositional explanations if they feel positively toward the people whose behavior is being scrutinized. For example, when asked to make attributions for the behavior of policy officers or Marines, conservatives made stronger situational attributions than liberals, resulting in a reversal of the ideo-attribution effect (see Skitka and Washburn (2016) for a review). These results are inconsistent with the assumption of stable individual differences in the propensity

to hold people personally responsible for their actions. Instead, they suggest that people make attributions based on whether they want to paint a given individual or group as responsible for their actions, and deserving of the positive or negative outcome associated with them, or irresponsible and thus undeserving of associated outcomes. To understand why some people are motivated to paint the poor and the unemployed in a bad light, while others are motivated to paint them in a good light, we need to turn to other lines of inquiry, including anti-immigrant sentiment or status threat reviewed below.

A second limitation of the ideo-attribution effect literature is its inability to account for important empirical correlations, especially in the European context. Studies of the ideo-attribution effect tend to focus on the United States and use respondents' subjective ideological identification as liberal or conservative to identify where people fit on the ideological spectrum. This empirical strategy is trickier in a European context. Indeed, decades of empirical work have shown that mass opinions are structured along an economic/redistributive dimension and a non-economic/cultural dimension: people who could be identified as liberal on economic and redistributive issues might nevertheless be identified as conservative on non-economic and cultural issues (Kitschelt 1997).⁴ Based on the ideo-attribution effect literature, one might expect differences in free riding beliefs to overlap with the economic/redistributive dimension. Indeed, people who tend to interpret behavior as dispositional are plausibly more likely to believe that the the rich should not be taxed (they “deserve” their wealth) and the poor should not be helped (they “deserve” their poverty). In contrast, people who emphasize structural factors are more likely to find the poor deserving and the rich undeserving and consequently more likely to support income redistribution and generous social transfers.

Yet, recent research shows that the structure of economic ideology is more complicated. Against all expectations, free riding beliefs are only poorly correlated with survey items traditionally used to measure economic/redistributive ideology. Instead, as mentioned in the introduction, free riding beliefs better correlate with survey items used to measure differences in non-economic/cultural ideology, specifically differences in liberal-authoritarian values. Figure 1 illustrates this point. Using

⁴ In some countries, electoral competition, and the party system more generally, roughly reflect this two-dimensional structure. In other countries, partly due to differences in electoral rule, the latter is subsumed into a single left/right dimension leaving some voters better represented than others (Rodden 2018).

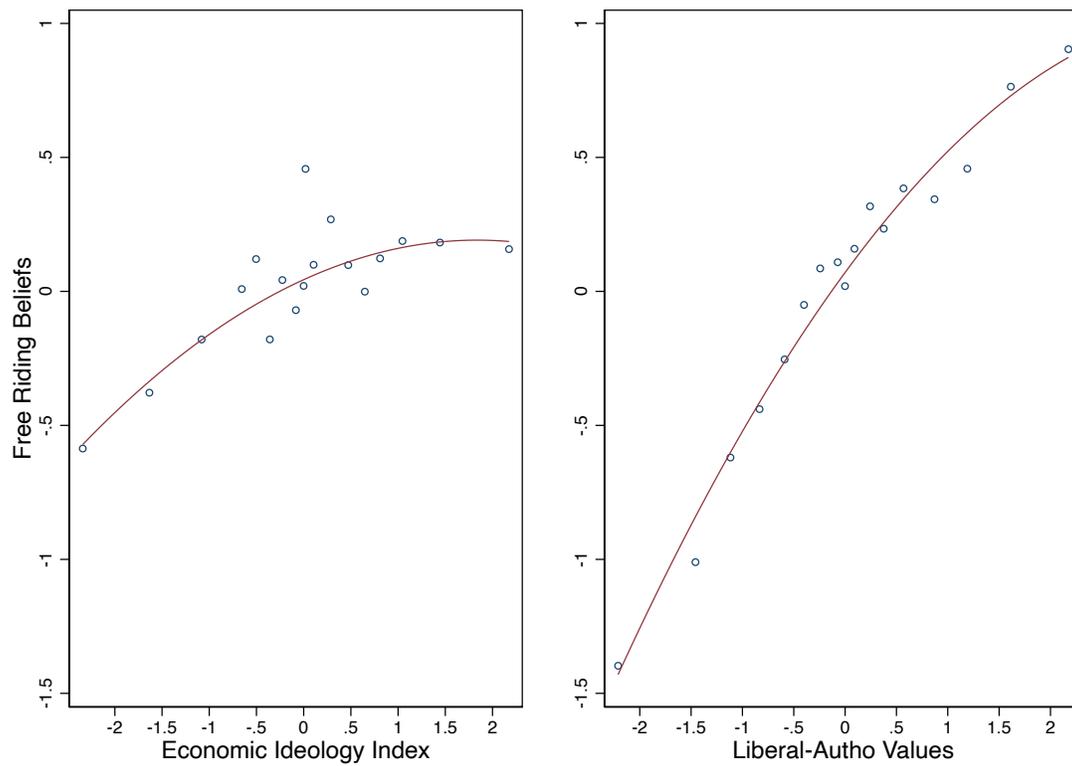
binned scatter plots, it visually represents the correlation between free riding beliefs, LAVs and economic ideology as traditionally measured in the literature. To measure free riding beliefs, I combine answers to the 8 survey items listed in Table 1. These items ask about the perceived work ethic of the unemployed (whether due to “laziness” or moral hazard), beliefs about the prevalence of shirking and the perceived deservingness of welfare recipients. I also include two items that ask about attitudes towards social transfers to the unemployed and the poor. Both an exploratory and confirmatory factor analysis show that all items load on the same latent variable. I consequently fit an Item Response Model to the data and use it to score individual on this latent variable: the higher the value the more someone perceives recipients as free riders whose benefits need to be cut. To measure economic/redistributive ideology and LAVs, I follow a similar procedure. Note that results are substantively the same if I combine items into separate additive indices.

As shown in Figure 1, answers to questions about income redistribution, big business and working people are much less informative of free riding beliefs than answers to questions about the death penalty and discipline in schools. Notice also how any evidence of a correlation between free riding beliefs and economic ideology is mostly driven by a small subset of left-wing respondents (score 1 SD or more below the mean). For the rest of the sample, more anti-redistribution economic attitudes does not translate into more anti-redistribution free riding beliefs. In contrast, the relationship between LAVs and free riding beliefs is positive and monotonic. These patterns contradict simple expectations derived from the literature on the ideo-attribution effect. Can alternative lines of inquiry do a better job?

Anti-Immigrant Sentiment

To the reader familiar with attitudinal survey data, the correlation between free riding beliefs and LAVs suggests turning one’s attention to a third attitudinal construct, namely anti-immigrant sentiment. Simply put, free riding beliefs are mostly capturing hostility to immigrants, perceived to be over-represented among benefit recipients (Alesina and Glaeser 2006). The correlation documented in Figure 1 might consequently be an artifact of the well-know correlation between LAVs and anti-immigrant sentiment, with anti-immigrant sentiment doing most of the heavy lifting. This would suggest a strong parallel between the situation in Europe and that in the United States

Figure 1: Free Riding Beliefs and Its Ideological Correlates: Great Britain (2016)



Individuals scores are computed using an IRT model. Scores have been standardized. Higher values is associated with conservative beliefs and attitudes (i.e., high prevalence of free riding, authoritarian values, right-wing on economic issues).

Binned scatter plots, The variable on the X-axis is split into 20 bins. The line represents a quadratic fit computed using the full underlying data.

Data: British Social Attitude Survey, 2016 (unweighted)

Table 1: Items Used in Figure 1

<p><i>Free Riding Beliefs</i></p> <p>Benefits for unemployed people: too low and cause hardship vs. too high and discourage job seeking The welfare state encourages people to stop helping each other If welfare benefits weren't so generous, people would learn to stand on their own two feet Many people who get welfare don't really deserve any help Most unemployed people could find a job if they really wanted one Most people on the dole are fiddling Gov't responsibility: good standard of living for the unemployed More spending on unemployment benefits</p>	<p><i>Economic/Redistributive Ideology</i></p> <p>Management will always try to get the better of employees if it gets the chance There is one law for rich and one for poor Working people do not get their fair share of nation's wealth Big business benefits owners at the expense of workers It is the responsibility of the government to reduce the differences in income Government should redistribute income from the better-off to those who are least well-off</p>
<p><i>Liberal-Authoritarian Values</i></p> <p>People who break the law should be given stiffer sentences For some crimes, the death penalty is the most appropriate sentence Young people today do not have enough respect for traditional British values Schools should teach children to obey authority</p>	

Table 2: Items Used in Figures 2 and 3

<p><i>Free Riding Beliefs</i></p> <p>Most unemployed people do not really try to find a job Many manage to obtain benefits/services not entitled to Employees often pretend they are sick to stay at home Social benefits/services make people lazy Social benefits/services make people less willing to look after themselves/family Social benefits/services make people less willing to care for one another</p>	<p><i>Anti-Immigrant Sentiment</i></p> <p>Immigration bad or good for country's economy Country's cultural life undermined or enriched by immigrants Immigrants make country worse or better place to live Allow many/few immigrants of different race/ethnic group from majority Allow many/few immigrants from poorer countries outside Europe Allow many/few immigrants of same race/ethnic group as majority</p>
<p><i>Liberal-Authoritarian Values</i></p> <p>Schools should teach children to obey authority People who break the law should receive much harsher sentences Terrorist suspect in prison until police satisfied</p>	

where, research shows, racial minorities are believed to be over-represented among welfare recipients (Gilens 1999) and where racial and ethnic cues are used to infer propensity to free ride (Fong and Luttmer 2009).

If anti-immigrant sentiment is indeed driving the relationship documented in Figure 1 (right panel), the positive correlation between free riding beliefs and LAVs should disappear once controlling for anti-immigrant sentiment. Unfortunately, the dataset used for Figure 1 does not include items that ask about immigration and immigrants. As an alternative, I turn to the European Social Survey, which includes items similar to those listed in Table 1 and also includes a battery of items asking about anti-immigrant sentiment. The items used are listed in Table 2.

Using the same procedures used to compute Figure 1 (i.e. EFA and CFA, followed by an IRT model), I compute three attitudinal scores, one for free riding beliefs, one for anti-immigrant sentiment and one for LAVs. Figure 2 presents binned scatter plots examining average free riding beliefs according to one's score on the LAV scale, with and without controlling for anti-immigrant sentiment. I reproduce the same analysis using anti-immigrant sentiment, with and without controlling for LAVs. The top panel uses British data and the bottom panel presents the same analysis for Germany. In Great Britain, the two attitudinal variables explain roughly a similar share of the variance

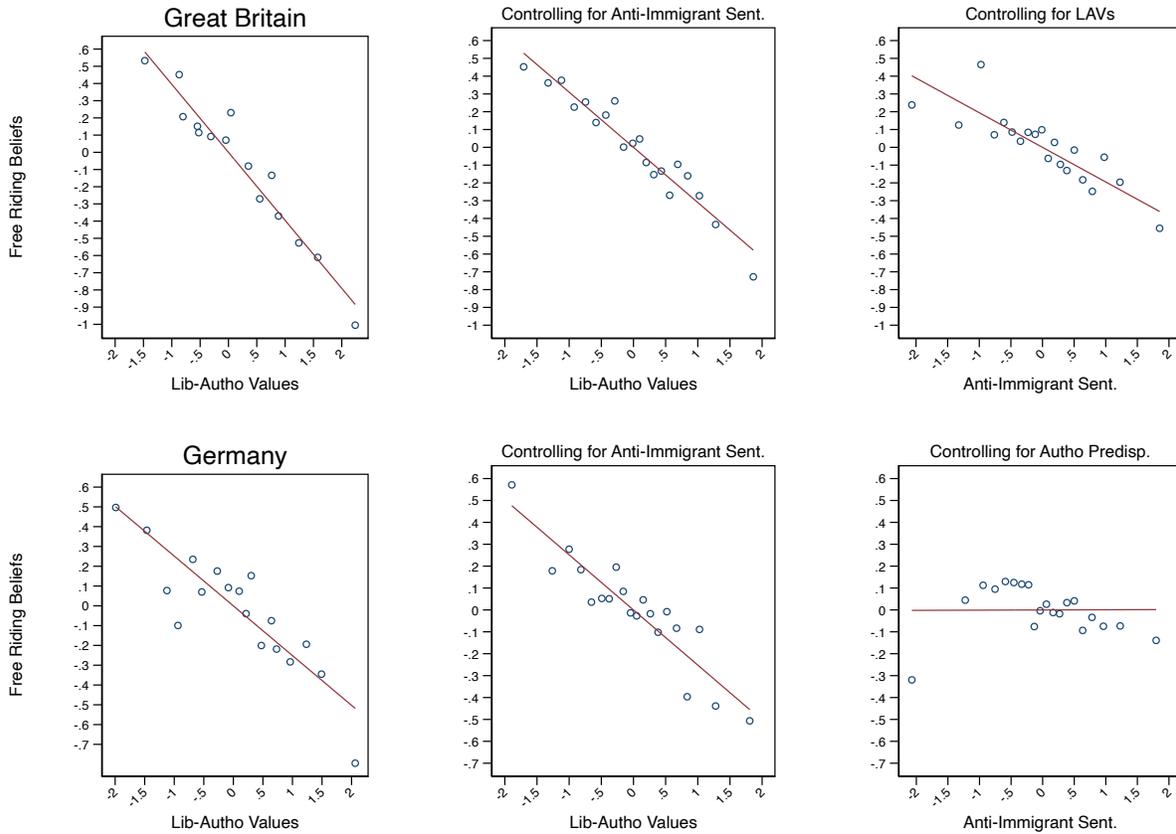
in free riding beliefs. In Germany, free riding beliefs are correlated with LAVs only: the correlation between anti-immigrant sentiment and free riding beliefs disappears once the LAVs index is included as a control. In most Western democracies the pattern is similar to the one in Germany, with some notable exceptions including France and Denmark, which are closer to the British case (not shown). In other words, the correlation between LAVs and free riding beliefs is not an artifact of differences in anti-immigrant sentiment. If anything, in many countries, the converse is true: the correlation between anti-immigrant sentiment and free riding beliefs is itself an artifact of differences in LAVs. Yet, the literature emphasizing the role of immigration-induced diversity has little to say about this robust correlation between free riding beliefs and LAVs.

Status Threat and Economic Change

As discussed earlier, accusations of free riding can be interpreted as attempts to present the poor and the unemployed in an unfavorable light. Conversely, the defense against such accusations, can be interpreted as attempts to present them in a more favorable light. Why are some people motivated to paint the poor and the unemployed in a bad light, while others are motivated to paint them in a good light? Part of the answer, studies suggest, could lie in human beings' drive for status maximization broadly defined.

According to Corneo and Grüner (2002), individuals positively value the material payoff that others receive to the extent that it does not threaten their own relative position. Echoing this argument, Kuziemko et al. (2012) have shown how support for a minimum wage increase was the lowest among individuals earning a little over the minimum wage. They argue that those close to the bottom of the income distribution fear that any increase will automatically lump them with the bottom, something they call "last-place aversion. Ethnographic work also finds that by presenting "other" recipients as "scroungers," low-income individuals build a symbolic boundary that distinguishes them from norm-violating free riders and increases the perceived social distance between themselves and others at the bottom of the desert ladder. According to Lamont and Molnar (2002), this process of social distancing provides an "essential medium through which

Figure 2: Free Riding Beliefs and Anti-Immigrant Sentiment



Individuals scores are computed using a IRT model. Scores have been standardized using country-specific mean and SD. Higher values is associated with conservative beliefs (i.e., high prevalence of free riding, authoritarian values, anti-immigrant sentiment).

Binned scatter plots, The variable on the X-axis is split into 20 bins. The line represents a quadratic fit computed using the full underlying data.

For an idea of the substantive relationship: in Great Britain, one SD increase in the grid score is roughly equal to half a SD increase on the free riding beliefs score (no controls).

people acquire status” and self-esteem.⁵ Jointly, these arguments suggests that being economically and socially proximate to the poor and the unemployed can increase the need to present the latter as undeserving. This prediction runs counter to the more intuitive expectation that disadvantaged individuals will be more sympathetic towards the poor because their similar living conditions lead them to empathize with the struggles and difficulties faced by others in the same situation.

To measure social proximity to the poor, I turn to the BSAS used in Figure 1. I use an occupation-centric social stratification scale developed by Lambert and Prandy (2012). To compute this scale, Lambert and his colleagues rely on census data to build a contingency table where the rows indicate the range of one partner’s occupation and the columns the range of the other partner’s occupation. The cells contain the number of occurrences of each combination in the population.⁶ Lambert and co-authors then use this data to rank occupations from least to most socially proximate based on the likelihood that members of two occupations live under the same roof. Different scales are calculated for men and women, “since holding the same occupation may have different implications for the persons’ social position, depending on their gender” (Bergman and Joye 2001: 36).⁷ On one hand of the scale are professional occupations such as doctors or lawyers (who are attributed to higher scores on the scale) and on the other are low-skill occupations such as janitors or care workers (who are attributed the lower scores). Using detailed occupational information, I match BSAS respondent to their occupational scores.

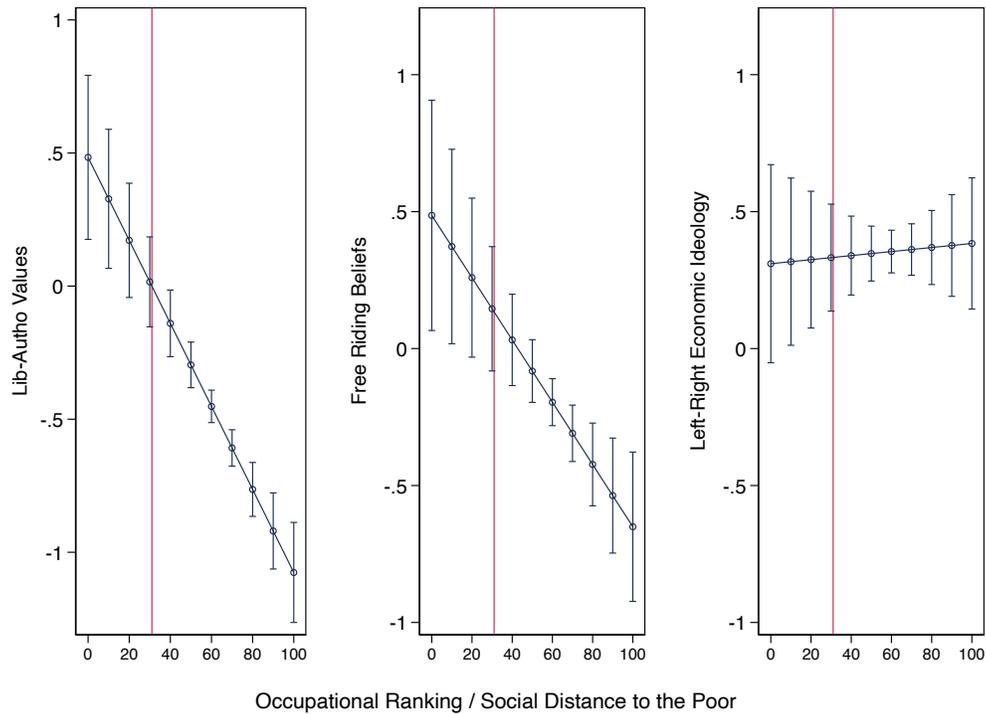
A simple analysis of the correlation between this social ranking measure and free riding beliefs (not shown) reveals that individuals who are ranked lower are more likely to hold anti-welfare free riding beliefs. In other words, identification with the poor does not increase with social proximity to the poor, to the contrary. These results point to status threat as an important line of inquiry for understanding how free riding beliefs form. But before reaching such conclusion, what about LAVs? Indeed, social status is partly a function of education, and education is itself associated with more liberal values. The correlation between social ranking and free riding beliefs could thus be an

⁵ See also Chase and Walker (2013); Gubrium and Lødemel (2014); Garthwaite (2016).

⁶ Partnership is defined here as both marriage and cohabitation.

⁷ For issues relating to “diagonal” effects, e.g. a farmer will most likely be married to a farmer, and multidimensionality, see the website: Lambert, P.S. and Prandy, K. (2011) CAMSIS project webpages: Cambridge Social Interaction and Stratification Scales, retrieved August 2013 from <http://www.camsis.stir.ac.uk/>.

Figure 3: Social Distance and Free riding Beliefs: Tertiary Education Only



See Figure 1 and Table 1 for information on scores

The vertical line marks the 1st percentile of the occupational ranking variable within the sub-sample of university-educated respondents. As one might expect, few highly educated individuals end up in the “lowest” occupations.

artifact of differences in LAVs across groups of difference educational attainment.

To provide a better test of the status threat hypothesis, I consequently zoom in on individuals with a tertiary degree (N=621). Figure 3 plots attitudes as a function of occupational ranking holding education level constant (in this case all respondents have a tertiary degree). For informative purposes, I start the analysis using the economic/redistributive ideology scores (see Table 1). The right panel plots predicted scores as a function of occupational ranking: as shown, there is no correlation between the two measures. The middle panel in Figure 3 plots predicted free riding beliefs as a function of occupational ranking. Even *among* people with a tertiary degree, the extent to which individuals end up in occupations socially proximate to the poor predicts how unsympathetic they are toward welfare recipients. However, as the left panel shows, we still cannot reject the null that this finding is an artifact of the correlation between free riding beliefs and LAVs: people who are socially more proximate to the bottom of the social ladder are also more author-

itarian.⁸ Again, the status threat line of inquiry has little to say about this robust and potentially confounding correlation.

Why are answers to questions about the death penalty, sentencing, and discipline in school most consistently informative of answers to questions about free riding among benefit recipients? Answering this question I show, opens up a new and fruitful line of inquiry.

2 A New Approach to The Formation of Free Riding Beliefs

When discussing attitudes toward policies that redistribute to the poor and the unemployed, researchers have most often emphasized beliefs about recipients' own individual failings (e.g. low effort level, failure to make the right decision when faced with a choice) or beliefs about factors beyond recipients' control (e.g. a global pandemic). Also important are beliefs about the behavioral consequences of social insurance, what economists call moral hazard (e.g. benefits make people lazy). As briefly discussed earlier, both types of items (see Tables 1 and 2 for examples) load on the same latent dimension. Instead of talking of "deservingness perceptions," a term commonly found in the literature, I have bundled all these considerations under the umbrella term of "free riding beliefs", i.e. beliefs about the prevalence of free riding among recipients. Indeed, jointly, beliefs about recipients' moral flaws, their locus of control and moral hazard shape perceptions of the share of welfare recipients who are free riding, i.e. relying on public funds without contributing to the cost of maintaining it through work and social contributions.

Thinking of deservingness perceptions in terms of beliefs about the prevalence of free riding also has the benefit of precision. Deservingness judgments are ubiquitous: does a student deserve a higher or a lower grade? does the perpetrator of an offense deserve tougher or lighter punishment? Is a successful businessman, heir or lottery winner deserving of his millions? The expression "deservingness perceptions" does not discriminate between these different scenarios.⁹ In contrast,

⁸ These results are robust to relaxing the assumption of a linear relationship between the main variables of interest: using both kernel and locally linear regressions, I find a strong relationship between social distance and these two outcomes.

⁹ All these deservingness judgments have in common the following rule of thumb: "if she cannot be deemed responsible for her situation then undeserving of the good or bad outcome associated with it."

the expression “free riding beliefs” better conveys the fact that, in the case that interests us, deservingness judgments apply to a situation of mutual help and interdependence. As I discuss next, this emphasis opens the door to a line of inquiry that is distinct from the ones reviewed above and better speaks to the robust correlation between free riding beliefs and LAVs.

The Social Solidarity Problem and How To Solve It

In the context of redistributive politics in Western societies, agreement and disagreement over who deserves what usually points to one of two ideal-typical resource allocation problems. The first of the two problems, the *inequality-among-equals* problem, was famously discussed by Rousseau in the *Discours on Inequality*. To simplify somewhat, the inequality-among-equals problem is at least partially solved when the majority accept inequality as fair and the rich as deserving. In this paper, I am interested in a different resource allocation problem, i.e. that of the provision by the group of basic material security for its members, lest the group not survive temporary material shocks. I call this ideal-typical problem the *social solidarity* problem. In practice, the social solidarity problem is at least partially solved when most group members are willing *contributors* to some form of social insurance and do not feel resentment towards those who take more than they contribute.¹⁰ In Western societies, this means consenting to paying (usually high) taxes to fund (often generous) social insurance programs without resenting net-beneficiaries of this large-scale resource pooling. The absence of resentment is fostered by the belief that net-beneficiaries are not free riding, i.e. are not willingly taking advantage of the shared resource without contributing to its maintenance.

Behavioral economists and evolutionary theories have sought to unpack the social mechanisms and cognitive modules that underpin such “solution” to the social solidarity problem. Using the tools of game theory and laboratory-based controlled experiments, they have identified a set of behavioral traits, norms and cues that help explain why some situations produce high levels of contributions and low levels of resentments, while others produce the opposite. These behavioral traits, norms and cues, researchers argue, play a key role in helping solve social dilemmas, i.e.

¹⁰ Research has documented the range of emotions underpinning resource sharing and cooperation, most importantly outrage at being taken advantage of (contributing when others are not), shame for failing to “carry one’s weight” (not contributing when others are), or outrage at seeing someone “jump the queue” (receiving the same access despite “slacking” behavior).

situations in which people's interest as members of a collective clashes with their interest as isolated actors. If we think of the successful and continued provision of social solidarity from this lens, then we can reasonably expect that some of these behavioral traits, norms and cues are also going to be relevant for redistributive politics, at least the subset of policies that relate to the most redistributive features of social insurance.¹¹ While it is beyond the scope of this paper to review the expansive literature on social dilemmas and cooperation, a few findings are worth mentioning as they shed a new light on the nature and relevance of free riding beliefs for people's social policy preferences (see also Ostrom and Walker (2003)).

One finding is the importance of the reciprocity norm (Axelrod 1980; Ostrom and Walker 2003). This norm is both simple to describe and surprisingly difficult to theorize. Simply stated, people abiding by this norm behave as conditional cooperators: they willingly contribute to a collective endeavor if they feel others are not free riding (positive reciprocity). They punish free riders by either ceasing to cooperate or by excluding them from accessing the goods generated by cooperation (negative reciprocity). Behavior attached to the reciprocity norm is thus inherently two-faceted and can be presented in one of two lights. The more positive light casts it as a form of conditional altruism: people's default position is to help others unless others are "antisocial" (Henrich et al. 2001; Fong, Bowles and Gintis 2006). Viewed in a negative light, it is a form of conditional punishment: people's default position is to deny help to others unless they are prosocial. This reciprocity norm has shown to help solve the previously mentioned social solidarity problem by turning all group members into willing cooperators, despite the incentive to free ride (Berg, Dickhaut and McCabe 1995; Falk, Fehr and Fischbacher 2003).

To unpack the relationship between the reciprocity norm and this cooperative equilibrium, behavioral economists have relied on an ideal-typical game called the public goods game (PGG) (Fehr and Gächter 2000; Fischbacher, Gächter and Fehr 2001; Ostrom and Walker 2003; Bechtel and Scheve 2014).¹² Results from studies with repeated iterations of the PGG show that, left to their

¹¹ Indeed, policies that redistribute by interfering with the production of market income are unlikely to be affected by concerns of free riding: market institutions' main claim to fame, at least in the collective imaginary, is having overcome conflict between individual and collective interest. In this context, the consent is based on the previously mentioned proportionality principle, which has little to say about free riding.

¹² In this setup, participants independently (and privately) choose how many of their private tokens to put into a public pot. The collectively pooled tokens are multiplied by a fixed factor and then divided among players. Each subject gets to keep the tokens they do not contribute on top of their cut of the common pot. The multiplying factor is chosen to create

own devices, individuals start by behaving prosocially, but ultimately revert to the selfish strategy. Detailed follow-up studies suggest the following dynamic: a significant share of players start with optimistic priors about what others will do (i.e. cooperate) and consequently, also behave in a prosocial way. Yet, after one or several rounds, having observed that not all players behave in such a way, players update their priors in a more pessimistic direction. Over multiple iterations, the final result is that nobody will do what is good for the group as a whole. What the PGG captures is the group's joint failure to solve a social dilemma. Everyone would be better off cooperating, but *failure to develop the shared understanding that all will cooperate* ultimately means that, in line with the tit-for-tat logic of the reciprocity norm, nobody does.

A striking finding coming from this line of work is that under certain conditions, the exact opposite outcome occurs and the group converges to full cooperation. These conditions include information on prior behavior (i.e. players' reputation), the ability to communicate and, most importantly, the ability to punish beyond simply refusing to cooperate. In most designs, the option to punish implies taking resources away (often at a personal price) from those who behave selfishly and free ride. Ultimately, the emergence of cooperation happens through a four step process. First, cooperation is jump-started by a group of "altruistic" individuals willing to take the risk of assuming that others will cooperate. Second, the introduction of punishment limits the share of individuals who decide to free ride in the first place and increases the share of individuals who believe that others will cooperate. Third, with punishment, a subset of individuals (which may or may not include the "altruistic" optimists in round one) pay the price of punishing people who free ride, forcing this latter group to change their behavior. Fourth, after punishment is observed, it leads more to update their priors about the relative share of cooperators and free riders. Core to this dynamic is the existence of individuals willing to punish even if at their own expense, resulting in the shared expectations that free riding will not go unpunished (this is why many fewer people free ride once punishment is introduced).

Of course, the cooperative equilibrium in a PGG is far from capturing the full complexity of all

a tension between collective interest and self-interest. On the one hand, the group's total payoff is maximized when everyone contributes all of their tokens to the public pot. On the other hand, each player could walk away contributing nothing (keeping their tokens) yet still receive their share of the collective pot. Ultimately, each player faces the same dilemma: contributing more tokens only makes sense if others are also contributing, but the more others contribute, the more it makes sense to contribute zero tokens and "free ride" on others' contributions.

the social mechanisms that go into the successful provision of social solidarity on the scale reached by most Western democracies. Yet, it does provide a blueprint for identifying a key set of behavioral parameters and traits—some common across individuals and others not—relevant for how people reason about social solidarity. One common parameter, for example, is a *shared* understanding of the reciprocity norm, specifically the shared understanding that support for generous social spending is the right thing to do conditional on the belief that most people are not free riding.

Shared agreement with the reciprocity norm cannot explain why people form different free riding beliefs. Based on the PGG's stylized set up, we can hypothesize additional behavioral parameters better suited to explaining such individual-level differences. One is whether or not someone starts by expecting the best from others or whether they start by expecting the worst. According to work on generalized trust, people differ significantly on this point (Ostrom and Walker 2003) and significant differences have also been documented not only across individuals but also across cultural groups (Henrich et al. 2001; Barrett et al. 2016).

Another parameter, which has received less attention, is the willingness to pay to see free riders identified and punished. A socially optimal outcome is one in which nobody free rides. As showed by Eleonor Ostrom, in a small enough group, this can be achieved without centralized monitoring. Part of the reason is that a subset of individuals appears especially sensitive to free riding and more willing to sacrifice their own time and resources to identify and punish free riders. In large welfare states, monitoring is performed by a central bureaucracy. The latter can make mistakes, i.e. unfairly denying benefits (unfair punishment of someone deserving) to some or unfairly granting benefits to others (unfair reward of someone undeserving). There is a trade-off between the two types of errors meaning that attempts to minimize one can end up increasing the other (Dickson, Gordon and Huber 2009). I hypothesize that people differ in terms of their error sensitivity (or put differently, their marginal rate of substitution between unfair punishment and unfair rewards). These differences have been famously documented in the case of the criminal justice system (Xiong, Greenleaf and Goldschmidt 2017): while some fret over innocents found guilty, an unfair punishment, others abhor the thought that guilty individuals could be found innocent, and unfair reward. I hypothesize that these differences extend to all contexts approximating a social dilemma, i.e. a situation in which an individual can win big by playing it selfish (stealing, lying or shirking), espe-

cially if others are being prosocial (trusting, helping, contributing). Put differently, the presence of free riders is likely to be more “painful” for some people than for others, making them more willing to “pay” for actions that will minimize the prevalence of free riding by making it harder and more risky to shirk, even if it means some cooperators are unfairly caught in the process.

To summarize, I have argued that individuals differ in how they reason when faced with a social dilemma. Before we discuss how these differences affect the formation of free riding, a brief reminder that our goal is to identify a mechanism of belief formation that could explain why, empirically free riding beliefs correlate with LAVs, as most commonly measured. The next section examines this point in details.

Free Riding and Liberal-Authoritarian Values

Jointly, differences in baseline trust and error sensitivity shape how people approach social dilemmas. On one end of the spectrum, people approach them with low baseline trust (often limited to a small group of socially proximate individuals) and a higher sensitivity to unfair rewards. On the other end, people approach them with high baseline trust (which extends to even socially distant individuals), and a higher sensitivity to unfair punishment. Note how these parameters reinforce each other: given pessimistic priors, one will be more concerned about the failure to punish free riders than the failure to reward cooperators. The same type of reinforcing dynamics applies to people with “optimistic” priors and who are less sensitive to free riding.

What evidence do we have for these two ideal-types? Anyone having taken an introductory course in political theory might recognize the famous contrast between a Hobbesian approach to the social contract and other approaches put forward by liberal [philosophers such as Lockes. Isaiah Berlin talks of “philosophers with an optimistic view of human nature,” among which he includes Locke and Smith, who believe “that there should be a large area for private life over which neither the State nor any other authority must be allowed to trespass.” He contrasts this group to Hobbes “and those who agreed with him” who, concerned with men’s ability to “destroy one another, and mak[e] social life [im]possible,” favor instituting “greater safeguards [...] to keep them in their places,” even if this means “increasing the area of centralized control and decreasing that of the

individual” (Berlin 1958: 7). While Hobbesian thinkers are concerned with protecting the group from individuals’ selfish impulses (i.e. minimizing free riding with an emphasis on punishment, or negative reciprocity), Liberals see a more pressing problem to address: carving a space for individual autonomy to minimize the claims made by the group on their own members (i.e. minimizing false negatives with an emphasis on positive reciprocity).

Social psychologists have shown that these scholarly debates over order-promoting institutions and principles echo differences found in the general public. Findings from Moral Foundation Theory are here particularly helpful. Indeed, a close read of Jonathan Haidt’s work on morality suggests that people systematically differ in terms of the “moral matrices” they rely on to navigate social dilemmas.¹³ Moral Foundation Theory scholars describe two ideal-typical mindsets that echo the Hobbesian versus Liberal distinction. One mindset, they argue, is built on the belief that “suppressing selfishness” requires “strengthening groups and institutions” and “binding individuals into roles and duties in order to constrain their imperfect natures.” This binding approach rests on a pessimistic understanding of human nature and focuses on the “group as the locus of moral value” (Graham, Haidt and Nosek 2009: 1030). The other mindset seeks to favor prosocial behavior by “protecting individuals directly (often using the legal system) and by teaching individuals to respect the rights of other individuals” (Turiel 1983; Shweder et al. 1997). Such a perspective is rooted in a more optimistic conception of human nature, according to which humans are inherently prosocial once protected from harm and injustice.

To sum up, there appears to be a core tension in Western societies between Hobbesian and (for lack of a better term) liberal approaches to social dilemmas. In its ideal-typical form, the Hobbesian approach assumes narrow baseline trust and obsesses over undeserving free riders being unfairly rewarded for their behavior. In its ideal-typical form, the Liberal approach assumes expansive baseline trust and obsesses over deserving pro-social actors being unfairly punished. These differences in the moral matrix people rely on when faced with a social dilemma help explain why people differ in their perceptions of the prevalence of free riding, i.e. their free riding beliefs. If someone

¹³ In this approach, the moral domain is defined as “interlocking sets of values, practices, institutions, and evolved psychological mechanisms that work together to suppress or regulate selfishness and make social life possible” (Graham, Haidt and Nosek 2009: 70). In other words, “moral matrices” help suppress free-riding and encourage cooperative behavior to overcome social dilemmas.

thinks that people do not opportunistically take advantage of a situation, then one is more likely to believe that most people are cooperating and not shirking, and less likely to be concerned about moral hazard. With regards to error sensitivity, I hypothesize that individuals who are particularly sensitive to one type of error discount existing evidence that the opposite errors are also occurring. Indeed, to justify focusing resources on minimizing one type of error, they end up disregarding evidence that the other type of error also needs addressing. In other words, people with a more liberal (Hobbesian) mindset, because of their optimistic (pessimistic) priors will be less (more) inclined to assume that recipients are cheating or shirking. Furthermore, their sensitivity to free riding mean that they will overestimate its prevalence.

We have now all the building blocks necessary to explain the correlation between LAV items on the one hand and free riding beliefs on the other. I have argued that, when answering questions about sentencing, discipline in school, moral hazard or recipients' moral failings, respondents are tapping into a set of considerations and ideas that differ in systematic ways across respondents depending on how they reason about how to best maximize cooperation and minimize free riding. I have conceptualized these differences as differences in baseline trust and error sensitivity. Low baseline trust, and higher sensitivity to free riding tend to go together and predispose people to express support for a harsher penal order and child-rearing practices that limit individual autonomy, as well as over-estimate the prevalence of free riding among welfare recipients. High baseline trust and lower sensitivity to free riding also tend to go together and predispose people to oppose a harsh penal order, favor education practices that enhance individual autonomy and under-estimate the prevalence of free riding.

How does this affect belief formation? People are exposed to informational signals about the state of the world and many of these signals are open to interpretation. When signals are ambiguous, people need to interpret them before incorporating them into their own basket of considerations on a given issue. Hobbesians will be more inclined to interpret ambiguous facts as evidence of free riding. First, this might better align with their own pre-existing considerations (i.e. motivated reasoning). Second, given their low trust, they will tend to interpret ambiguous signals assuming the worse case scenario. Third, given their preference for minimizing free riding, they will need to justify to themselves that it is the fair thing to do (punishment is fair because free riding is preva-

lent). People with Liberal predisposition will have the opposite response to ambiguous signals. This can lead people with different ways of reasoning about social dilemmas to end up with increasingly different posterior beliefs as they interpret the same media or elite signals very differently, which then reinforces their differences in beliefs.

Starting from the robust correlation between LAV items and free riding beliefs items, I have proposed a framework theorizing what distinguishes people who provide authoritarian and anti-welfare answers to these items on the one hand, from people who provide liberal and pro-welfare answers on the other. One important difference, I have argued, is how people reason about solutions to social dilemmas and the monitoring of free riding. People closer to the authoritarian end of the value divide have low baseline trust and favor minimizing free riding even if it means sometimes unfairly punishing pro-social actors. People closer to the liberal end of the value divide have high baseline trust and prefer minimizing the latter type of errors even if it means letting some free riders get away with their behavior. In the right context, these differences in baseline trust and error sensitivity can result in large differences in empirical beliefs about the prevalence of free riding. In the next section, I provide preliminary evidence to support this argument, with a particular emphasis on error sensitivity, one of the more novel implications derived from this line of inquiry.

3 Preliminary Evidence

In this section, I first turn to evidence that people differ in terms of error sensitivity. Second, I endogenously manipulate the perceived cost of free riding and examine whether differences in error sensitivity help predict people's response to this manipulation. To do so, I use survey data collected from a sample of British respondents. Respondents were recruited on an online panel and match the British population on education, age, gender and work status. Note however that this sample does not constitute a representative sample of the British population.

3.1 Differences in Error Sensitivity

Beliefs about the prevalence of free riding, I have argued, are best thought of in terms of beliefs about the prevalence of undeserved rewards: free riders are individuals who are unfairly recognized as deserving, when in fact they are not. Most importantly, I have hypothesized that people differ in terms of error sensitivity, i.e., preference for minimizing the number of undeserved rewards even if at the cost of increasing undeserved punishment.

To examine this expectation, I asked respondents a set of questions covering *five* different institutional domains: social transfers and errors in granting/denying benefits, refugee programs and errors granting/denying refugee status, the justice system and errors punishing/releasing blue collar criminals, the justice system and errors punishing/releasing white collar criminals, and finally, retail banking and errors granting/denying a mortgage. Below, I reproduce item wording using the example of social transfers:

The Department for Work and Pension (the DWP) is in charge of paying social benefits to people who qualify. When giving social benefits, the DWP can make mistakes. But which mistake do you think is WORSE?

- Deny benefits to someone who qualifies.
- Give benefits to someone who does not qualify.
- Both are equally bad
- Do not know

Thinking about mistakes the DWP makes when giving social benefits, which one do you think is MOST COMMON?

- Deny benefits to someone who qualifies.
- Give benefits to someone who does not qualify.
- Both are equally bad
- Do not know

This set of questions was repeated four more times, varying the institutional realm each time

and minimizing differences in wording and structure. Specifically, I asked respondents about the following four institutions and related errors:

- When granting refugee status, the Home Office can make mistakes. But which mistake do you think is worse? most common? how often?
 - Deny refugee status to an asylum seeker who qualifies
 - Give refugee status to an asylum seeker who does not qualify
- (Starts by defining blue-collar crimes) When judging blue-collar crimes, all systems of justice make mistakes. But which mistake do you think is worse? most common?
 - Convict an innocent person
 - Let a guilty person go free
- (Starts by defining white-collar crime) When judging white-collar crimes, all systems of justice make mistakes. But which mistake do you think is worse? most common?
 - Convict an innocent person
 - Let a guilty person go free
- When granting loans, all banks make mistakes. But which mistake do you think is worse? most common?
 - Deny a loan to someone who qualifies
 - Give a loan to someone who does not qualify

If people have preferences over the type of error to minimize then these preferences should be unrelated to the institutional domain under consideration. In other words, if someone is more sensitive to one type of error in one domain, then she is also likely to be sensitive to the same type of error in another domain. If this is true in two domains, then this should be especially likely in a third domain and so on and so forth. Empirically, this implies that patterns of answers to the “which is worse?” items can be modeled assuming that 1) responses to all items are governed by a latent trait θ and that, 2) as a person’s locations on θ increase, the probability of choosing a given type of error as “worse” increases (or at least does not decrease). In practice, this amounts to fitting Mokken’s monotonely homogeneous (MH) model to the “which is worse?” items (more below). Failure to fit such model to the “which is worse?” items would go against the expectation that differences in error sensitivity shape responses to these items.

There remains the possibility that success in fitting such model might just be an artifact of question format, which encourages respondents to mechanically provide similar answers across domains. As a result, I fit the same model to the “most common?” items. If question wording is

driving my results, then I should be able to fit the same model to the latter set of items. However, if my argument is correct, this should not be possible. Indeed, a domain-invariant dislike for one type of error over the other does not imply a domain-invariant propensity to assume that one type of error is more prevalent than the other. That is because beliefs are shaped by domain-invariant individual factors (such as error sensitivity, as previously discussed), domain-specific contextual factors (such as elite discourse, social stereotypes about target groups and domain-specific individual experiences) and the interaction between the two.

A word on the choice of the institutional domains listed above. The justice system items were designed to vary target groups holding the institutional domain constant. If, instead of error sensitivity, emotional affect toward blue and white collar individuals affect responses to the “which is worse?” items, then the correlation across the two items will be low, thus undermining my expectation regarding domain-invariant error sensitivity. The item on banking was designed to capture attitudes toward an issue where individual and collective interests do not clash: to cheat on one’s loan application has little consequences for everyone’s ability to get one. Any evidence that people answer questions for this block of item in similar ways as for the other four blocks would indicate that the patterns of answers might be, again, merely an artifact of question format.

Let’s first examine simple descriptive statistics.¹⁴ As show in Table 3, up to three times more people chose the DNK response option when asked about error prevalence than when asked about which type of error is the worst. Between 40 and 60% of respondents think both errors are equally bad.

To fit a Mokken’s MH model, three assumptions need to hold: unidimensionality, monotonicity and local independence. Unidimensionality means that survey answers can be used to order respondents on a single latent dimension θ . Monotonicity means that the more θ increases, the more likely a respondent provides responses typical of the higher attribute level. Local independence implies that, for individuals with the same θ value, answers to survey questions are uncorrelated. Violations of this third assumption can have serious consequences when using the survey questions

¹⁴ For the analysis in this section, I have dropped individuals who took the survey in less than 7 min. This duration represents the 10th percentile in terms of survey duration. However, I keep individuals who despite being below this cut-off, have successfully answered the screening question.

Table 3: Descriptive Results

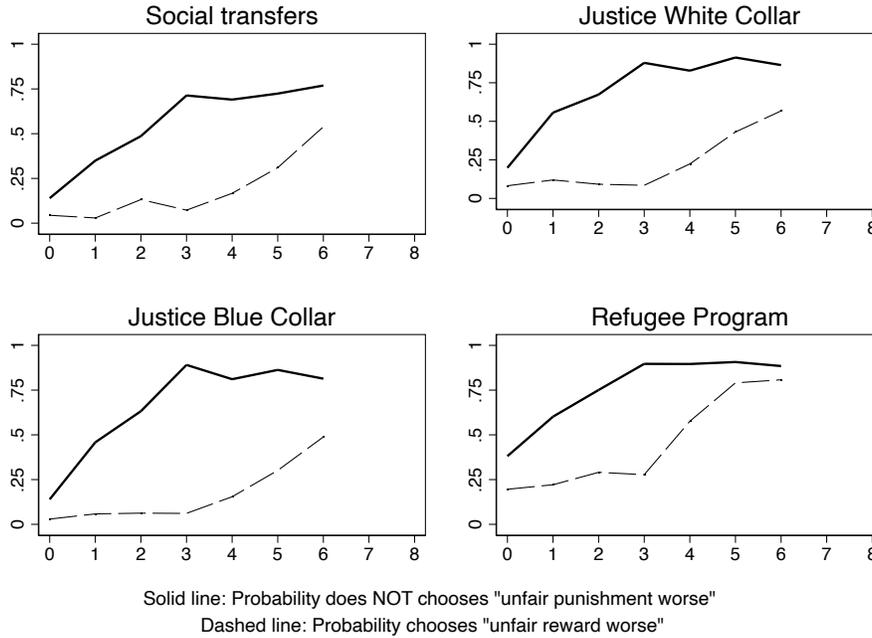
Which one is worse ?				
	Undeserved Punishment	Both	Undeserved Rewards	DNK
Social benefits	0.41	0.44	0.13	0.02
Refugee program	0.23	0.41	0.31	0.05
Blue collar crime	0.29	0.57	0.10	0.04
White collar crime	0.27	0.55	00.14	0.04
Bank loans	0.09	0.52	0.35	0.04
Which one is most common ?				
Social benefits	0.35	0.32	0.23	0.09
Refugee program	0.18	0.37	0.34	0.12
Blue collar crime	0.10	0.33	0.41	0.15
White collar crime	0.09	0.31	0.44	0.16
Bank loans	0.23	0.34	0.31	0.12

to approximate individual values of θ , something I come back to later in this section. As stated earlier, in this part of the analysis, I am most interested in the unidimensionality and monotonicity assumptions.

To test for unidimensionality, I use a procedure examining which items, if any, could be described using a Mokken scale. When applied to the “which is worse?” items, the algorithm returns one Mokken scale ($H=0.35$) using all items with the exception of the banking item: knowing that someone finds it worse to deny a loan to someone who meets the criteria versus grant a loan to someone who does not says nothing about their value on θ . When applied to the “most common?” items, the algorithm returns two Mokken scales ($H=0.39$ and $H = 0.32$) combining the justice system items on the one hand and the refugee and social transfers items on the other. Based on this first analysis, unidimensionality only applies to the “which is worse?” items, this after dropping the item on bank loans.

Next, I turn to the monotonicity assumption. Figure 4 shows cumulative trace lines of each of the remaining “which is worse?” items as a function of the sum of respondents’ answer to the other three items. With the exception of a small kink for the “justice blue collar” item, higher scores imply higher proportion of people who chose a higher attribute. Figure ?? in the appendix shows the same analysis for the “most common?” items and reveals repeated violations of the monotonicity assumption. Specifically, a higher score on three items does not imply a higher likelihood of choosing the higher attribute (in this case undeserved rewards as worse) on the fourth item.

Figure 4: Testing for Monotonicity: “Which is Worse?” items



Probability is computed using proportion of responses for a given score.
Scores (x-axis) are the sum of answers to other three items (ranging from 0 to 9)

Jointly, these results are suggestive of the existence of individual differences in error sensitivity. In light of the difficulty of fitting a similar model to the “not common?” items, these results do not appear to be an artifact of survey design.

3.2 Error Sensitivity, Resource Pooling and Resource Scarcity

I have argued that people rely on very different interpretative lenses when faced with a social dilemma, i.e. a situation in which individual “selfish” behavior can undermine the collective interest. What are possible contextual shocks that could activate these differences? One, explored in this next study, is scarcity, especially if applied to a pooled resource. In such a context, individuals pay increasingly attention to free riding and are more likely to punish free riders by excluding them from accessing the shared resource.

In Great Britain, the NHS is an ideal case to test the role of individual-level differences in cognitive processes: it is a common pool resource that has been presented as overused and overstretched.

In the second part of the survey, I experimentally primed respondents to think about the finances of the NHS, conveying the need to ration services and increase revenue. I then asked respondents who they think should be denied access to the NHS, if anyone. Among the options are refugees, non-citizens, smokers, people who have paid little in taxes and the rich.

For this test, I randomly varied the order of two blocks of survey questions. One set of items—the exclusion block— asks respondents who should be denied access to the NHS. First, respondents are reminded that “priority access to care mostly depends on the seriousness of a patient’s condition.” They are then asked whether they think other factors should be taken into consideration beyond priority care and offers them to select all factors that apply among a suggested list. The response options are as follow: “ YES, it would be right to limit someone’s access to priority care if ...”

- They can afford to pay for similar treatment in the private sector
- Their illness is due to a heavy smoking or drinking habit
- They are not British citizens
- They have contributed little in taxes
- They are illegal immigrants
- Some other reason
- NO, access to priority care should ONLY be decided by the seriousness of someone’s condition

Across treatment and control, the category least likely to be denied priority access are high-income individuals (16%). In total, 60% of respondents chose to exclude one or both categories of immigrants. With regards to smokers, on average, 28% elected to exclude this group. This number is 18% for people who have not paid enough in taxes. In this analysis, I am particularly interested in the immigrant, smoker and low tax-contributor categories. Smokers and people who do not pay enough taxes can be construed as free riders who “take more than they contribute” and are doing so out of personal choice (choosing not to stop smoking, choosing not to work). Historically, the British welfare state was explicitly designed for citizens only: being a citizen gives access to economic security, without economic security one cannot be a full citizen. More generally, group boundaries are constitutive of social solidarity. On the on hand, membership (i.e. group

boundaries) facilitates resource pooling. On the other, resource pooling is what gives membership its value. In light of this, it is unsurprising that research repeatedly finds that being an immigrant counts as a negative when assessing the deservingness of a welfare recipient.

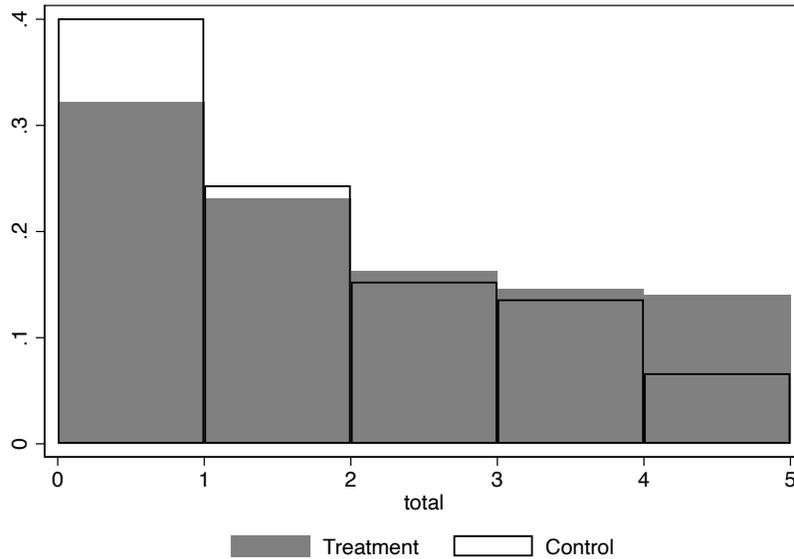
The second block of items —the scarcity block— primes respondents to think about the financial situation of the NHS. It reminds respondents that “(the) NHS is paid for by taxes and is free for all” and that “(b)ecause of growing demands on the NHS, many worry that the NHS will soon not have enough money and resources to provide everyone with the treatment they need when they need it.” Then respondents are asked their opinion on the likelihood that, in the next decade, taxes to fund the NHS will go up, the quality of care will go down or co-pay will be introduced. Individuals who answer the scarcity block before the exclusion block are the treatment group. Others constitute the control group. The order manipulation increases the perceived cost of free riding in the treatment group compared to the control group. I expect fewer people to support universal access in the treatment group and I expect respondents in this group to want to deny priority care to a larger category of “undeserving” individuals.

If people with a more Hobbesian approach to social dilemmas have a stronger dislike for “unfairly rewarding the undeserving”, then I expect larger treatment effects for individuals with more authoritarian responses to LAV items. First, authoritarians are more likely to find immigrants, smokers and low tax contributors less deserving. Second, when resources are scarce, they are also more likely to exclude these individuals from priority access. I also repeat this analysis using an index combining the error sensitivity items discussed in the first part of the analysis. Individuals who find it worse to unfairly grant refugee status, social benefits and freedom to —respectively— the “wrong” migrants, jobless individuals or criminals; will be more likely to deny access to care to more “undeserving” groups. This will apply not only to immigrants but also to smokers and low tax-contributors.¹⁵

Figure 5 compares the total number of groups excluded (none, 1 and up to 5) in the control and treatment groups. Individuals who have to think about the financial state of the NHS are less likely

¹⁵ Reliance on this index assumes local independence, which is a strong assumption to make. In the absence of a purposefully developed measure of error sensitivity, results presented next need to be interpreted with caution, alongside the analysis relying on LAV items.

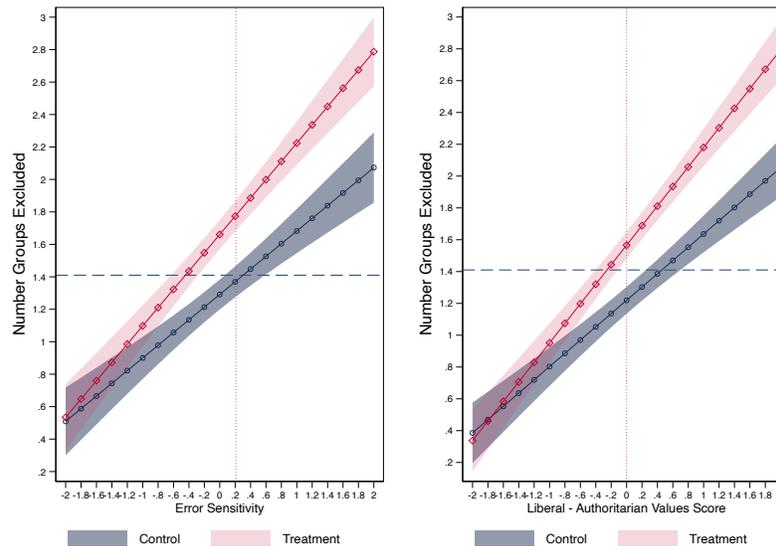
Figure 5: Number of Groups Denied Priority Access



to support universal access and, on average, 0.35 more groups are denied priority access in the treatment condition ($t = 4.69$). I then examine if the effect size varies with error sensitivity and LAVs. As shown in Figure 6, people identified as more authoritarian and more sensitive to people getting undeserved rewards are both more likely, as a baseline, to exclude more groups and also more affected by the treatment.

A closer analysis of the data (see appendix) shows that answers to the exclusion questions are nested meaning that people who chose to exclude smokers or people who did not pay enough in taxes also exclude immigrants, while the reverse is not true. In other words, people who exclude three or more groups are mostly extending their punitive behavior to groups beyond immigrants. In other words, the results are unlikely to be driven by anti-immigrant sentiment alone. In contrast, an emphasis on Hobbesian and Liberal approaches to cooperation can explain these results: immigrants are less deserving because they are outside the boundaries of solidarity, especially so in times of scarcity (i.e. they stand at the bottom of the pecking order in this worldview); the same applies to smokers who are responsible for wasting limited resources and thus behaving in a non-cooperative fashion.

Figure 6: Number of Groups Denied Priority Access: Heterogeneity



Variables on the x-axes were measured earlier in the survey.

The horizontal line captures the average number of groups excluded. The vertical line plots the median error sensitivity and grid preferences score.

Larger values on the x-axis indicates higher sensitivity to false positives and more authoritarian LAVs.

Discussion

In this paper, I highlighted the robust —yet overlooked— correlation between free riding beliefs and what is often called "liberal-authoritarian" values. Existing lines of research on perceptions of deservingness have little to say about this correlation. Moreover, I have shown that evidence in support of existing theories is partly an artifact of the correlation between LAVs on the one hand and non-economic/cultural ideology, anti-immigrant sentiment and social ranking on the other.

What is the data generation process underpinning the correlation between LAVs and free riding beliefs? My preliminary answer to this question emphasizes differences in how people reason about how to maximize pro-social behavior and minimize free riding. This line of inquiry points to a relationship between scarcity and social solidarity, mediated by error sensitivity. I showed that, in times of scarcity, boundaries of solidarity narrow at the expense not only of outsiders but also people whose actions can be construed as free riding. This narrowing is markedly stronger among people most sensitive to free riding, who also happen to provide more authoritarian response patterns to LAV items.

This evidence, of course, is only suggestive. Providing a more extensive test will require developing a better measure of error sensitivity and providing evidence that individual-level differences in error sensitivity are time-invariant. Furthermore, while I have leveraged the correlation between LAV items and free riding beliefs to propose a new line of inquiry, I have not directly tested the hypothesized differences in information processing explaining how this correlation emerges. One important assumption to test is the one stating that people have an intuitive sense of moral dilemmas: everyone, irrespective of their background or cognitive skills, should be able to intuitively perceive the tension between individual and collective interest.

The argument presented in this paper raises one important question: where do Hobbesian-binding versus Liberal-individualizing orientations to social dilemmas come from? A natural tendency is to tie them to fundamental personal traits such as personality types (e.g. the “authoritarian personality”). In other words, people are born closer to one side of the spectrum than the other.¹⁶ Based on my reading of the existing evidence, I more cautiously tie these differences to shared cultural traits, themselves shaped by group-specific ecological factors.¹⁷ The evidence that social distance measures that leverage occupational endogamy are highly correlated with both LAVs and free riding beliefs suggest the importance of occupational socialization and sorting.

The framework’s potential payoffs extend beyond understanding the formation of free riding beliefs. Assuming the argument presented in this paper is one step in the right direction, it could also help improve the way we conceptualize and theorize non-economic/cultural attitudes, i.e., the “second-dimension” of electoral politics in post-industrial democracies. While there is some consensus regarding the nature of the first economic/redistributive dimension, this consensus does not extend to the second dimension. Some researchers boil the latter down to whatever non-economic issue is most salient in their country and period of interest, e.g. immigration (Ford and Jennings 2020), government intervention in favor of minority groups (Miller and Schofield 2003), nationalism (Shayo 2009), abortion or gay rights (Weeden and Kurzban 2014). Others empha-

¹⁶ Haidt argues that their coexistence within the human population has provided humans with evolutionary advantages.

¹⁷ For more on this topic, see the research on class and parenting styles (Gelfand 2019), studies of the relationship between demographic transitions, household formation patterns and cultural values (Lesthaeghe and Neidert 2007), or evidence that economic insecurity affects preferred “leadership style” (Safra et al. 2017; Ballard-Rosa et al. 2017).

size the shared “cultural preferences” or “values” that underpin people’s attitudes toward these issues, distinguishing between “cosmopolitan liberals” and “parochial authoritarians” (Kitschelt and Rehm 2014).¹⁸ Unlike the “first dimension,” researcher still disagree on what types of issue and proto-ideological constraints hold these issues together. Furthermore, the socio-structural bases and behavioral motives underpinning the second dimension have proven hard to theorize and document.¹⁹

The argument presented in this paper contributes to developing a more fully fledged theory of second dimension attitudes. First, it can help explain why LAVs items and anti-immigrant sentiment items then to go together (Kitschelt and Rehm 2014). For Hobbesians, large groups and porous borders are more likely to be perceived as moral hazard making them more inclined to perceive immigrants as welfare shoppers. More generally, group boundaries are constitutive of social solidarity. On the one hand, membership (i.e. group boundaries) facilitates resource pooling. On the other, resource pooling is what gives membership its value: people derive meaning and status from their membership in a protective resource pool. In light of this, it is less surprising that low status individuals are more invested than high status in maintaining the welfare state as an exclusive club good. The argument presented in this paper also suggests that non-economic issues such as abortion, gay rights or women’s right, which do not have the features of a social dilemma, might be less constitutive of the second dimension than issues such as law and order, immigrants’ access to the welfare state, or support for welfare reform that seek to minimize moral hazard. As shown by Busemeyer, Rathgeb and Sahm (2022) attitudes on these particular issues distinguish far-right voters from others. In other words, this paper only scratches the surface of a larger research agenda: I hope to have convinced my reader that this line of inquiry is worth investing time and resources.

¹⁸ Other expressions include “communitarian” versus “liberal” (Bornschiefer 2010), “materialist” versus “post-materialist” (Inglehart 2007) or “open” versus “closed” (Johnston, Lavine and Federico 2017).

¹⁹ Contrast this with the first dimension: in this case, assuming that it this dimension is about inequality and redistribution and that people’s beliefs and preferences on this dimension are shaped by material interest goes a long way.

References

- Alesina, A. and E.L. Glaeser. 2006. *Fighting Poverty in the US and Europe: a World of Difference*. Oxford University Press, USA.
- Alesina, Alberto and Stefanie Stantcheva. 2020. Diversity, immigration, and redistribution. In *AEA Papers and Proceedings*. Vol. 110 pp. 329–34.
- Amodio, David M, John T Jost, Sarah L Master and Cindy M Yee. 2007. “Neurocognitive correlates of liberalism and conservatism.” *Nature neuroscience* 10(10):1246–1247.
- Axelrod, Robert. 1980. “More effective choice in the prisoner’s dilemma.” *Journal of Conflict Resolution* 24(3):379–403.
- Ballard-Rosa, Cameron, Mashail Malik, Stephanie Rickard and Kenneth Scheve. 2017. “The economic origins of authoritarian values: evidence from local trade shocks in the United Kingdom.” *Unpublished manuscript, Department of Political Science, University of North Carolina at Chapel Hill*. <https://goo.gl/T6XG3T>.
- Barrett, H Clark, Alexander Bolyanatz, Alyssa N Crittenden, Daniel MT Fessler, Simon Fitzpatrick, Michael Gurven, Joseph Henrich, Martin Kanovsky, Geoff Kushnick, Anne Pisor et al. 2016. “Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment.” *Proceedings of the National Academy of Sciences* 113(17):4688–4693.
- Bechtel, Michael M and Kenneth Scheve. 2014. “Public Goods, Reciprocity, and the Causal Effect of Expected Cooperation in Representative Samples.”.
- Berg, Joyce, John Dickhaut and Kevin McCabe. 1995. “Trust, reciprocity, and social history.” *Games and economic behavior* 10(1):122–142.
- Bergman, Manfred Max and Dominique Joye. 2001. “Comparing social stratification schemas: CAMSIS, CSP-CH, Goldthorpe, ISCO-88, Treiman, and Wright.”.
- Berlin, Isaiah. 1958. Two concepts of liberty. pp. 33–57.
- Bornschieer, Simon. 2010. “The new cultural divide and the two-dimensional political space in Western Europe.” *West European Politics* 33(3):419–444.
- Busemeyer, Marius R, Philip Rathgeb and Alexander HJ Sahn. 2022. “Authoritarian values and the welfare state: the social policy preferences of radical right voters.” *West European Politics* 45(1):77–101.
- Cavaille, Charlotte. 2021. *Fair Enough: Demand for Redistribution in the Age of Inequality*. Unpublished Manuscript.
- Chase, Elaine and Robert Walker. 2013. “The co-construction of shame in the context of poverty: Beyond a threat to the social bond.” *Sociology* 47(4):739–754.

- Corneo, Giacomo and Hans Peter Grüner. 2002. "Individual preferences for political redistribution." *Journal of Public Economics* 83(1):83–107.
- Dickson, Eric S, Sanford C Gordon and Gregory A Huber. 2009. "Enforcement and compliance in an uncertain world: An experimental investigation." *The Journal of Politics* 71(4):1357–1378.
- Falk, Armin, Ernst Fehr and Urs Fischbacher. 2003. "On the nature of fair behavior." *Economic inquiry* 41(1):20–26.
- Fehr, Ernst and Simon Gächter. 2000. "Cooperation and punishment in public goods experiments." *American Economic Review* pp. 980–994.
- Fischbacher, Urs, Simon Gächter and Ernst Fehr. 2001. "Are people conditionally cooperative? Evidence from a public goods experiment." *Economics Letters* 71(3):397–404.
- Fong, Christina M and Erzo FP Luttmer. 2009. "What determines giving to Hurricane Katrina victims? Experimental evidence on racial group loyalty." *American Economic Journal: Applied Economics* 1(2):64.
- Fong, Christina M, Samuel Bowles and Herbert Gintis. 2006. "Strong reciprocity and the welfare state." *Handbook of the Economics of Giving, Altruism and Reciprocity* 2:1439–1464.
- Ford, Robert and Will Jennings. 2020. "The changing cleavage politics of Western Europe." *Annual Review of Political Science* 23:295–314.
- Garthwaite, Kayleigh. 2016. "Stigma, shame and 'people like us': an ethnographic study of foodbank use in the UK." *Journal of poverty and social justice* 24(3):277–289.
- Gelfand, Michele. 2019. *Rule makers, rule breakers: Tight and loose cultures and the secret signals that direct our lives*. Scribner.
- Gilens, Martin. 1999. *Why Americans hate welfare: Race, media, and the politics of antipoverty policy*. University of Chicago Press.
- Graham, Jesse, Jonathan Haidt and Brian A Nosek. 2009. "Liberals and conservatives rely on different sets of moral foundations." *Journal of personality and social psychology* 96(5):1029.
- Gubrium, Erika K and Ivar Lødemel. 2014. "Not Good Enough': Social Assistance and Shaming in Norway'." *The Shame of It. Global Perspectives on Anti-Poverty Policies* pp. 85–110.
- Häusermann, Silja and Hanspeter Kriesi. 2015. "What do voters want? Dimensions and configurations in individual-level preferences and party choice." *The politics of advanced capitalism* pp. 202–230.
- Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis and Richard McElreath. 2001. "In search of homo economicus: behavioral experiments in 15 small-scale societies." *American Economic Review* pp. 73–78.

- Hetherington, Marc J and Jonathan Weiler. 2005. Authoritarian Disposition and Political Choice. In *Annual Meeting of the Midwest Political Science Association, Chicago, April*. pp. 7–10.
- Inglehart, Ronald. 2007. “Postmaterialist values and the shift from survival to self-expression values.” *The Oxford Handbook of Political Behavior* .
- Jensen, carsten and Michael Bang Petersen. 2014. “Cognitive Bias and the Politics of Health Care.”.
- Johnston, Christopher D, Howard G Lavine and Christopher M Federico. 2017. *Open versus closed: Personality, identity, and the politics of redistribution*. Cambridge University Press.
- Jost, John T, Jack Glaser, Arie W Kruglanski and Frank J Sulloway. 2003. “Political conservatism as motivated social cognition.” *Psychological bulletin* 129(3):339.
- Kitschelt, Herbert. 1997. *The radical right in Western Europe: A comparative analysis*. University of Michigan Press.
- Kitschelt, Herbert and Philipp Rehm. 2014. “Occupations as a Site of Political Preference Formation.” *Comparative Political Studies* p. 0010414013516066.
- Kuziemko, Ilyana, Ryan W Buell, Taly Reich and Michael I Norton. 2012. “Last-place Aversion”: Evidence and Redistributive Implications. Technical report National Bureau of Economic Research.
- Lambert, P.S P.S and K. Prandy. 2012. “CAMSIS project webpages: Cambridge Social Interaction and Stratification Scales.” <http://www.camsis.stir.ac.uk/>.
- Lamont, Michèle and Virag Molnar. 2002. “The study of boundaries in the social sciences.” *Annual review of sociology* 28:167–195.
- Lesthaeghe, Ron and Lisa Neidert. 2007. The Political Significance of the “Second Demographic Transition” in the US: A Spatial Analysis. In *meetings of the Population Association of America, New York, March*. pp. 28–30.
- Miller, Gary and Norman Schofield. 2003. “Activists and partisan realignment in the United States.” *American Political Science Review* 97(2):245–260.
- Morgan, G Scott, Elizabeth Mullen and Linda J Skitka. 2010. “When values and attributions collide: Liberals’ and conservatives’ values motivate attributions for alleged misdeeds.” *Personality and Social Psychology Bulletin* 36(9):1241–1254.
- Ostrom, Elinor and James Walker. 2003. *Trust and reciprocity: Interdisciplinary lessons for experimental research*. Russell Sage Foundation.
- Petersen, Michael Bang, Daniel Sznycer, Leda Cosmides and John Tooby. 2012. “Who deserves help? evolutionary psychology, social emotions, and public opinion about welfare.” *Political psychology* 33(3):395–

418.

- Rodden, Jonathan. 2018. "Keeping Your Enemies Close: Electoral Rules and Partisan Polarization."
- Safra, Lou, Yann Algan, Teodora Tecu, Julie Grèzes, Nicolas Baumard and Coralie Chevallier. 2017. "Childhood harshness predicts long-lasting leader preferences." *Evolution and Human Behavior* 38(5):645–651.
- Sennett, Richard and Jonathan Cobb. 1993. *The hidden injuries of class*. WW Norton & Company.
- Shayo, Moses. 2009. "A model of social identity with an application to political economy: Nation, class, and redistribution." *American Political Science Review* 103(2):147–174.
- Shweder, Richard A, Nancy C Much, Manamohan Mahapatra and Lawrence Park. 1997. "The big three of morality (autonomy, community, divinity) and the big three explanations of suffering." *Morality and health* pp. 119–169.
- Skitka, Linda J and Anthony N Washburn. 2016. "Are conservatives from Mars and liberals from Venus? Maybe not so much." *Social psychology of political polarization* pp. 78–101.
- Sniderman, Paul M, Phillip E Tetlock and Richard A Brody. 1993. *Reasoning and choice: Explorations in political psychology*. Cambridge University Press.
- Turiel, Elliot. 1983. *The development of social knowledge: Morality and convention*. Cambridge University Press.
- Uunk, Wilfred and Wim van Oorschot. 2019. "Going with the Flow? The Effect of Economic Fluctuation on People's Solidarity with Unemployed People." *Social Indicators Research* 143(3):1129–1146.
- Weeden, Jason and Robert Kurzban. 2014. *The Hidden Agenda of the Political Mind: How Self-Interest Shapes Our Opinions and Why We Won't Admit It*. Princeton University Press.
- Xiong, Moulin, Richard G Greenleaf and Jona Goldschmidt. 2017. "Citizen attitudes toward errors in criminal justice: Implications of the declining acceptance of Blackstone's ratio." *International Journal of Law, Crime and Justice* 48:14–26.