CESifo WORKING PAPERS

9381
2021

October 2021

# Global Evidence on Misperceptions and Preferences for Redistribution

*Jennifer Feichtmayer, Klaus Gründler*

CESifo

**Impressum:**

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9381*

# Global Evidence on Misperceptions and Preferences for Redistribution


## Abstract

Individuals often hold erroneous beliefs about their socio-economic status relative to others. We develop a new machine learning technique to measure these misperceptions and use large-scale international survey data to compute status misperception for 241,757 households from 97 countries (24 OECD, 73 non-OECD). We show that status misperception is a widespread phenomenon across the globe. Upward-biased perceptions are associated with lower preferences for redistribution and have direct consequences for welfare provision via the tax and transfer system. The effect accounts for approximately 9% of the variation in redistribution preferences, is independent of socio-demographic characteristics, robust to measurement errors in social surveys, and occurs similary when we change the underlying micro data or examine party preferences.


JEL-Codes: D310, H530, I300, C430.

Keywords: misperceptions, machine learning, socio-economic status, preferences, redistribution, welfare provision, taxes and transfers.

*Jennifer Feichtmayer*
*FAU Erlangen-Nürnberg*
*Nürnberg / Germany*
*jennifer.feichtmayer@fau.de*

*Klaus Gründler**
*ifo Institute – Leibniz Institute for Economic Research at the University of Munich / Germany*
*Gruendler@ifo.de*

*corresponding author

October 20, 2021
We would like to thank David Autor, Kai Gehring, Arye Hillman, Ilyana Kuziemko, Tommy Krieger, Erzo F.P. Luttmer, Niklas Potrafke, Panu Poutvaara, Regina Riphahn, and Heinrich Ursprung for their valuable comments and discussions. We would also like to thank participants of the Spring Meeting of Young Economists (SMYE) 2021 in Bologna (virtual), the EALE SOLE AASLE World Conference 2020 in Berlin (virtual), the 29th Silvaplana Workshop on Political Economy 2020 in Pontresina, and the 8th Conference for Social and Economic Data (KSWD) in Berlin for excellent comments and feedback. We are also grateful for discussions with participants in research seminars at various universities and research institutes.

# 1 Introduction

*"Magic mirror on the wall, who is the fairest one of all?"*
—Brothers Grimm

How much do individuals know about their own socio-economic background relative to others? And do erroneous perceptions about the relative social standing influence individuals' preferences? Standard models from many sub-fields of economics are built on the assumption that individuals are correctly informed about the world they live in. This assumption, however, has often proved to be wrong (e.g. Kuziemko et al., 2015; Alesina et al., 2018a; Stantcheva, 2020c). A prime example is peoples' support to transfer resources from the rich to the poor. From a rational viewpoint, a relatively poorer individual gains more from equalizing tax and transfer policies and should therefore demand more of it (Romer, 1975; Meltzer and Richard, 1981). This prediction, however, is far from what observational data suggests.[1]

Aiming to reconcile predicted with observed preference formation, experimental studies have brought forward the hypothesis that individuals' preferences are grounded on erroneous perceptions about their relative income position (Cruces et al., 2013; Engelhardt and Wagener, 2017; Karadja et al., 2017). Individuals who think they are richer demand less redistribution than would be rational and vice versa. Although experimental studies delivered many insights about individual-level heterogeneity for single countries, the evidence differs greatly across countries in the extent to which perceptions are biased, and in the extent to which these biases affect peoples' preferences.

In this paper, we take two steps to advance on the existing evidence. First, we consider individuals' relative social standing, which is often more decisive for preference formation than income alone (Corneo and Grüner, 2000, 2002; Clark et al., 2008; Keely and Tan, 2008; Brown-Iannuzzi et al., 2014; Hvidberg et al., 2020). Second, we take into account heterogeneity in country-specific characteristics—e.g. in the form of institutions, political history, and cultural norms—by analyzing perception biases and preferences on a global scale. We use large-scale survey data for more than 250,000 individuals from 97 countries (24 OECD, 73 non-OECD) and employ machine learning algorithms to derive individual-level measures of misperceptions that are harmonized across countries. We find that erroneous beliefs about the relative social position are widespread among citizens, but also uncover heterogeneity in the extent of perception bias across countries and across individuals. Our results show that biased perceptions have large effects on preference

[1] The standard political economy model predicts that rising income disparities in the OECD should have led to greater support for equalizing policies. Empirical data, however, point to remarkable cross-country differences in how redistribution preferences have developed over the past 30 years. Data from the World Value Survey, for instance, shows that redistribution preferences declined in Italy and Spain between the early 1980s and the post-2010 period, while preferences increased in Germany and Sweden, and remained constant in Japan and the United States.

formation. Individuals who overestimate their status are significantly less supportive of transferring resources from the rich to the poor and vice versa. The evidence from a global perspective reveals that countries differ in the size but not in the direction of this effect, pointing to a general pattern in human behavior.

We move beyond previous work by focusing on social status rather than on income alone. The idea that individuals are motivated by their pursuit of social status goes back to the earliest writings known to mankind. Many classical economic theories have been centered around what Thomas Hobbes described as men's *"continual competition for honor and dignity"* (Hobbes, 1651).[2] More recently, the social status has been increasingly taken into account by economists in both theoretical and empirical studies (see, e.g., Luttmer, 2005; Solnick and Hemenway, 2005; Falk et al., 2020; Hvidberg et al., 2020). Our stylized theoretical framework shows that the use of social status helps mitigate conceptual and statistical problems that arise when using annual income data to study the effects of relative well-being on preferences. First, the social status is more reflective of family background and lifetime income, leading to a more accurate measure of individuals' disposable resources and long-run expectetions thereof. Accounting for these factors is important when inter-household family redistribution results in individuals being wealthier than reflected by their own income (e.g. Cox, 1987; Cox et al., 2004) or when expectations about future income influences the formation of preferences (Bénabou and Ok, 2001; Alesina and Giuliano, 2011; Rueda and Stegmueller, 2019). Second, measuring misperceptions based on the social status alleviates methodological concerns, including life-cycle biases (Haider and Solon, 2006) and under- or over-reporting in social surveys (Mayer et al., 2015; Mittag, 2019).

The main challenge when examining the social status of individuals is to aggregate its multiple characteristics into a uni-dimensional index. Previous techniques have largely been unsuccessful in providing reliable individual-level classifications of social status and suffer from conceptual and methodological shortcomings (Oakes and Rossi, 2003; Marks, 2011; Moreno-Maldonado et al., 2018). The reason is that theory provides no guidance on how the function that maps socio-economic characteristics into an index of social status might look like. We develop a new machine learning algorithm that "learns" this function based on observed socio-economic characteristics and that transfers the task of aggregation into a non-linear optimization problem. Our approach is built on Support Vector aggregation, which has been shown to outperform traditional aggregation techniques (Gründler and Krieger, 2021a). Unlike previous methods, our algorithm avoids ad-hoc assumptions and provides a distribution of classification outcomes for each individual to account for measurement uncertainty in survey data (Bound et al., 2001).

[2] Classical economic theories that include status competition as a fundamental element are, for instance, the work of Karl Marx, Thorstein Veblen, and James Duesenberry. A similar perspective can be found in the work of John Stuart Mill, to whom the quote *"men do not desire to be rich, but richer than other men"* is attributed.

We then compare individuals' relative standing in society with self-reported perceptions about their status to measure the degree of misperception.

We uncover widespread misperceptions of status across the globe. More than 40% of all individuals in the OECD and about 37% in the full sample perceive to be in the middle of the status distribution. This perception bias towards the middle class is the result of substantial misjudgments of the own status at the top and the bottom of the distribution. We find that many individuals at the bottom of the distribution overestimate their social status, while individuals at the top tend to underestimate their status. We also document large variation in the extent and direction of perception bias across individuals in each country, and find that older and less educated individuals have a higher propensity to overestimate their social status. In contrast, individuals that are well-informed about recent political events and supporters of left-wing political parties are less likely to overestimate their status.

In the second part of the paper, we explore whether biased perceptions of social status influence peoples' preference formation. We focus on preferences for redistribution, which are at the foundation of political economy. In the presence of incomplete information, individuals need to make inferences about their relative position in society. When people overestimate this position, they underestimate the gains in status that can be realized via redistribution and vice versa. In a similar vein, individuals often signal their status via consumption (Bilancini and Boncinelli, 2012), and hence they feel worse off when others around them consume more (Luttmer, 2005). While greater consumption of peers is positively associated with redistribution preferences, this effect should be lower when individuals overestimate their status and vice versa. Both theories suggest that upward-biased perceptions should be negatively related to redistribution preferences.

Our results show that individuals who overestimate their status are significantly less supportive of redistribution. The relationship is particularly strong in OECD countries and slightly weaker if measured based on all available 97 countries. The "misperception effect" is robust to changes in the employed estimation strategy and is similar in size regardless of whether we use our benchmark dataset obtained by the World Value Survey (WVS) or the Life in Transition Survey (LiTS), which we use for cross-validation. There is also little heterogeneity in the effect across individual-level characteristics. Upwards-biased perceptions decrease support for redistribution irrespective of income levels, gender, age, education, party preference, and religion. We reveal large effects of misperceptions also when we examine individuals of single countries, showing that the misperception effect is a global phenomenon. To address measurement errors, we run multiple imputation analyses drawing on the distribution of our misperception measure and find that our results are not driven by measurement uncertainty in social surveys. We also account for potential reporting biases by investigating revealed preferences using information on the political parties individuals vote for in national elections. Finally, the results show that by influencing preferences, biased perceptions ultimately affect tax and transfer systems. Status misperceptions therefore have important consequences for economic policy.

**Contribution to the literature:** We most strongly connect to the literature that examines perceptions and misperceptions of economic variables based on survey data. Using a representative telephone survey for the U.S., Blinder and Krueger (2004) show that perceptions about key economic policy questions are driven mostly by ideology and less so by actual knowledge about economics. In several papers, Stantcheva (2020a,b) documents that people have perceptions about economic policies that starkly differ from those economists usually have about certain topics. More broadly, it has been shown that there is large divergence in opinions on the economy between economic experts and the general public (Sapienza and Zingales, 2013). Misperceptions are also common regarding the functioning of income redistribution, particularly with respect to individuals' understanding of tax policies (Stantcheva, 2020c). Households have been shown to systematically underestimate their marginal tax rate (De Bartolome, 1995; Ballard and Gupta, 2018) because individuals often approximate complex and non-linear tax schedules with less complicated linear ones (Rees-Jones and Taubinsky, 2020).

Some papers specifically investigate erroneous perceptions of individuals about the own ranking in the income distribution. These studies have been conducted for Buenos Aires (Cruces et al., 2013), Sweden (Karadja et al., 2017), and Germany (Engelhardt and Wagener, 2017). The results differ in the measured ability of individuals to infer their relative income rank and the extent to which misperceptions influence peoples' preferences. While the Swedish data suggests that 92% of those respondents that missed their income position by more than 10 percentage points underestimate their position, the data for Germany and Argentina shows that about half of individuals overestimate their income position. Informing individuals about their true income position has strong effects on preferences for Argentina, but less so for Germany and Sweden.

We contribute to the literature on misperceptions and preferences in two ways. First, we consider the broader social standing, which is flatter over the life-cycle and less prone to measurement error (Weiss and Fershtman, 1998). Hence, the socio-economic status has been shown to be more decisive for the formation of redistribution preferences than income (Keely and Tan, 2008). Second, we consider the external validity of single-country studies by examining status misperception using a unified framework for 241,757 households in 97 countries (24 OECD, 73 non-OECD), which are representative for roughly 90% of the world population. The global perspective allows us to account for the many sources of unobserved heterogeneity across countries that may influence preferences, perceptions, and the relationship between the two (e.g. cultural values, institutional frameworks, geographic and historical factors). As our dataset spans the period 1981–2016, we can also account for period-specific shocks that may have influenced individuals' preferences at a given point in time (e.g. economic crises, natural disasters, epidemics, periods of political instability) and for cohort-fixed effects.[3] Using large-scale international survey data also substantially increases statistical power, alleviating increasing concerns about underpowered results in economic research (Ioannidis et al., 2017).[4] Our survey-based measures show patterns that are closely comparable to misperceptions identified in experiment-based studies for Germany and Sweden.

Our paper also speaks to the literature showing that perceptions are better predictors for preferences than officially reported statistics. Recent work has shown that individual's perceptions about intergenerational mobility differ from actual chances to climb the income ladder (Alesina et al., 2018b) and that individuals greatly overestimate the share of migrants in their home country (Alesina et al., 2018a). Both sources of perception bias influence redistribution preferences. Perceptions about the own standing in society have also been shown to influence individuals' fairness views (Hvidberg et al., 2020). Other papers have documented that misperceptions can have real economic consequences. Gründler and Köllner (2017) show that government redistribution is related much more to perceived inequality measures than to official inequality numbers. Scholars have also examined how misperceptions about the tax and transfer system affects the impacts of the Earned Income Tax Credit (Chetty et al., 2013) and the Child Tax Credit (Feldman et al., 2016) on labor market decisions and earnings. Our paper echos the real-world consequences found in earlier studies, providing evidence that cross-country variations in the ability to infer the own social position is related to the generosity of welfare systems.

More generally, our paper is connected to the large literature that examines how individuals form economic preferences and, in particular, preferences for redistribution (see, e.g., Alesina et al., 2018b; Kuziemko et al., 2015; Lockwood and Weinzierl, 2015; Durante et al., 2018; Alesina and Giuliano, 2011). We complement these studies by showing that redistribution preferences can partly be traced back to biased perceptions about objective reality. We also show that this is a universal phenomenon that can be found in very different societies across the globe.

The standard political economy model of redistribution suggests that greater inequality leads to more government redistribution (Romer, 1975; Meltzer and Richard, 1981), but this link is often weakly reflected in the data. Several papers have studied the causes of this "redistribution puzzle", stressing the role of subjective upwards mobility (Bénabou and Ok, 2001), cultural factors (Alesina and Giuliano, 2011; Luttmer and Singhal, 2011), last-place aversion (Kuziemko et al., 2014), and differences in political power (Ursprung

[3]The long-run effects of macroeconomic environments on redistribution preferences are well-documented. For instance, Guiliano and Spilimbergo (2014) show that individuals who experienced a recession when young have greater preferences for redistribution policies.

[4]The experimental studies for Buenos Aires and Germany include 1,100 respondents, the study for Sweden is based on 1,001 respondents. Our analysis exceeds this number by a factor of 240. Also, we include 2,643 respondents from Germany, 3,031 respondents from Sweden, and 3,124 respondents from Argentina.

and Breyer, 1998). We contribute to this literature by showing that upwards-biased status perceptions are systematically related to lower redistribution preferences. The implication is that the standard political economy model may provide more reliable predictions when formulated on income perceptions rather than on realized incomes.

Finally, we contribute to the measurement of social status by providing a new technique that draws on machine learning algorithms. Our technique addresses concerns levied against conventional techniques that aim to aggregate socio-economic characteristics into a uni-dimensional index. By shifting the problem of classification into a non-linear optimization problem, our approach discards ex ante assumptions on the aggregation function and allows for non-linear relationships between the variables. It also enables the construction of confidence intervals to address measurement uncertainty in survey data.

**Organization:** Section (2) describes how misperceptions influence the formation of preferences and how the use of social status can tackle problems of using annual income data to examine preferences based on economic self-interest. Section (3) discusses previous measures of social status and presents our machine learning technique to compute status and status misperceptions. Section (4) shows perceptions about status and misperceptions across the globe. Section (5) reports the empirical results. Section (6) concludes.

# 2 Status misperceptions and preferences

Standard theories in political economy and public finance describe how agents form preferences based on perfect information. In many cases, however, accessing information is costly or restricted, and the formation of preferences depends on individuals' ability to acquire and process data. Hence, erroneous perceptions are likely to have consequences for equilibria in many models with behavioral economic components. In this section, we describe how erroneous perceptions influence redistribution preferences, discuss empirical problems of using income to study preferences related to economic self-interest, and describe how using social status can mitigate some of these problems.

## 2.1 The basic model

The simplest possible illustration of the basic "workhorse" model for redistribution preferences (e.g. Romer, 1975; Meltzer and Richard, 1981) is a static economy in which individuals $i$ care only about their own consumption $c_i$, i.e.

$$U_i = U(c_i). \tag{1}$$

Individuals differ in their ability, which is reflected in a productivity parameter $\phi$ that determines individuals' pretax income $y(\phi)$. Normalizing wage rates to 1 and considering one unit of labor that is inelastically supplied, the income of individuals equals their productivity, i.e. $y_i = \phi_i$.

The government imposes a linear tax $t$ on income and pays lump sum benefits $r$. The government budget constraint requires that all transfers are covered by taxes, i.e.

$$r = \bar{y}t = \bar{\phi}t, \tag{2}$$

where $\bar{y}$ and $\bar{\phi}$ are average income and average productivity of all individuals in the economy. Suppose that the distortionary cost of taxation per person is $wt^2$ (Alesina and Giuliano, 2011), which yields the budget constraint of individual $i$

$$y_i = c_i = (1-t)\phi_i + r - wt^2 = (1-t)\phi_i + \bar{\phi}t - wt^2. \tag{3}$$

Each individual prefers a tax rate that maximizes consumption. Re-arranging the first-order condition

$$0 = \frac{\partial c_i}{\partial t} = \phi_i - \bar{\phi} - 2wt \tag{4}$$

yields individual's preferences for redistribution

$$t_i = \frac{\bar{\phi} - \phi_i}{2w}. \tag{5}$$

The critical determinant of individuals' redistribution preferences is the distance between the average productivity and the own productivity. Individuals prefer higher redistribution when the distance $(\bar{\phi} - \phi_i)$ is larger and vice versa.

## 2.2 The role of misperceptions

### 2.2.1 Misperceptions and preferences for redistribution

Under perfect information, $\bar{\phi}$ is known to individuals, in which case the preferred tax rate can easily be derived. In many cases, however, information is imperfect and costly to acquire. Hence, individuals need to make inferences about $\bar{\phi}$ using information they can observe. Sophisticated individuals may apply Bayes' rule to infer the income distribution from the entire population. A well-documented phenomenon, however, is that statistical inferences are drawn based on heuristics that are very imprecise (Kahneman and Tversky, 1972; Kahneman et al., 1982). In particular, "naïve" individuals will simply use the information about the income distribution within their social reference group $S_i$ as if it were representative of the entire population (Cruces et al., 2013). This reference group consists of friends, colleagues, neighbors, family members, and other individuals with repeated interactions.

The key question is how the mean productivity level of individual $i$'s social reference group $\bar{\phi}^{S_i}$ is related to the population mean $\bar{\phi}$. The literature on social segregation suggests that individuals' income is similar to the mean income level of the population sub-sample which individuals can observe (e.g. Reardon and Bischoff, 2011), but may deviate greatly from the total population mean. This observation suggests that a "naïve" individual $\tilde{i}$ that is relatively poor infers a mean productivity level $\bar{\phi}^{S^P_{\tilde{i}}}$ that is lower than the population mean $\bar{\phi}$. Hence, this individual has upwards biased perceptions about her or his rank in the distribution. Regarding preference, it follows that $\tilde{i}$ underestimates the distance between the mean productivity level and the own productivity level

$$t_{\tilde{i}} = \frac{\bar{\phi}^{S^P_{\tilde{i}}} - \phi_i}{2w} < t_i = \frac{\bar{\phi} - \phi_i}{2w}, \tag{6}$$

which results in downward-biased preferences for redistribution. The same logic applies to relatively rich individuals, which underestimate their status within their reference group and hence demand more redistribution than would be rational under perfect information.

### 2.2.2 Relative consumption concerns

Utility from consumption is often influenced as much by its contrast with a reference point of consumption as by consumption itself (Kahneman and Tversky, 1979). The argument of utility from relative consumption goes back to the theory of social interaction (Becker, 1974). Individuals aspire to improve their position in social rankings and get disutility if they are surrounded by others with higher status (Oswald, 1983, Luttmer, 2005; Carlsson et al., 2007). With relative consumption concerns, the utility function of Equation (1) becomes (e.g. Clark et al., 2008)

$$U_i = U(u_1(c_i), u_2(c_i|\bar{c})). \tag{7}$$

A simple way to represent (dis-)utility from consumption of others is modeling

$$u_2 = -n_i(1-t)(\bar{\phi} - \phi), \tag{8}$$

where $n_i \in (0,\, 1)$ is a preference parameter that denotes individual $i$'s disutility from consumption of peers. While a larger distance of $i$'s productivity from that of the society average results in greater disutility, the effect is mitigated by lower tax payments of $i$ relative to the society mean.

Augmenting the maximization problem of Equation (3) by relative consumption concerns of Equation (8), the preferred tax rate of $i$ adjusts to (see appendix A.1 for details)

$$t_i = \frac{(\bar{\phi} - \phi_i) + n(\bar{\phi} - \phi_i)}{2w}. \tag{9}$$

When individuals do not care about consumption of others ($n = 0$), Equation (9) becomes the standard problem of Equation (3). However, the preferred tax rate additionally increases with greater preference parameter $n$ when productivity of $i$ falls short of the average productivity level.[5] Again, a key question is whether $i$ correctly infers $\bar{\phi}$. When $i$'s social reference group has lower productivity than the total population mean, the tax-increasing effect of relative consumption concerns is mitigated.

### 2.2.3 The central hypothesis

Our simple model results in the following central hypothesis regarding the effect of misperceptions on preferences for redistribution:

**Hypothesis 1 (H1):** *Upward-biased (downward-biased) perceptions of $\phi_i$ decrease (increase) preferences for redistribution.*

This hypothesis follows from the baseline model (Section 2.2.1) and is reinforced by relative consumption concerns (Section 2.2.2).

## 2.3 Status and income

In a one-period model where individuals' budget constraint only depends on their own labor market participation and productivity, individuals' disposable income is $y_i = \phi_i$ and determines their preferences according to equations (5) and (9). However, in multi-period models, expectations about future productivity and income may influence the formation of preferences. Individuals' disposable resources may also be determined by other members of the family and income levels collected in surveys are often prone to misreporting. We next describe how using social status can mitigate these problems.

### 2.3.1 Multiple periods

When income ranks can change and redistribution policies are long-lasting, expectations about future positions on the income ladder influence current preferences for redistribution (Bénabou and Ok, 2001; Alesina and Giuliano, 2011). Over multiple periods, peoples' utility depends on consumption today ($t1$) and in one or more future periods ($t2$)

$$U_i = U(u_1(c_{it1}), u_1(c_{it2}), u_2(c_{it1}|\bar{c}), u_2(c_{it2}|\bar{c})). \tag{10}$$

When individuals expect to climb the income ladder ($c_{it2} > c_{it1}$ and $(c_{it2}|\bar{c}) > (c_{it1}|\bar{c})$), income in $t1$ poorly predicts redistribution preferences. Rather, preferences are formed based on the expected lifetime income. Annual income at a given point in time, however, is weakly correlated with lifetime income, giving rise to a "life-cycle bias" (Haider and Solon, 2006; Böhlmark and Lindquist, 2006). Evidence for the dominating role of expected

[5] This result is similar to studies showing that increased concern for relative consumption increases marginal tax rates (Boskin and Sheshinski, 1978; Ljungqvist and Uhlig, 2000; Aronsson and Johansson-Stenman, 2008).

lifetime income over annual income for the formation of preferences is provided by Rueda and Stegmueller (2019). In contrast to annual income, the social status is flatter over the life cycle and can be viewed as a measure of permanent income (Weiss and Fershtman, 1998).

#### 2.3.2 Productivity of other family members

Disposable income of individuals often depends (partly) on the income and productivity of other family members. In this case, the income level that is decisive for preferences $y_i^P$ may deviate from individuals' market income by the amount of family redistribution $y_i^F$, i.e.

$$y_i^P = \phi_i + y_i^F \neq y_i. \tag{11}$$

Consistent with the argument of equation (11), empirical evidence shows that compared to individual income, family income is the relatively stronger negative correlate of redistribution preferences (Alesina and La Ferarra, 2005). One way to address the role of family income in empirical studies is using household income instead of individual-level income. A remaining problem, however, is that individuals benefit from family income even though they may not live in the same household. A large body of literature has studied the role of private inter-household income transfers between family members (e.g. Cox, 1987; Cox et al., 2004). When private transfers are substitutes to public transfers (Becker, 1974), personal income or household income may underestimate preferences that are driven by economic self-interest. The social status, which is influenced by family background regardless of whether individuals share their household with other family members, reduces this bias.

#### 2.3.3 Reporting bias in social surveys

The implicit assumption when empirically examining Hypothesis (1) is that respondents in social surveys correctly report their income numbers. Recent validation studies show, however, that survey misreporting is pervasive when households are asked about their income (Mayer et al., 2015; Mittag, 2019). In the presence of misreporting, the social status is a more reliable measure than income, as it is based on several socio-economic characteristics and hence more difficult to manipulate.

#### 2.3.4 Need for a global setting

Our simple framework has some shortcomings. Support for redistribution may also depend on factors beyond economic self-interest, including considerations about fairness and assessments related to the causes of indigence and wealthiness (such as the well-documented perceptual differences about "luck versus effort", e.g. Alesina and Angeletos, 2005; Lefgren et al., 2016). These factors are influenced by cultural norms and vary systematically across countries (Gründler and Köllner, 2020). Redistribution preferences may also be driven by a country's institutions, political history, or geography (Gründler and Köllner, 2017). Accounting for these and other unobserved sources of cross-country heterogeneity requires examining the relationship between misperceptions and preferences in a global cross-national setting.

# 3 Measuring misperceptions on the globe

## 3.1 Social stratification and measurement of social status

The relative social position of individuals ("social stratification") is a fundamental building block of many classical theories in the social sciences (Hobbes, 1651; Marx, 1867; Weber, 1922; Duesenberry, 1949; Mills, 1651). Economic models that include relative status components have great overlap with sociological and psychological theories, where status considerations often are central themes (see, e.g., Saenger, 1945; Bradley and Corwyn, 2002; Chan and Goldthorpe, 2007). Modern sociologists build on the Weberian notion of status, which consists of prestige-based and resource-based factors. Prestige-based factors measure peoples' relative socio-economic standing in society. Resource-based factors instead include income, wealth, family affluence, and education (Chan and Goldthorpe, 2007). Although these components are correlated, a key lesson from prior studies is that they are linked differently to preferences and economic outcomes. Hence, the literature concludes that the social status is a multi-dimensional phenomenon and *"individual indicators can only capture part of its global meaning"* (Moreno-Maldonado et al., 2018, p.520). To address this argument, empirical studies usually aggregate several socio-economic characteristics to obtain composite measures of social status.

**Composite measures of social status:** A popular measure of social status is *Duncan's Socioeconomic Index* (SEI, Duncan, 1961), which combines data on education, income and occupational prestige. The SEI has been dominantly used in sociological and psychological research and was adjusted in several steps during the 1980s and 1990s (e.g. in Stevens and Featherman, 1981). A related measure is the *Nam-Powers-Boyd Occupational Status Scale* (OSS, Nam and Boyd, 2004). Originally developed at the Census Bureau in the late 1950s, the OSS is designed to reflect the average education and income of incumbents of each detailed occupation. Another widely used indicator is the *Hollingshead Index* (HI), which aggregates data on occupation, education, gender, and marital status (Hollingshead and Redlich, 1958; Hollingshead, 1971). Since the introduction of the Programme for International Student Assessment (PISA) in 2000, the *Index of Socioeconomic, Social,* *and Cultural Status* measures the socio-economic status of each child participating in the PISA study. The index combines information on parental income, education, occupational prestige, and social capital measured by the number of books available at home.

### 3.2 Unsolved challenges in the classification of social status

The most challenging task in the measurement of social status is aggregating its multiple characteristics into a uni-dimensional index. Previous indices that are based on conventional aggregation schemes have been criticized for methodological shortcomings (Oakes and Rossi, 2003; Marks, 2011). The problem of aggregation is finding a functional relationship

$$\mathfrak{F}^{\mathcal{S}} : \mathcal{X} \subseteq \mathbb{R}^m \to \mathcal{S} \subseteq \mathbb{R}, \tag{12}$$

that links the socio-economic attributes $\mathbf{x} = (x^1, \ldots, x^m) \in \mathcal{X}$ to the social status $S \in \mathcal{S}$. Naturally, each assumption on the functional relationship $\mathfrak{F}(\cdot)$ influences the classification outcome. As the "true" function is unknown, researchers have to make assumptions on the functional form and the relative importance of the underlying characteristics, defining weights $\omega_j$ (Decanq and Lugo, 2013). In most cases, the aggregation scheme is either additive

$$\mathfrak{F}^{\mathcal{S}}_{\text{Add}} = \omega_1 x_1 + \ldots + \omega_m x_m,\ \omega_j \geq 0, \tag{13}$$

or multiplicative

$$\mathfrak{F}^{\mathcal{S}}_{\text{Mul}} = x_1^{\omega_1} \times \ldots \times x_m^{\omega_m},\ \omega_j \geq 0, \tag{14}$$

or a combination of both. The choice of the aggregation scheme depends on the initial assumption about whether the attributes $\mathbf{x}$ are substitutable or represent necessary conditions of social status.

Conventional approaches for data aggregation face four major challenges: (i) little information can be acquired from theory about the relative importance of the socio-economic characteristics and their linkage, (ii) there is redundancy in Equations (13) and (14) due to correlations between the variables, (iii) there is no possibility to account for measurement uncertainty which typically accompanies social surveys, and (iv) both schemes result in distortions at the extreme ends at the top and the bottom.

Principal component analyses (PCAs) address some of the arbitrariness involved in the choice of $\omega_j$ (see Vyas and Kumaranayake, 2006). PCAs are statistical techniques to reduce the dimensionality of datasets based on correlations of variables. Data aggregation with PCAs, however, has two methodological disadvantages: first, it is built on the assumption that the first principal component is sufficient to predict the socio-economic status. Second, PCAs do not allow for non-linear relationships between the input variables and the social status.

## 3.3 Using machine learning to measure social status and misperceptions

We introduce a new procedure for measuring social status that tackles the challenges of conventional methods. Our algorithm (i) discards ex ante assumptions about the functional relationship between the characteristics and the social status, (ii) computes aggregation functions that are (highly) non-linear and (iii) produces confidence intervals that address measurement uncertainty in social surveys. Our method is built on machine learning techniques, which are designed to "learn" the functional relationship without being explicitly programmed (Breiman, 2001). Specifically, we use Support Vector Machines (SVMs), a machine learning technique of pattern recognition that finds the function

$$\mathfrak{F}(\mathbf{x}_i) \stackrel{!}{=} z_i \;\forall i = 1, \ldots, n, \tag{15}$$

which links a set of input variables $\mathbf{x} = (x_1, \ldots, x_m) \in \mathcal{X} \subseteq \mathbb{R}^m$ to an outcome $z \in \mathcal{Z} \subseteq \mathbb{R}$ for observations $i$ in sample $\mathcal{A} = \{(\mathbf{x}_i, z_i) | i = 1, \ldots, n\}$.

SVMs are supervised learning models, meaning that the learning algorithm classifies data points based on a set of training examples with known outcome. This training set is required for the algorithm to "learn the rule" that maps $\mathbf{x}$ onto $z$. The properties of SVM techniques are well-documented in the mathematical and computer science literature (see, e.g., Abe, 2010; Smola and Schölkopf, 2004; Vapnik, 1995, 1998). A brief introduction into the mathematical foundation of SVMs is provided in appendix A.2.1. The appendix also describes all technical features of our algorithm.

Our strategy to compute machine learning indicators of social status builds on "SVM aggregation" introduced by Gründler and Krieger (2016, 2021a,b), who designed a specific SVM-based algorithm to classify democratic institutions. This approach has been shown to outperform traditional techniques of data aggregation (Gründler and Krieger, 2021a).

### 3.3.1 Conceptualization, operationalization, and data source

For conceptualization, researchers need to define (i) the attributes $\mathbf{x} = (x_1, \ldots, x_m) \in \mathcal{X} \subseteq \mathbb{R}^m$ that underlie the social status and (ii) the relationship between these attributes. We follow the definition of the American Psychological Association (APA) to use a conceptualization that has proven to be of practical relevance. The APA's definition is closely related to the Weberian view, where the relative social position is influenced by (i) education, (ii) income, and (iii) occupational status (APA, 2007).

We use data from the World Value Survey (WVS) to obtain attributes of individuals that reflect our concept of social status. The WVS is the most extensive cross-country collection of data measuring individuals' beliefs, values, and well-being. At the time of this study, the WVS provides data on individuals from 97 countries (including 24

OECD-countries), covering 341,271 households that are representative for roughly 90% of the world population. The WVS covers a large set of questions related to the socio-economic background of respondents and includes the perceived socio-economic status of individuals, which is essential for computing the degree of misperception in the final step. Table (B-1) in the appendix lists our set of variables used to classify the social status and their numbers in the WVS's questionnaire.[6]

### 3.3.2 Training the learning algorithm

Our supervised learning machine is trained based on a subsample of individuals for which the social status can be determined with sufficiently large probability (the *training data*). Based on this data, the SVM algorithm uses the socio-economic characteristics discussed in Section (3.3.1) to construct an optimal (non-linear) hyperplane that classifies all observations in the dataset. We follow Falk et al. (2020) in employing a combination of education and income

$$p_{ij} = \text{education}_{ij} \times \text{income}_{ij}, \tag{16}$$

to identify the extreme ends of the status distribution and use the top and the bottom decile as training data for country $j$. We verify this strategy by a wisdom-of-the-crowd (WotC) approach based on text analyses of more than 2,500 Twitter Tweets on social status (see appendix A.2.1 for details on the label selection and A.2.2 for details of the WotC approach).

### 3.3.3 Aggregation

We classify the social status of individuals on the continuous interval $\mathcal{S} = \{0, 1\}$. The continuous scale ensures a detailed measurement of the social status. We draw random samples from our training data to compose the $\zeta$th training set $\mathcal{T}_{\zeta,j}$ for each country $j$. These training sets are used to obtain the machine learning classification function

$$\mathfrak{F}^{\mathcal{S}}_{j\mathcal{T}_{\zeta,j}} : \mathcal{X} \to \{0, 1\}. \tag{17}$$

We employ the classification function $\mathfrak{F}^{\mathcal{S}}_{j\mathcal{T}_{\zeta,j}}$ to compute a continuous index of social status for each observation of the dataset

$$S^{\mathcal{S}_{\{0,1\}}}_{ij_{\zeta}} = \mathfrak{F}^{\mathcal{S}_{\{0,1\}}}_{j\mathcal{T}_{\zeta,j}}(\mathbf{x}_{ij}) \ \forall (i, j). \tag{18}$$

[6]The selection of variables refers to (i) accurate measurement of our definition of status and (ii) availability of data in the WVS. In principle, the WVS includes a number of additional variables that potentially influence the social status. These variables, however, are available only for a (small) subset of households. As Support-Vector techniques require balanced panels, we cannot include this additional information in our analysis without a disproportionately large loss of observations.

Finally, we repeat each aggregation step for iterations $\zeta = 1, \ldots, 200$ to obtain a distribution $\Phi_{SS}$ of the status for each individual $i$. We use the mean of all realizations as a point estimate, which we henceforth refer to as the *"Support Vector Machines Socio-Economic Status (SVMSES)"*

$$\text{SVMSES}_{ij} = \frac{1}{\zeta_{\max}} \sum_{\zeta=0}^{\zeta_{\max}} S_{ij_\zeta}^{\mathcal{S}_{\{0,1\}}}.$$

We repeat this procedure for every wave included in the WVS (time subscripts are omitted to economize the description).

### 3.3.4 Properties and Interpretation

The social status, like any latent variable in the social sciences, can only be "objective" with regard to a specific contextual framework. Hence, when referring to the objective status, we henceforth mean objective in the sense of our definition of status. This interpretation is consistent with the large body of literature that uses observable socio-economic characteristics to derive measures of objective status (e.g. Jackman and Jackman, 1973).

The SVMSES delivers a continuous classification of the social status on the $\{0, 1\}$ interval. The closer the SVMSES to 0, the lower is the socio-economic status and vice versa. The SVMSES is a relative measure of status, reflecting the socio-economic position relative to other members of society. Due to the heterogeneity of countries in the WVS, we compute the SVMSES separately for each country, i.e. we specify the relevant peer group to be the population in country $j$.

We compute classifications based on traditional aggregation techniques to examine how our technique relates to alternative approaches. The comparison uncovers distinct differences in the classification outcomes that are in line with our motivation for introducing the machine learning technique (see Appendix A.3 for a comparison and description).

### 3.3.5 Computing the degree of misperception

To derive the degree of misperception for individual $i$ relative to other individuals in country $j$, we compare individuals' classified status with their self-perceptions of status. The WVS provides data on the subjective assessments of social status, where individuals are asked to classify their perceived status on a scale running from 1 to 5. We recode this variable so that higher values point to a higher level of perceived social status ($\text{Subjective}_i \in \{1, 5\}$). We then recode the SVMSES to match the co-domain of $\text{Subjective}_i$ and compute the degree of misperception via

$$\text{Mispercept}_i = \text{Subjective}_i - \text{SVMSES}_i^{\mathcal{R}},$$

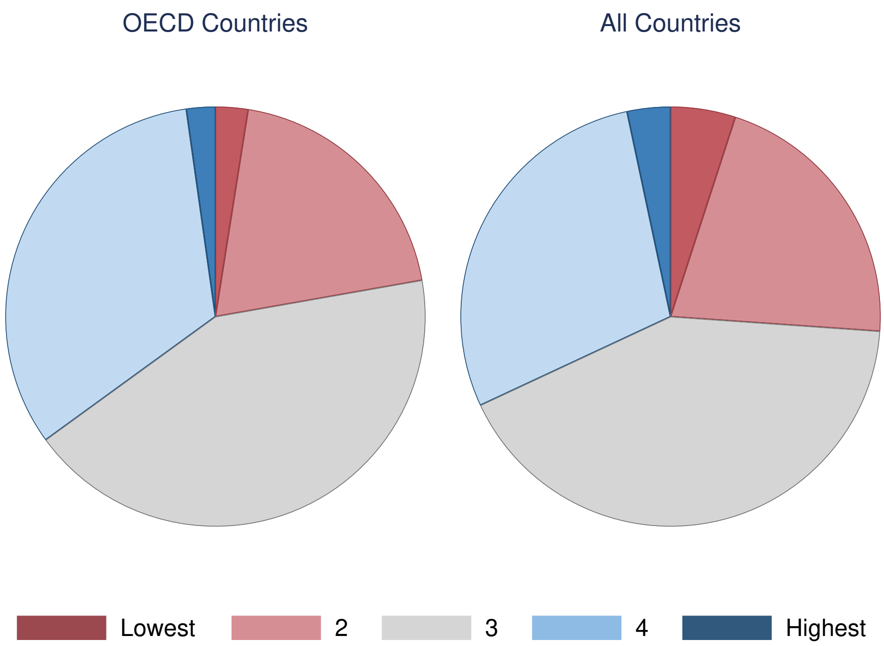
where positive (negative) values of $\text{Mispercept}_i$ indicate that individuals overestimates (underestimates) their individual socio-economic position.

# 4 A global view on perceptions and misperceptions of socio-economic status

## 4.1 Subjective perceptions of social status

Figure (1) shows the subjectively perceived social status by individuals for the sample of individuals living in an OECD country (left panel) and the full sample of individuals (right panel). Most individuals perceive to be middle class and only very few report to be at the very bottom or the very top. More than 40% of all respondents in the OECD and about 37% of the individuals in the full sample perceive to be exactly in the middle of status distribution. In contrast, only very few individuals think they belong to the upper 20% or the lower 20%.

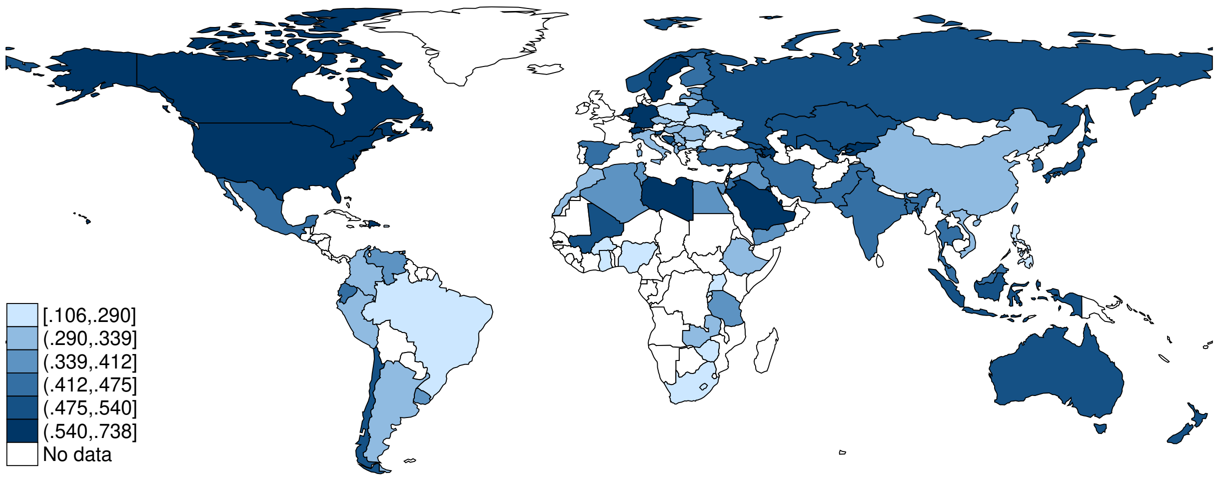
**Figure 1** SUBJECTIVE PERCEPTIONS ABOUT SOCIAL STATUS, OECD AND THE WORLD




*Notes*: Subjective perceptions about social status in the OECD (panel on the left-hand side) and the full sample of countries (panel on the right-hand side). The figure shows self-assessment of social status by individuals on a five-scale ladder (running from 1, the lowest possible level, to 5, the highest possible level) for the sample of OECD countries (left panel) and the world (right panel). The graph uses information on all individuals living in an OECD country (left panel, $N = 65,809$) and the full sample of all individuals (right panel, $N = 241,757$).

Figure (1) also shows that more individuals consider themselves being “almost at the top” (class 4) than “almost at the bottom” (class 2). Figures (C-1) and (C-2) in the appendix report individual’s self-positioning for each of the 24 OECD countries for which the WVS provides information on subjective assessments. These numbers show that the key patterns for the OECD re-appear for most countries, but they also reveal substantial cross-national differences. More than half of the population perceive to be in the upper two classes in the United States, Germany, and Switzerland, whereas in many Eastern

**Figure 2** GLOBAL DISTRIBUTION OF MISPERCEPTIONS




*Notes*: The figure shows the global distribution of status misperceptions, displaying the proportion of individuals with upwards-biased perception per country. The classes are obtained based on the sextiles of the national distribution of misperceptions.

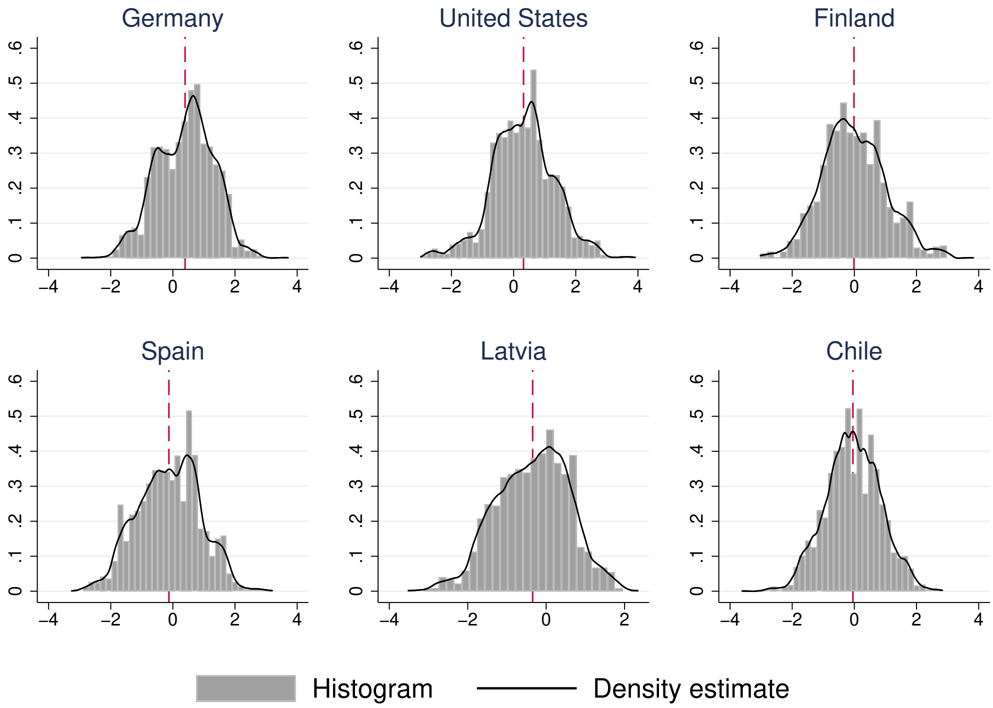
European countries, most individuals report to be either in the lowest classes or the middle class.

To measure the extent of within-country differences in perceptions, we compute Gini indices of self-reported social status (Figure C-4 in the appendix). Country-level differences in self-assessment are larger in non-OECD countries (average Gini index: 0.192) than in OECD countries (0.154), but there is heterogeneity also in the group of OECD countries. Differences in self-reported social status are pervasive in Poland, Finland, Italy, the United States, and Canada. At the other end of the spectrum, self-assessments are more homogeneous in Germany, the Netherlands, Switzerland, and South Korea. Self-perceptions also differ across occupation types (Figure C-3 in the appendix). In particular, perception gaps are sizable between manual and non-manual workers and between employees and managers or firm-owners.

## 4.2 Misperceptions of social status

In the next step, we compare the objective and the subjective status of individuals and explore the extent to which objective reality differs from individual's perceptions. Figure (2) shows the global distribution of misperceptions, illustrating the proportion of individuals with upwards-biased perceptions per country. The global average is 41.44%, meaning that the average individual on the globe underestimates her socio-economic status. The figure also reveals substantial cross-country heterogeneity in misperceptions. Perception biases vary by a standard deviation of 13.53 percentage points and reach from 10.70% (Uganda) to 73.87% (Saudi Arabia). Of the 3,828 country-pairs, more than half of the pairwise

**Figure 3** DISTRIBUTION OF MISPERCEPTIONS FOR SELECTED COUNTRIES



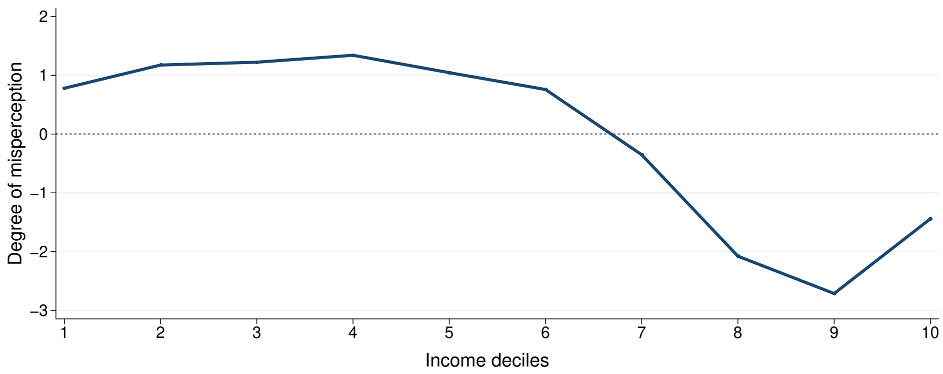

*Notes*: The figure plots the empirical distribution of $\text{Mispercept}_{ij}$ for selected countries. The figure shows the frequency (gray bars) and the kernel density (black line) of the degree of misperception in Germany, the United States, Finland, Spain, Latvia, and Chile. The Epanechnikov kernel is used for the estimation of the empirical density. Vertical red lines represent country averages.

comparisons deliver $t$-tests that are statistically significant at the 1% level. This heterogeneity suggests that there are country-specific characteristics (institutions, culture, history, geography) that influence a society's propensity to have erroneous self-beliefs, underlining the need to examine the consequences of misperceptions on a global scale.

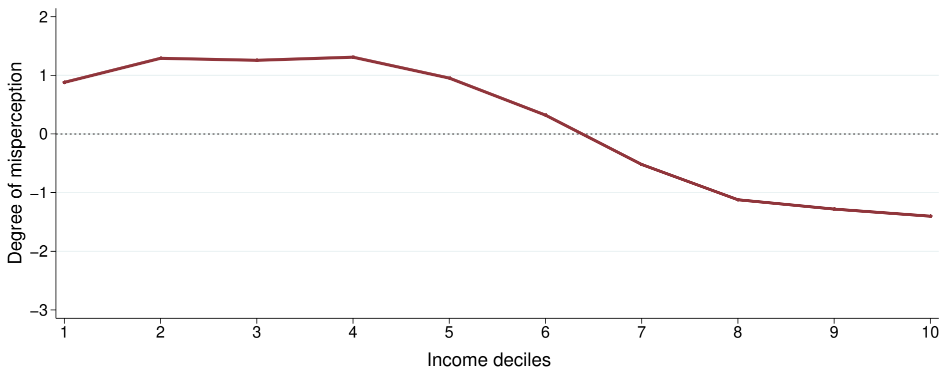
Table (B-2) in the appendix provides detailed summary statistics on the degree of misperception for each OECD country. These statistics show that the extent of bias differs substantially across individuals within countries. Figure (3) plots the distribution of individual-level misperceptions $\text{Mispercept}_{ij}$ for selected OECD countries in which the average individual overestimates (Germany and the United States), underestimates (Spain and Latvia), or correctly estimates (Finland and Chile) her status. Despite the differences in means, the figure shows that both over- and underestimation of the socio-economic status is widespread in each country. The figure also suggests that a substantial portion of individuals classify their status quite accurately. Similar pattern arise for the sumpsamples of OCED and non-OECD countries (Figure C-5 in the appendix). These patterns are consistent with the results of experimental studies (e.g. Karadja et al., 2017).

Experimental studies conducted for Argentina and Germany report correlations between incomes and misperceptions, with relatively poor individuals overestimating their own rank and vice versa (Cruces et al., 2013; Engelhardt and Wagener, 2017). A similar

**Figure 4** MISPERCEPTIONS ACROSS INCOME DECILES, UNITED STATES AND GERMANY




**(a)** Misperceptions and income distribution for the United States.



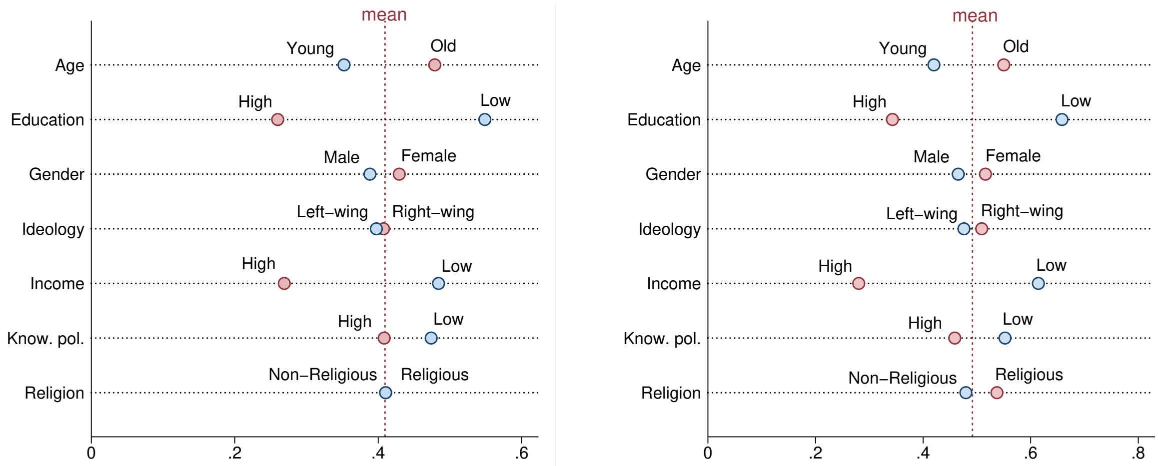

**(b)** Misperceptions and income distribution for Germany.

*Notes*: Misperceptions at different income deciles in the United States and Germany. The figure shows the average degree of Mispercept$_{ij}$ for different income deciles. Detailed numbers are presented in Table (B-3) in the appendix.

pattern is visible in our data. Figure (4) illustrates the average level of misperceptions across objective income deciles for the United States and Germany. The underlying data is reported in Table (B-3) in the appendix, which also shows the proportion of households with a positive bias. Individuals with relatively low social status tend to overestimate their position, while households with higher status tend to underestimate their position. Also, the overwhelming fraction of households underestimate their socio-economic status in the upper deciles. This is a general pattern that is visible for all OECD countries. The data for the OECD as a whole suggests that roughly 70% of the individuals in deciles 1–3 overestimate their individual status, whereas this proportion declines to approximately 20% in deciles 8–10 (Figure C-6 in the appendix).

Using the numbers for Germany (Engelhardt and Wagener, 2017) and Sweden (Karadja et al., 2017) for cross validation, we find that our approach delivers classifications of misperceptions that are close to the results of experimental settings. For Germany, the proportion of the population with a positive bias in the lowest decile is almost identical

**Figure 5** SHARE OF INDIVIDUALS WITH A POSITIVE BIAS




**(a)** Full sample.

**(b)** OECD countries.

*Notes*: Share of individuals with a positive bias for key socio-economic characteristics, all individuals in the sample (Panel a) and individuals living in the OECD (Panel b). The figure shows the share of individuals that over-estimate their status depending on age, education, gender, political ideology, income, self-reported knowledge of current political events, and religion. The vertical grey line represents the sample average.

(0.723 in our study versus 0.695 in the experiment of Engelhardt and Wagener) and the similarity persists also for other deciles. The results are also in line with patterns reported for Sweden (see Figure C-7 in the appendix).

In Figure (5), we analyze whether status misperception is related to socio-economic characteristics. The figure shows the share of individuals with a positive bias by social and economic factors. This comparison reveals that the propensity to overestimate the social status is higher for older and for less educated individuals. Moreover, overestimation is more prevalent among women and (particularly in the OECD) among individuals that favor right-wing political parties. We also observe that individuals that are better informed about current political events overestimate their status to a lesser extent.

# 5 Misperceptions and preferences for redistribution

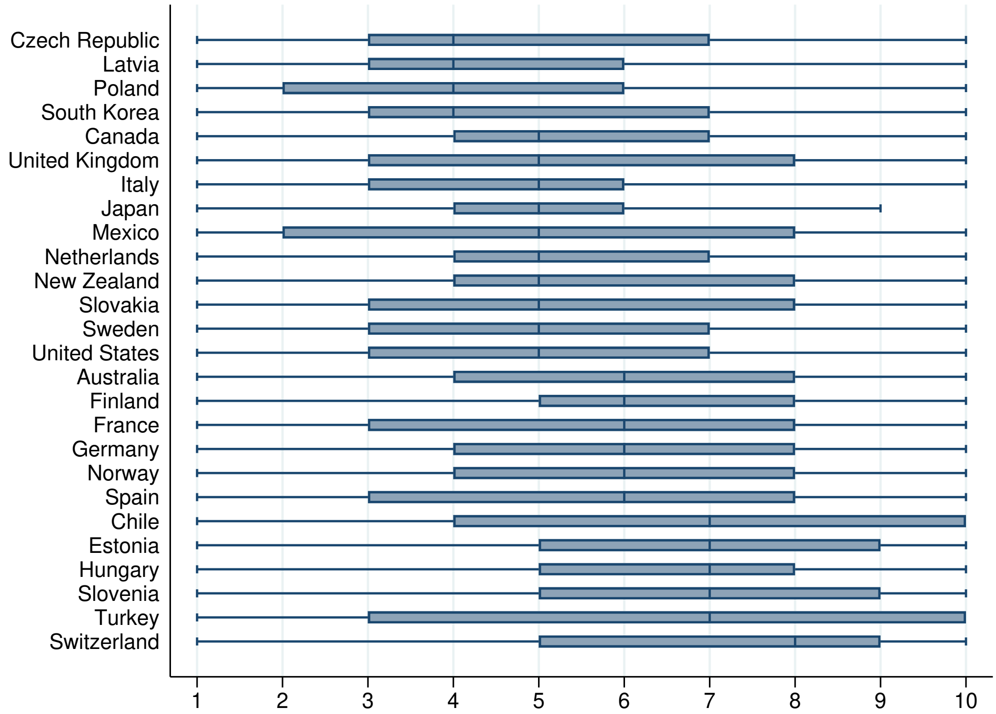
We next turn to the consequences of misperceptions and examine whether erroneous beliefs about the own status influence preference formation.

## 5.1 Data on redistribution preferences

We measure redistribution preferences on a scale running from 1 to 10, with larger values reflecting higher support for redistributive policies (appendix A.4 provides details of the coding and the underlying question from the WVS).

Figure (6) shows the distribution of redistribution preferences and the median for

**Figure 6** PREFERENCES FOR REDISTRIBUTION, OECD




*Notes*: The figures shows preferences for redistribution across individuals living in the OECD. The graph displays the median (shown by the vertical marker) and the distribution (shown by the boxes; 25th and 75th percentile) of the distribution of redistribution preferences per country. The "whiskers" (horizontal lines) show the minimum and the maximum values. Redistribution preferences are measured on a scale between 1 and 10; higher values reflect higher support for redistribution policies.

OECD countries, pointing to large within-country heterogeneity in preferences. With the exception of Japan, preferences in each country span the entire spectrum from 1 to 10, and there are marked differences between the 25th and the 75th percentile of the national distributions (the average standard deviation is 2.79). We also observe substantial differences in support for redistribution between countries. The median reaches from a level of 4 in the Czech Republic to a level of 8 in Switzerland. Differences in redistribution are even larger in the sample of non-OECD countries (see Figure C-8 in the appendix).

**Social status and income:** The key hypothesis underlying our theory in Section (2) is that social status is more reflective of redistribution preferences in terms of economic self-interest than income. We examine this hypothesis in Table (B-4) in the appendix. The table shows that both income and social status are negatively related to redistribution preferences, but the parameter estimate for social status exceeds that for income by a factor of 3.5 (columns 3 and 4). This result suggests that social status is the relatively stronger negative correlate of redistribution preferences.

## 5.2 Empirical strategy

### 5.2.1 Baseline specification

Our empirical specification brings our central hypothesis regarding the link between status misperceptions and redistribution preferences to the data (Hypothesis 1). An important challenge of the empirical model is to disentangle components of preferences that reflect biased perceptions from those that reflect cultural norms, institutions, or economic self-interest. We address these issues in our parsimonious model specification

$$\text{Pref}_{it} = \gamma \text{Mispercept}_{it} + \delta \text{SVMSES}^{\mathcal{R}}_{ijt} + \mathbf{X}_{it}\boldsymbol{\beta} + \eta_j + \zeta_t + \varepsilon_{it}, \tag{19}$$

where the dependent variable, $\text{Pref}_{it}$, denotes preferences for redistribution of individual $i$ in country $j$ at time $t$. Our key variable of interest is $\text{Mispercept}_{it}$, the extent of status misperception. To account for national cultural norms, political institutions and social security systems, we include country-fixed effects $\eta_j$. These effects also absorb any other source of cross-country heterogeneity in time-invariant unobservables and eliminate the influence of differences in the design or the wording of questionnaires across countries, which can occur in the translation process. The model also includes time-fixed effects $\zeta_t$ to eliminate methodological differences between the waves of the WVS and to account for period-specific shocks and trends in preferences.

In Section (2.3), we described how using the social status to determine individual's expected benefit or loss from redistribution policies can mitigate conceptual problems regarding the use of annual income data. We address these argument by including the socio-economic status $\text{SVMSES}^{\mathcal{R}}_{ijt}$ into the model. The taste of redistribution has been shown to depend also on other socio-economic and demographic characteristics of individuals (Alesina and Giuliano, 2011; Corneo and Grüner, 2002; and Fong, 2001). We include these factors in the vector of individual characteristics $\mathbf{X}_{it}$.[7] Standard-errors are robust to arbitrary heteroskedasticity and adjusted for clustering at the national level.

### 5.2.2 Identification

Identifying an effect of misperceptions on redistribution preferences based on Equation (19) is challenging because there may be unobserved confounders that influence both self-perceptions and preferences. To the extent that these confounders correlate with other observable individual-level characteristics, we absorb them by our control variables. To the extent that confounding occurs because of country-specific characteristics, we eliminate them by our country fixed effects. We also eliminate global trends in preferences by

[7] Our list of controls includes age, gender, marital status, children in the household, a dummy variable for divorced, a dummy variable for widowed, a dummy variable for retired, and a dummy variable if the respondent is the chief wage earner. Further individual-level characteristics are available in the WVS, but with substantially fewer observations.

including wave-fixed effects.

A threat to the validity of our results is that preferences and self-perceptions differ across sub-national regions and cohorts, and that trends in preferences may be country-specific and cohort-specific. We address these concerns in a set of robustness analyses by successively augmenting Equation (19) by fixed effects for first-level administrative regions ($\rho_r$) and birth years ($\psi_a$). We also follow Guiliano and Spilimbergo (2014) by accounting for region-specific cohort effects, implemented by fixed effects for interactions between sub-national regions and birth years ($\rho_r \times \psi_a$). Finally, we control for time-varying unobservables on the country level by including fixed effects for interactions between countries and years ($\eta_j \times \zeta_t$). The least parsimonious model is given by

$$\begin{aligned}\text{Pref}_{it} =& \gamma\text{Mispercept}_{it} + \delta\text{SVMSES}^{\mathcal{R}}_{it} + \mathbf{X}_{it}\boldsymbol{\beta} + \eta_j + \zeta_t + \rho_r + \psi_a \\ &+ \rho_r \times \psi_a + \eta_j \times \zeta_t + \varepsilon_{it}.\end{aligned} \tag{20}$$

Including fixed-effects for cohorts ensures that the estimates are not driven by changes in political preferences over the lifetime, e.g. when people become more conservative when they get older (e.g. Roth and Wohlfahrt, 2018). Country-year fixed effects absorb specific macroeconomic conditions of countries that affect all individuals in the sample. Including fixed effects for regions, countries, country-years and region-cohorts also addresses potential socialization effects and trends of national welfare systems that may shape redistribution preferences.

As a complementary strategy to rule out that our estimates are counfounded by national social security systems and cultural norms, we also estimate models using a sub-sample of first- and second-generation migrants that have been socialized in welfare regimes other than that of their country of residence.

A remaining concern is that there are individual-specific personality traits that influence both misperceptions and beliefs. We control for socio-economic, national and regional characteristics and also include variables measuring individual's view of the world that may correlate with unobserved personality traits. Nevertheless, we cannot fully rule out that the results are influenced by unobserved individual-level factors.

## 5.3 Results

Table (1) reports our baseline results for the OECD countries, including data from 65,809 individuals living in 24 countries. The model in Column (1) associates preferences with the subjectively perceived social status of individuals. The parameter estimate is negative and statistically significant at the 1% level ($t = 13.13$). Numerically, the coefficient suggests that a one unit increase in perceived social status is related to a 0.428 unit decrease in preference for redistribution, measured on a scale from 1 to 10. This result underlines that self-perceptions of individuals are significantly correlated with support for redistribution. Subjective perceptions of status consist of two components, objective criteria and erroneous beliefs. A drawback of Model (1) is that the estimated parameter captures both components; it does not allow to separate between the two. In Column (2), we separately account for the objective status and the extent to which individuals held misperceptions. Consistent with motives of self-interest, the objective status is negatively related to the support of equalizing policies. The coefficient on misperceptions is also negative and highly statistically significant ($t = 9.51$). A one unit increase in our measure of misperceptions is associated with a reduction in redistribution preferences by 0.302 units. This result underscores that parts of the negative correlation between self-perceptions and redistribution preferences are due to erroneous beliefs.

Column (3) includes economic and demographic controls. The parameters of these variables are in line with earlier studies on preferences (see, e.g., Luttmer and Singhal, 2011). Women and older individuals have higher preferences for equality, while married couples and parents are less supportive of income redistribution. In contrast, being retired, widowed or divorced does not correlate with preferences. The parameter estimate for $\text{Mispercept}_i$ retains its economic and statistical significance when we account for other predictors of preferences.

In Table (B-5) in the appendix, we re-estimate our baseline results using data for all available households ($N$ = 241,757) and all 97 countries in our sample. Doing so has little influence on inferences. The parameter estimate for status misperception declines by about 50% in the full sample ($\gamma = -0.145$), indicating that the misperception effect is weaker in non-OECD countries compared to OECD countries.

Taken together, the results are in line with our hypothesis that upward-biased perceptions of social status are negatively correlated with preferences for redistribution. The results suggest that about 9% of the variation in redistribution preferences can be explained by (biased) perceptions of status.

## 5.4 Robustness

Table (2) addresses threats to the validity of our baseline results. Row (1) shows the baseline outcome of our preferred specification (Column 2 of Table 1) as a benchmark and compares the result with those of less parsimonious specifications, estimating variants of Equation (20). In Row (2), we directly account for individual status variables and replace our indicator of social status ($\text{SVMSES}_{ij}^{\mathcal{R}}$) by its underlying components. Our baseline model controls for unobserved heterogeneity in time-invariant characteristics at the country-level, but it is also conceivable that there are persistent differences in wealthiness, cultural norms, local governments, and geographic conditions across sub-national regions within countries. We address the concern that these factors may be systematically

**Table 1** STATUS MISPERCEPTION AND REDISTRIBUTION PREFERENCES—BASELINE ESTIMATES, OECD COUNTRIES

Dependent variable: Subjective preferences for income redistribution, $\text{Pref}_{ij}$

| | (1) Perceptions | | (2) Misperceptions | | (3) Misperceptions and Controls | |
|---|---|---|---|---|---|---|
| | Coefficient | SE | Coefficient | SE | Coefficient | SE |
| $\text{Subjective}_i$ | -0.428*** | (0.0326) | | | | |
| $\text{Mispercept}_i$ | | | -0.302*** | (0.0318) | -0.308*** | (0.0319) |
| $\text{SVMSES}_{ij}^{\mathcal{R}}$ | | | -0.523*** | (0.0389) | -0.514*** | (0.0396) |
| Age | | | | | 0.004** | (0.0016) |
| Female | | | | | 0.173*** | (0.0407) |
| Divorced | | | | | -0.027 | (0.0770) |
| Married | | | | | -0.099** | (0.0464) |
| Widowed | | | | | -0.122 | (0.0727) |
| Child in household | | | | | -0.082* | (0.0417) |
| Chief wage earner | | | | | -0.035 | (0.0494) |
| Retired | | | | | 0.100 | (0.0512) |
| Observations | 65,809 | | 65,809 | | 65,770 | |
| Countries | 24 | | 24 | | 24 | |
| Country-fixed effects | Yes | | Yes | | Yes | |
| Wave-fixed effects | Yes | | Yes | | Yes | |
| R-squared | 0.091 | | 0.097 | | 0.098 | |
| F-Stat | 206.4*** | | 272.6*** | | 170.6*** | |

*Notes*: Robust standard errors adjusted for clustering by countries are in parentheses. The dependent variable is $\text{Pref}_{ij}$, which gives the subjective preference for income redistribution on a scale running from 1 (low) to 10 (high). $\text{SVMSES}_{ij}^{\mathcal{R}}$ denotes the SES measured with the SVM-algorithm. In order to rule out issues with multi-collinearity, the table does not include further measures of socio-economic status, such as income, education, and occupational status. "Age" gives the age of respondents in years, "Female" is a dummy variable that equals 1 if the respondent is female, "Divorced", "Married", "Widowed", and "Retired" are dummy variables that control for the corresponding status, "Child in household" equals 1 if a child lives in the household of the respondent, and "Chief wage earner" denotes whether the respondent is the chief wage earner of the household. The data refers to questions X001, X011, X007, X041, C006, and S002 of the consolidated dataset of the WVS and are recoded accordingly. All models include fixed effects for countries and waves.

*** Significant at the 1 percent level,
** Significant at the 5 percent level,
* Significant at the 10 percent level

correlated with misperception in Row (3) by including fixed effects for first-level administrative regions (ADM1).[8] Row (4) accounts for country-year fixed effects to alleviate concerns about time-varying unobservables on the country-level and regional differences in trends. Row (5) augments our baseline model by including fixed effects for birth years of respondents to account for differences in perceptions across age cohorts. In Row (6), we include region-specific cohort effects. Row (7) includes all fixed effects, estimating the fully specified model of Equation (20).

To address the concern that the results are driven by unobserved personality traits, we add a richer set of control variables that are potentially correlated with personality and fundamental views of the world in Row (8). We include individuals' political preferences, self-reported knowledge about current political developments, and religiousness. To rule out that the results are driven by observations at the very top or bottom (the "super-rich" and the "super-poor"), we also exclude the top-10% and bottom-10% of the social status distribution in Row (9).

A key assumption for the validity of our results is that fixed effects for countries, years, sub-national regions, region-cohorts and country-years absorb any elements of preferences that are determined by national cultural norms, political institutions, and social security systems. This assumption is plausible, given the large inertia of these factors. To further alleviate concerns about socialization effects and a potential influence of national security systems on (mis-)perceptions of individuals and their taste for redistribution, we restrict the sample in Row (10) to observations from first-generation migrants. This "epidemiological approach" eliminates correlations between institutions and national cultural norms with individuals' perceptions, as migrants have been socialized in institutional and cultural environments different to those of their country of residence. Because of the restricted number of first-generation migrants available in the WVS for OECD countries, Row (10) is based on all available countries in the sample. In Row (11), we specifically explore OECD countries by enlarging the sample to first- and second-generation migrants.

We observe almost no changes in the estimated parameter of misperceptions across Rows (1)–(11). The coefficients are all negative and statistically significant at the 1% level. Except for the estimates in Row (8.) that are based on all available countries, the parameters are statistically indistinguishable from the baseline estimates in Row (1).[9]

Another threat to the validity of our results comes from possible measurement errors in the WVS data. We address this concern in Table (B-6), where we present results of multiple estimations using all 200 SVM-iterations based on which $\text{Mispercept}_j$ and $\text{SVMSES}_i^{\mathcal{R}}$ are computed. We use multiple imputation (MI) regression techniques to account for the empirical distribution of our measures for social status and misperceptions.

[8] ADM1 regions are administrative entities one level below the central government (in many countries, these are referred to as "states").

[9] A series of Wald tests that compare the parameter estimates of Rows (2)–(11) with those of Row (1) all return p-values (substantially) larger than 0.5.

**Table 2** STATUS MISPERCEPTION AND REDISTRIBUTION PREFERENCES—ROBUSTNESS CHECKS

| Dependent variable: Subjective preferences for income redistribution, $\text{Pref}_{ij}$ | | | | |
|---|---|---|---|---|
| Specification | Coefficient of Misperception | SE | R-squared | $N$ |
| 1. Baseline (OECD) | -0.302*** | (0.0318) | 0.097 | 65,809 |
| 2. Baseline (OECD) with status variables | -0.281*** | (0.0302) | 0.100 | 65,809 |
| 3. Baseline (OECD) with sub-national FE | -0.291*** | (0.0292) | 0.113 | 57,974 |
| 4. Baseline (OECD) with country-year FE | -0.299*** | (0.0336) | 0.104 | 65,809 |
| 5. Baseline (OECD) with birth-year FE | -0.302*** | (0.0314) | 0.098 | 65,570 |
| 6. Baseline (OECD) with region-cohort FE | -0.296*** | (0.0315) | 0.286 | 51,493 |
| 7. Baseline (OECD) with all FE | -0.290*** | (0.0332) | 0.289 | 51,493 |
| 8. Baseline (OECD) with more controls | -0.329*** | (0.0441) | 0.092 | 24,651 |
| 9. Baseline (OECD), exclude 1st & 10th decile | -0.299*** | (0.0342) | 0.095 | 52,647 |
| 10. Baseline (all), migrants (1st) | -0.196*** | (0.0606) | 0.179 | 3,092 |
| 11. Baseline (OECD), migrants (1st & 2nd) | -0.327*** | (0.0799) | 0.145 | 4,209 |

*Notes*: Robust standard errors adjusted for clustering by countries are in parentheses. The dependent variable is $\text{Pref}_{ij}$, which gives the subjective preference for income redistribution on a scale running from 1 (low) to 10 (high). All specifications include fixed effects for countries and years. The specification in Row 1. reports our baseline estimate for the OECD of Table (1) for comparison. In Row 2., we replace our composite measure of social status by the individual variables underlying this indicator. These variables are explained in detail in Table (B-1) and include education, income, age, several variables to account for occupational status, and friendship networks. In Row 3., we include fixed effects for sub-national regions (first-level administrative regions) to account for differences in income, wealth, culture, and institutions within countries. Row 4. includes country-year fixed effects, Row 5. accounts for differences across cohorts by including fixed effects for birth years. In Row 6., we include a full set of cohort-region fixed effects. We include all fixed effects in Row 7., estimating the full model of Equation (20). In Row 8., we include a richer set of control variables, accounting for political ideology, self-reported knowledge about politics, and religion. In Row 9., we exclude the bottom-10% and the top-10% of the distribution of our status indicator. Note that we also re-classify the bottom-10% and top-10% of the status distribution. In Row 10., we restrict the sample to first-generation migrants as an alternative strategy to disentangle effects of misperceptions from those of national cultural norms and institutions. Because of data availability, we assess the misperception effect for first-generation migrants based on all available countries. We repeat the analysis of Row 10. using data on first-generation and second-generation migrants for the OECD countries in Row 11.

*** Significant at the 1 percent level,
** Significant at the 5 percent level,
* Significant at the 10 percent level

Each specification represents the results of 200 estimations that are combined by the Rubin (1987) method. Hence, the regressions rely on 13,069,000 (OECD) and 48,351,400 (all countries) data points. Row (1) reproduces the baseline results of Table (B-5) for comparison, while the subsequent rows show the outcomes of the MI variants of the fully specified (Row 2) and the reduced (Row 3) baseline models. Rows (4–6) show results using all available countries.

The outcomes of the MI regressions underscore that our results are not driven by measurement issues in the data. In each specification, misperceptions of social status are negatively related to redistribution preferences, and the effect is again significant on the 1% level. The parameter estimates of the fully specified MI regression for the OECD is -0.252, which is virtually identical to the baseline outcome (-0.251). Parameter estimates are also close to the baseline outcome when we run MI regressions on the large sample that includes all available countries (-0.126 and -0.128).

## 5.5 Heterogeneity

We next examine heterogeneity in the effect of biased perceptions on preferences to assess the probability that (unobserved) personality traits and preferences influence our estimates. Previous studies show that individuals' economic preferences vary systematically with observable socio-demographic factors (e.g. Falk et al., 2018). We explore whether the coefficient on misperceptions differs across key socio-economic characteristics that are potentially correlated with personality traits. To this end, we augment our baseline model by a set of interactions between misperceptions and socio-economic characteristics (denoted with $\mathrm{C}_{it}$) via

$$\mathrm{Pref}_{it} = \gamma \mathrm{Mispercept}_{it} + \psi\{\mathrm{Mispercept}_{it} \times \mathrm{C}_{it}\} + \pi \mathrm{C}_{it} + \delta \mathrm{SVMSES}_i^{\mathcal{R}} + \eta_j + \zeta_t + \varepsilon_{it}. \quad (21)$$

Table (3) shows the results, where each row represents a single regression with two variables of interest: the parameter estimate of $\mathrm{Mispercept}_i$ and the interaction term between $\mathrm{Mispercept}_i$ and the respective socio-economic variable.[10] We also report p-values on a test of whether the baseline parameter estimate and the coefficient in the rows of Table (3) are equal. We examine potential heterogeneity caused by income (Row 1), gender (Row 2), age (Row 3), the highest education level (Row 4), party affiliation, measured with a dummy that is 1 if respondents support right-wing candidates (Row 5), and religion, measured with a dummy variable that is 1 if respondents report that they consider themselves religious (Row 6). The interaction term between misperceptions and the relevant socio-economic variable is statistically indistinguishable from zero in each

[10]The socio-economic variables also enter in the regressions but are truncated to economize space. These variables are all not significantly related to redistribution preferences.

**Table 3** STATUS MISPERCEPTION AND REDISTRIBUTION PREFERENCES—EFFECT HETEROGENEITY

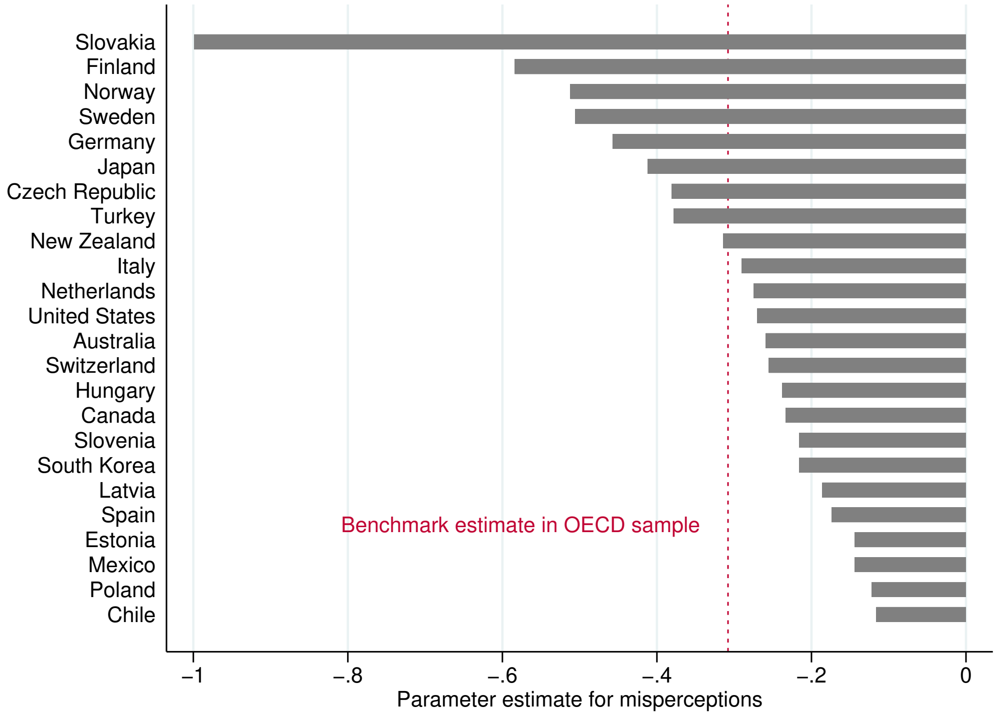
Dependent variable: Subjective preferences for income redistribution, $\text{Pref}_{ij}$

| Specification | Coefficient | (SE) p-val. | R-squared | $N$ |
|---|---|---|---|---|
| 1. By income | | | | |
| $\text{Mispercept}_i$ | -0.325*** | (0.0466) | 0.098 | 65,809 |
| $\text{Mispercept}_i \times \text{Income}_i$ | 0.009 | (0.0061) | | |
| *p-value on Wald test of equal parameters* | | 0.625 | | |
| 2. By gender | | | | |
| $\text{Mispercept}_i$ | -0.299*** | (0.0309) | 0.098 | 65,770 |
| $\text{Mispercept}_i \times \text{Female}_i$ | 0.012 | (0.0164) | | |
| *p-value on Wald test of equal parameters* | | 0.917 | | |
| 3. By age | | | | |
| $\text{Mispercept}_i$ | -0.267*** | (0.0451) | 0.097 | 65,809 |
| $\text{Mispercept}_i \times \text{Age}_i$ | -0.001 | (0.0009) | | |
| *p-value on Wald test of equal parameters* | | 0.450 | | |
| 4. By education | | | | |
| $\text{Mispercept}_i$ | -0.254*** | (0.0579) | 0.097 | 65,809 |
| $\text{Mispercept}_i \times \text{Education}_i$ | -0.010 | (0.0078) | | |
| *p-value on Wald test of equal parameters* | | 0.415 | | |
| 5. By party preference | | | | |
| $\text{Mispercept}_i$ | -0.269*** | (0.0348) | 0.113 | 57,885 |
| $\text{Mispercept}_i \times \text{Right-wing}_i$ | 0.0272 | (0.0308) | | |
| *p-value on Wald test of equal parameters* | | 0.355 | | |
| 6. By religion | | | | |
| $\text{Mispercept}_i$ | -0.295*** | (0.0334) | 0.095 | 64,841 |
| $\text{Mispercept}_i \times \text{Religious}_i$ | -0.032 | (0.0223) | | |
| *p-value on Wald test of equal parameters* | | 0.833 | | |

*Notes*: Robust standard errors adjusted for clustering by countries are in parentheses. The dependent variable is $\text{Pref}_{ij}$, which gives the subjective preference for income redistribution on a scale running from 1 (low) to 10 (high). All specifications include fixed effects for countries and years. The table reports heterogeneity in the effect of status misperceptions on redistribuition preferences, including interaction terms between status misperceptions and socio-economic characteristics (Equation 21). In Row (1), we examine whether the effect depends on the income level of individuals. In Row (2), we examine the effect separately for women and men; in Row (3), we explore whether the effect depends on the age of individuals; in Row (4), we examine differences across education levels. In Row (5), we study whether the misperception effect differs across political preferences, interacting misperceptions with preferences for right-wing parties. In Row (6), we explore differences between religious and non-religious individuals. The row entitle "*p-value on Wald test of equal parameters*" reports the p-value for a test of equality between the baseline estimates and the parameter estimates of the individual Row.

*** Significant at the 1 percent level,
** Significant at the 5 percent level,
* Significant at the 10 percent level

**Figure 7** MISPERCEPTION EFFECTS ACROSS OECD COUNTRIES




*Notes*: The figure shows the estimated coefficient on misperceptions across OECD countries. The figure reports results of equation (19) for OECD countries. The empirical specifications are identical to the model reported in Column (2) of Table (1).

regression. The coefficient on misperception remains negative and statistically significant at the 1% level in each model. We also do not observe any change in the size of the estimated parameter on misperceptions. For each model, the Wald test does not reject the null of equality of the baseline coefficients on misperceptions and those of Rows (1)–(6) of Table (3).

Taken together, the results provide no indication of effect heterogeneity. Rather, the misperception effect materializes irrespective of individual's socio-economic background.

## 5.6 Misperception effects across countries

A key advantage of our global perspective is that it delivers estimates with high external validity. An important question is whether there are differences in the misperception effect across countries. Figure (7) shows results from estimations that only take into account data from individual OECD countries. The underlying benchmark model is reported in Column (2) of Table (1). Upwards-biased perception decrease preferences for redistribution in each country. There are, however, large differences across countries. The effect is largest in Slovakia and in the Nordic countries, about the size of our benchmark estimate in the United States, the Netherlands, and Australia, and smallest in Latin America, Spain, Estonia, and Poland.

Taken together, the cross-country heterogeneity underscores the importance to account for unobserved country-level characteristics when examining the relationship between perceptions and preferences. Unobserved heterogeneity in the form of culture, institutions, and political history restrict the external validity in studies using data from singly countries. Despite the differences in the size of the estimate parameters, the coefficients are negative and statistically significant in each case, pointing to a general pattern in human behavior regardless of the country-specific background.

## 5.7 Alternative household data

A threat to the validity of our results is that the results may be driven by our selection of the World Value Survey as data source for our statistical analysis. In Table (B-7) in the appendix, we re-estimate our baseline models using micro data from the Life in Transition Survey (LITS). The LITS is compiled by the European Bank for Reconstruction and Development (EBRD) and has been collected in three waves (2006, 2010, and 2016). The dataset includes subjective assessments about income levels on a ten-scale ladder for all waves, but includes data on net household incomes only for the third wave.

Table (B-7) in the appendix shows that the results based on the LITS are almost identical to those obtained from the WVS. Similar to the WVS results, there is heterogeneity across countries. The cross-country patterns suggested by the LITS data is similar to those obtained by the WVS data (see Figure C-9 in the appendix). This similarity is reassuring, because it is unlikely that there are systematic cross-country biases in the survey design or methodology that are identical in the WVS and in the LITS.

## 5.8 Revealed preferences

A drawback of redistribution preferences measured via social surveys is that they rely on self-reported preferences by individuals. It is unclear whether these self-reported levels reflect "true" preferences or only reporting behavior. Individuals may wish to be perceived open-minded and caring for others, while they are in fact unwilling to share own resources with other individuals. A strategy to tackle this problem is to look at peoples' voting behavior (Luttmer and Singhal, 2011), because by making their electoral choice, individuals reveal their true preferences ("revealed preferences").

We exploit the WVS's information on voting behavior to construct two variables of party preference. The first variable measures respondent's self-positioning on a political scale running from 1 (very left-wing) to 10 (very right-wing). This variable reduces the reporting bias, but individuals may nonetheless respond that they are more left-wing (right-wing) than they actually are. The second variable, our preferred measure, uses information on the political party individuals would vote for in national elections. We focus on OECD countries, where suffrage is universal. The WVS lists 351 parties for the

**Table 4** STATUS MISPERCEPTION AND POLITICAL PREFERENCES

Dependent variables: Row (1): Left-wing party preferences
Row (2): Party individuals voted for is left-wing

| | (1) Perceptions | | (2) Misperceptions | | (3) Misperceptions and Controls | |
|---|---|---|---|---|---|---|
| | Coefficient | SE | Coefficient | SE | Coefficient | SE |
| **1. Self-reported political preferences** | | | | | | |
| Subjective$_i$ | -0.169$^{***}$ | (0.0496) | | | | |
| Mispercept$_i$ | | | -0.244$^{***}$ | (0.0442) | -0.233$^{***}$ | (0.0449) |
| SVMSES$^{\mathcal{R}}_{ij}$ | | | -0.111 | (0.0667) | -0.128$^{*}$ | (0.0649) |
| Observations | 58,501 | | 58,501 | | 58,468 | |
| Countries | 24 | | 24 | | 24 | |
| Country-fixed effects | Yes | | Yes | | Yes | |
| Wave-fixed effects | Yes | | Yes | | Yes | |
| R-squared | 0.031 | | 0.034 | | 0.042 | |
| **2. Ideology of parties individuals vote for** | | | | | | |
| Subjective$_i$ | -0.0422$^{***}$ | (0.0118) | | | | |
| Mispercept$_i$ | | | -0.0507$^{***}$ | (0.0090) | -0.0493$^{***}$ | (0.0090) |
| SVMSES$^{\mathcal{R}}_{i}$ | | | -0.0357$^{*}$ | (0.0173) | -0.0317$^{*}$ | (0.0156) |
| Observations | 28,252 | | 28,252 | | 28,232 | |
| Countries | 21 | | 21 | | 21 | |
| Country-fixed effects | Yes | | Yes | | Yes | |
| Wave-fixed effects | Yes | | Yes | | Yes | |
| R-squared | 0.151 | | 0.152 | | 0.159 | |

*Notes*: Robust standard errors adjusted for clustering by countries are in parentheses. The dependent variable are: self-reported party preference on a scale running from 1 (very right-wing) to 10 (very left-wing)(Row 1) and a dummy variable that is 1 (0 otherwise) if the political party respondents vote for has a left-wing ideology (Row 2). SVMSES$^{\mathcal{R}}_{ij}$ denotes the SES measured with the SVM-algorithm. All models include fixed effects for countries and waves. The model specification is identical to that of Equation (19) but replaces our baseline redistribution measure with our two measures of party preferences. Column (1) reports results for subjective perceptions, Column (2) presents estimates for misperceptions of social status, and Column (3) includes control variables. The set of control variables replicates that of our baseline table.

*** Significant at the 1 percent level,
** Significant at the 5 percent level,
* Significant at the 10 percent level

included OECD members states. We carefully analyze each of these parties and manually classify them as left-wing, center, or right-wing using information from the Database of Political Institutions 2017 (Scartascini et al., 2017) and (if parties are not included in the DPI) the official party programs. In appendix A.5, we provide a detailed description of our coding and show summary statistics. The relationship between right-wing and left-wing parties is balanced in our sample: 14,409 individuals vote for right-wing parties, 14,344 individuals vote for left-wing parties, while another 1,844 individuals vote for center parties without a clear ideology.

In Table (4), we report the results from empirical specifications identical to those of our baseline models (Table B-5), but where we replace redistribution preferences by our two measures of party preferences. The results strongly corroborate our baseline finding on redistribution preferences. The higher the perceived social status, the lower is the preference for left-wing parties. The effect of misperceptions is negative and statistically significant at the 1% level in each regression. The coefficient on the objective status is also negative, but it is weaker in terms of statistical significance than the parameter for misperceptions. This result suggests that a bias in perception is the relatively stronger negative correlate of left-wing preferences than the objective status.

## 5.9 Misperceptions and the tax and transfer scheme

By influencing individual preferences, biased perceptions of status may ultimately impact the tax and transfer systems chosen by governments. To examine the political consequences of misperceptions for fiscal policies, we construct a panel data set where we compute the average level of misperception for each country-year observation included in the WVS. The panel consists of 97 countries observed in five non-overlapping 5-year intervals between 1990 and 2014.[11]

The generosity of the tax and transfer system is measured via the pre-post approach, which gauges governmental intervention in the income distribution by assessing the difference of inequality before and after taxes and transfers (Lupu and Pontusson, 2011). This measure is computed based on Gini coefficients, i.e.

$$\mathrm{R}_{jt} = \mathrm{Gini(M)}_{jt} - \mathrm{Gini(N)}_{jt}, \tag{22}$$

where Gini(M) and Gini(N) are inequality of market and disposable incomes in country $j$ at time $t$. We use inequality data from the Standardized World Income Inequality Database (SWIID) compiled by Solt (2016). The SWIID provides income inequality series that maximize cross-country comparability for the broadest possible sample of observa-

[11] We assign each wave of the WVS to a 5-year period, as the waves span over multiple years: Wave 2 (1990-1994) is assigned to the 1990-1994 period, Wave 3 (1995-1998) to the period 1995-1999, Wave 4 (1999-2004) to the period 2000-2004, Wave 5 (2005-2009) to period 2005-2009, and Wave 6 (2010-2014) to period 2010-2014.

**Table 5** STATUS MISPERCEPTIONS AND THE NATIONAL TAX AND TRANSFER SYSTEM

Dependent variable: Effective redistribution (macro level), $R_{jt}$

| | Full sample | | OECD countries | | Non-OECD countries | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $\text{Mispercept}_{jt}$ | -0.00326 | -0.00319 | -0.0112** | -0.0100* | -0.00331 | -0.00326 |
| | (0.00230) | (0.00229) | (0.00504) | (0.00556) | (0.00205) | (0.00203) |
| Gini (market) | 0.193*** | 0.197*** | 0.415*** | 0.444*** | 0.0927 | 0.0956* |
| | (0.0530) | (0.0519) | (0.0928) | (0.0762) | (0.0571) | (0.0563) |
| Democracy | | 0.00488 | | 0.136*** | | 0.00243 |
| | | (0.00457) | | (0.0329) | | (0.00400) |
| Observations | 174 | 174 | 58 | 58 | 116 | 116 |
| Countries | 82 | 82 | 24 | 24 | 58 | 58 |
| R-Squared | 0.21 | 0.21 | 0.41 | 0.48 | 0.13 | 0.14 |
| F Stat | 7.620 | 5.682 | 11.68 | 15.06 | 2.400 | 1.654 |
| F p-val | 0.000 | 0.001 | 0.000 | 0.000 | 0.099 | 0.187 |

*Notes*: Robust standard errors adjusted for clustering by countries are in parentheses. The dependent variable is $R_{jt}$, which is a measure of effective redistribution, gauged via the difference between inequality before and after taxes and transfers. Gini (market) reflects the level of inequality of market incomes, and democracy is a measure of the quality of democratic institutions from Gründler and Krieger (2021b). All estimations include country-fixed effects. The estimations cover the period between 1990 and 2014 (in five non-overlapping 5-year intervals representing the waves of the WVS data), but data for most countries is available only for the post-1995 period. Availability of data for the covariates results in a reduction of $N$ from 97 to 82.

*** Significant at the 1 percent level,

** Significant at the 5 percent level,

* Significant at the 10 percent level

tions. The algorithm of the SWIID uses more than 10,000 data series on inequality and consolidates them into Gini indices that are comparable to the data from the Luxembourg income Study (LIS). While the LIS is considered the gold standard in inequality data, most of the countries surveyed in the WVS are not included. Appendix A.6 features a discussion of the pre-post approach and the (dis)advantages of the SWIID and its alternatives.

We estimate the effect of misperceptions on welfare provision based on the following econometric model

$$R_{jt} = \alpha + \gamma \text{Mispercept}_{jt} + \rho \text{Gini(M)}_{jt} + \phi \text{Democracy}_{jt} + \eta_j + \varepsilon_{jt}, \tag{23}$$

where $R_{jt}$ is our pre-post measure of redistribution and $\text{Mispercept}_{jt}$ is the standardized country mean of misperception. The specification controls for the economic model of self-interest in voting behavior by including market inequality. We also control for political institutions, which are a prerequisit for voters' preferences to channel to policy measures, by including the level of democratization (taken from Gründler and Krieger, 2021b). To account for unobserved cross-country heterogeneity, Equation (23) also includes country-fixed effects $\eta_j$.

Table (5) shows the results derived from all available observations and also presents estimates based on the sub-samples of OECD and non-OECD countries. The results show that upwards-biased perceptions are negatively related to income redistribution in all samples, but the correlation is statistically significant only for OECD countries. Consistent with the economic voting model, higher levels of market inequality are accompanied by more generous welfare states. We also observe that redistribution is higher in countries with a higher quality of democratic institutions. This result is in line with standard political economy model, which predicts that higher suffrage increases redistribution.

Our macroeconomic results provide suggestive evidence that misperceptions directly translate to welfare provision, at least in the democratized OECD countries. Erroneous beliefs about the social position are hence correlated with real-world consequences and provide an important step towards a better understanding of cross-country differences in income redistribution.

# 6 Conclusions

The willingness to redistribute resources from the rich to the poor is one of the defining characteristics of societies. Traditional economic theories assume that individuals form redistribution preferences by maximizing consumption under a set of constraints, and that the aggregation of individual preferences determines the equilibrium level of taxes and transfers. However, a burgeoning literature shows that individuals often have redistribution preferences that are only weakly correlated with the objective state of their pocketbook. We argue that an important step towards understanding this "redistribution puzzle" is to study whether individuals hold erroneous perceptions about their position in the social ladder and, consequently, misperceive their potential gains from redistribution.

Running a new machine learning algorithm on large-scale survey data for 241,757 households in 97 countries (24 OECD, 73 non-OECD), we uncover that status misperceptions are widespread on the globe. In OECD countries, about 70% of the households in the bottom-30% of the income distribution overestimate their individual status. Our empirical results show that individuals that overestimate their social status have significantly lower redistribution preferences and vice versa. The effects are statistically significant at the 1% level, independent of key socio-economic characteristics, and robust to potential measurement errors in the survey data.

While our results support prior experimental evidence for single countries, we find substantial cross-country differences in the extent to which status misperception translates to preference formation. This result suggests that unobserved country-level characteristics (institutions, culture, political history, geography) play an important role for the transmission of biased perceptions into preferences. However, the effects only differ in their size and not in their direction of influence. Upwards-biased perceptions are negatively related to the taste for redistribution in all countries, pointing to a general pattern in human behavior.

Our results complement the standard economic policy model that describes how redistribution preferences are driven by economic self-interest: if voters over-estimate their individual socio-economic position, they under-estimate their gains from income redistribution. Consequently, they may vote for less redistribution than would be economically rational, and a naive look at data on voting behavior may be "biased" in the sense that it does not directly reveal motives of economic self-interest behind voting. Our results also suggest that biased perceptions translate to policy outcomes.

The origins of status misperception are still poorly understood, and their investigation poses a promising avenue for future research. While some authors hypothesize a major role of limited information and limited cognitive ability (Cruces et al., 2013), other factors such as media consumption, the social environment, and cultural influences may as well be decisive for the formation of biased perceptions.

# A. Supplementary notes

## A.1 Preferences for redistribution with relative consumption concerns

When relative consumption concerns play a role for the formation of preferences (see Section 2.2), the utility function in our simple model adjusts to

$$U_i = U(u_1(c_i), u_2(c_i|\bar{c})), \tag{24}$$

including the direct utility from consumption and the indirect utility of consumption relative to the rest of society $u_2(c_i|\bar{c})$.

A simple way to represent disutility from consumption of others, $u_2$, is modeling

$$u_2 = -n_i(1-t)(\bar{\phi} - \phi), \tag{25}$$

where $n_i \in (0, 1)$ is a preference parameter that denotes individual $i$'s disutility from consumption of other members of society. The larger the distance between the average productivity and the own productivity level, the higher is the disutility from relative consumption. Disutility from relative consumption, however, is mitigated by higher tax payments of the average member of society relative to $i$. This mechanisms is captured by the term $(1-t)$.

In the augmented model, the utility function to maximize becomes

$$u_i = (1-t)\phi_i + \bar{\phi}t - wt^2 - n_i(1-t)(\bar{\phi} - \phi) \tag{26}$$

with first-order condition

$$0 = \frac{\partial u_i}{\partial t} = \phi_i - \bar{\phi} - 2wt + n(\bar{\phi} - \phi). \tag{27}$$

Re-arranging gives the preferred tax rate

$$t_i = \frac{(\bar{\phi} - \phi_i) + n(\bar{\phi} - \phi_i)}{2w}. \tag{28}$$

Equation (28) shows that for $n > 0$, the tax rate increases when the productivity level of $i$ falls short of the productivity level of the peer group.

### A.2.1 Support Vector Regression and its use to classify social status

Our approach to measure the SES uses supervised machine learning techniques to find patterns in the data. Specifically, we use Support Vector Machines (SVMs) to accomplish our classification task. In this appendix, we briefly describe the principles of SVM regressions. A detailed description can be found in Vapnik (1995, 1998), Steinwart and Christmann (2008), and Schölkopf and Smola (2001).

**A brief introduction into Support Vector Regression:** Support Vector Machines are designed to find an unknown funtional relationship $\mathfrak{F} : \mathcal{X} \to \mathcal{Z}$ that links input variables $\mathbf{x} = (x_1, \ldots, x_m)' \in \mathcal{X} \subseteq \mathbb{R}^m$ to realizations $z \in \mathcal{Z} \subseteq \mathbb{R}$ using observations $i$ of a sample of data $\mathcal{S} = \{(\mathbf{x}_i, z_i)|i = 1, \ldots, n\}$, i.e.

$$\mathfrak{F}(\mathbf{x}_i) \stackrel{!}{=} z_i, \ \forall i = 1, \ldots, n.$$

The main advantage of this technique compared with traditional tools of statistical modelling is that SVMs can be trained to find relationships in the data without requiring prior assumptions or explicit programs (Breiman, 2001).

Support Vector Machines are supervised learning techniques, meaning that they use a set of labeled observations to compute a function that is able to predict the output for all observations. In contrast, unsupervised learning techniques aim at identifying clusters of observations that are similar in terms of their covariates. Initially, SVMs have been used for classification problems where the output variable comes from a countably finite set, providing non-linear extensions of the General Portrait Algorithm (GPA). Vapnik (1995, 1998) augments these techniques to be feasible for estimation of real-valued functions. These "Support Vector Regressions" are constructed to find a function

$$\mathfrak{F} : \mathcal{X} \subseteq \mathbb{R}^m \to z \in \mathcal{Z} \subseteq \mathbb{R}$$

so that the deviation of the predicted outcomes from the observed labels does not deviate by more than a specified level $\varepsilon$:

$$|\mathfrak{F}(\mathbf{x}_i) - z_i| \stackrel{!}{\leq} \varepsilon, \ \forall i = 1, \ldots, n.$$

In the linear case when the regression function is a hyperplane, we have

$$\mathfrak{F}(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b, \ \mathbf{x} \in \mathbb{R}^m, b \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^m$$

and our primary goal is to minimize the slope of the norm $\mathbf{w}$. This means solving the quadratic optimization problem

$$\min_{\mathbf{w}\in\mathbb{R}^m, b\in\mathbb{R}} \frac{1}{2}\|w\|^2 \text{ s.t. } \begin{cases} z_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b \leq \varepsilon & \forall i \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b - z_i \leq \varepsilon & \forall i. \end{cases} \tag{29}$$

The regression line can then be computed using the solution $(\mathbf{w}^*, b^*)$ of the optimization. For many applications, it is impossible to solve this optimization problem due to its constraints, and researchers often want to model non-linear relationships. Vapnik (1995, 1998) meets both challenges by (i) including slack variables $(\zeta_i^+, \zeta_i^-) \in \mathbb{R}_+^2$ $(i = 1, \ldots, n)$ to relax the auxiliary conditions and (ii) using the method of Boser et al. (1992) that extends the GPA to enable the estimation of non-linear classification functions. The idea of this approach is to use a non-linear function $\Phi : \mathcal{X} \to \mathcal{H}$ that maps the input characteristics into a higher-dimensional space, called the "Reproducing Hilbert Space" denoted with $\mathcal{H}$. The logic behind this data transformation is that the optimization problem can be transferred to the adjusted sample $\mathcal{S}_{\mathcal{H}} = \{(\Phi(\mathbf{x}_i, z_i)|i = 1, \ldots, n)\}$, and be solved in $\mathcal{H}$.

The problem of Equation (29) then adjusts to

$$\min_{\mathbf{w}\in\mathcal{H}, b\in\mathbb{R}, (\zeta_i^+, \zeta_i^-)\in\mathbb{R}_+^2} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\zeta_i^+, \zeta_i^-) \text{ s.t. } \begin{cases} z_i - \langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle - b \leq \varepsilon + \zeta_i^+ \\ \langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle + b - z_i \leq \varepsilon + \zeta_i^- \\ \zeta_i^+, \zeta_i^- \geq 0. \end{cases} \tag{30}$$

This problem can easily become computationally infeasible when the dimensionality of $\mathcal{H}$ is large. This problem can be circumvented by using the corresponding dual program of Equation (30)

$$\max_{\alpha^+, \alpha^-} -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-)\langle \Phi(\mathbf{x}_i)\Phi(\mathbf{x}_j)\rangle_{\mathcal{H}} - \varepsilon\sum_{i=1}^{n}(\alpha_i^+ - \alpha_i^-) + \sum_{i=1}^{n} y_i(\alpha_i^+ - \alpha_i^-)$$
$$\text{s.t.} \sum_{i=1}^{n}(\alpha_i^+ - \alpha_i^-) = 0 \text{ and } \alpha_i^+, \alpha_i^- \in [0, C], \tag{31}$$

where $\alpha^+ = (\alpha_1^+, \ldots, \alpha_n^+)$ and $\alpha^- = (\alpha_1^-, \ldots, \alpha_n^+)$ are the Lagrangian multipliers of the primal program. The closed form solution of Equation (31) is

$$\mathfrak{F}(\mathbf{x}) = \sum_{i=1}^{n}(\alpha_i^+ - \alpha_i^-)\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x})\rangle_{\mathcal{H}} + b_{\mathcal{H}}^*. \tag{32}$$

The main challenge to obtain Equation (32) is to find $\Phi : \mathcal{X} \to \mathcal{H}$, which is unknown. This challenge is often tackled by using the "kernel trick". This method replaces the unknown inner product $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x})\rangle_{\mathcal{H}}$ with a kernel $\mathfrak{k} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.

The non-linear regression function then becomes

$$\mathfrak{F}(\mathbf{x}) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \mathfrak{k}(\mathbf{x}_i, \mathbf{x}) + b_{\mathcal{H}}^*. \tag{33}$$

The shape of the function in Equation (33) depends only on those observations where the Lagrangian multipliers $(\alpha_i, \alpha_i^*)$ are non-zero. These observations are called "Support Vectors" and coin the name of the approach.

**Notes on the socio-economic characteristics underlying social status:** The SVM approach requires labels for individuals with (very) high and low social status to select a set of training data and uses socio-economic characteristics of individuals to find the optimal hyperplane that divides the training data. Our list of socio-economic characteristics follows the definition of the APA and is shown in Table (B-1). The selection of variables refers to (i) accurate measurement of our definition of status and (ii) availability of data in the WVS. In principle, the WVS includes a number of additional variables that potentially influence the social status. These variables, however, are available only for a (small) subset of households. As Support-Vector techniques require balanced panels, we cannot include this additional information in our analysis without a disproportionately large loss of observations. Education and incomes are directly measured in the WVS, occupational status is gauged with the help of five different proxies, including full-time employment, unemployment, self-employment, work as manager or employer, and enrollment as full-time student. We also include two additional variables that are important for occupational success: (i) age measured in years and (ii) the strength of friendship ties to account for social capital (Chan and Goldthorpe, 2007; Ellison et al., 2007).

A concern about our variable selection is that the characteristics may be correlated. An advantage of SVMs over traditional methods is that including correlated or irrelevant variables (in the sense that they do not provide any additional information not captured in the remaining variables) does not influence the classification. In SVM applications, the marginal effect of the variable in the classification function simply approximates zero.

**Details on the selection of labels ad the training data:** We use the upper and the lower decile of the distribution of $p_{ij}$ as training data for country $j$. To translate the information into a form that can be processed by the learning machine, we assign the label 1 to the top decile and the label 0 to the bottom decile. The rationale is that individuals with very low (high) levels of income and education possess the lowest (highest) possible levels of status. Conventional aggregation techniques produce similar classifications at the extreme ends of the distribution, and there is little dispute about who belongs to the "high society" and who not. Classifying individuals between the very top and the very bottom, however, is a much more challenging task, which for reasons described in Section

(3.2) cannot be adequately fulfilled by Equation (16). We use income and education, as both variables are only moderately correlated (31.44%, $N = 277,393$). The multiplicative combination ensures that high levels of one variable cannot compensate for low levels of the other dimension and, in particular, that $p_{ij}$ becomes zero if one factor is zero. The selection of training data is decisive for the classification procedure. The SVM algorithm thus includes several control mechanisms to prevent misclassification due to mislabeling.

**Preventing missclassification caused by mislabeled training data:** The accuracy of the classification hinges on the quality of the priming data. Four methodological features are implemented to prevent mislabeled training data. First, the distribution $\Phi_{S^{\mathfrak{S}}}$ accounts for measurement uncertainty in the data. Uncertainty may arise in the WVS survey if data for the characteristics $\mathbf{x} = (x_1, \ldots, x_m)$ is prone to measurement errors. This is a general problem of social surveys and may hence also affect the training data. Second, we use $\mathfrak{F}^{\mathfrak{S}}_{j_{\mathcal{T}_{\zeta,j}}}$ to classify *all* observations, including the initially labeled training data ("priming data"). Our robustness checks show that this procedure eliminates potentially mislabeled training data. Third, we draw subsamples of all training data in each iteration $\zeta$ that are random in their composition and their size. Fourth, we also run robustness checks for several regularization parameters $C$, which balances between the pros and cons of a smaller and a larger margin hyperplane. A small value of $C$ causes the optimizer to look for a larger-margin separating hyperplane and vice versa.

**Technical details about the specification of the SVM algorithm:** Our Support Vector Machine is implemented with regulaization parameter $C = 1$, as recommende by Mattera and Haykin (1999). In addition, we follow Gründler and Krieger (2021b) in calibrating the margin parameter $\varepsilon = 0.05$ and using the Gaussian RBF to transfer the data into the higher-dimensional feature space.

### A.2.2 Wisdom-of-the-crow approach using Twitter data to assess the labeling rule of dimensions for the SVM algorithm

An important step of our SVM-based approach to combine socio-economic characteristics into an index of social status is to label those individuals in the dataset that can be classified as "very low status" and "very high status" with sufficient probability (the labels $z_i \in \mathcal{Z} \subseteq \mathbb{R}$ discussed in appendix A.1.1). We follow the approach of Falk et al. (2020) using a combination of education and income to accomplish this task. We believe that this approach is reasonable, as it mitigates problems related to the observation that individuals with relatively high status may have high educations levels but do not necessarily rank at the top of income distribution (a prime example would be a university professor) and vice versa (an illustrative example for the opposite case would be, for instance, a sport professional). People with neither education or income, however, can reasonably be thought of as having low social status, while there is good reason to conjecture that individuals with high education and high income are at the top of the status distribution.

To verify our method, we use a wisdom-of-the-crow (WotC) approach using Twitter data. WotC approaches aggregate opinions from a large sample of individuals to answer a question. The underlying assumption (empirically validated in a number of studies conducted particularly in computer science) is that swarm intelligence outperforms the assessment of single individuals, even experts. The statistical advantage is that by collecting opinions from a group of individuals eliminates outliers. A special advantage in our setting is that it reflects the collective opinion of society about social status.

We collect 2,519 Twitter Tweets that included the term "social status" and preprocess the included text in three steps: (i) We transform the text to only include lowercase words, (ii) filter stop words, and (iii) tokenize the raw input. This procedure leaves us with 11,435 distinct words. We compute frequency analyses and illustrate the result of the frequency analysis in a word cloud, shown in Figure (8).

The figure shows that individuals perceive the two dimension that we use to identify our labels (income and education, marked in red) to be key defining features of social status. This result provides strong support for the labeling rule used in our SVM model.

**Figure 8** WISDOM-OF-THE-CROWD APPROACH TO VALIDATE LABELS


*Notes*: Wisdom-of-the-crow (WotC) approach using Twitter data to validate the labeling rule of the SVM algorithm. The figure shows a word cloud of words mentioned in Twitter Tweets on the term "social status". The size of the words represents the frequency a particular word has been mentioned together in a Tweet with the term "social status". The graph is based on 11,435 distinct words included in 2,519 Tweets. The data is pre-processed to exclude stop words. Words referring to our labeling rule are marked in red color.

## A.3 Comparison of the SVM-classification of social status and traditional techniques

We compare the classification outcome of our SVM-based procedure with traditional aggregation techniques employed to classify the social status of individuals. To this end, we analyze the correlation of the SVMSES with the most commonly used aggregation techniques described in Section (3.2), namely additive and multiplicative aggregation schemes, as well as principal component analyses (PCA). The computation of the corresponding indices is based on the identical conceptualization and operationalization as used for the SVMSES. The only difference between the indices is the rule applied to aggregate the raw data into an index. To make the computation of the additive and the multiplicative index feasible, the dummy variables for student status and unemployment are re-coded so that 1 refers to "not student" and "not unemployed". This re-coding ensures that there is a negative relationship between these variables and the social status. The multiplicative technique also uses a margin to prevent the index from becoming 0 in cases when one or

**Table 6** CORRELATIONS BETWEEN SVMSES AND TRADITIONAL AGGREGATION TECHNIQUES

| | **SVMSES** | **Additive** | **Multiplicative** | **PCA** |
|---|---|---|---|---|
| SVMSES | 1.0000 | -0.0247 | 0.5683 | 0.7781 |
| Additive | -0.0247 | 1.0000 | 0.3362 | -0.2856 |
| Multiplicative | 0.5683 | 0.3362 | 1.0000 | 0.5074 |
| PCA | 0.7781 | -0.2856 | 0.5074 | 1.0000 |

*Notes*: Table displays the correlation of the SVMSES and traditional techniques applied to aggregate raw data into composite measures, inclulding the additive approach (with $\omega_1, \ldots, \omega_m = 1$), the multiplicative approach (with $\omega_1, \ldots, \omega_m = 1$), and principal component analyses (PCA). These techniques are described in Section (3.2). The raw data is identical for all indices.

more dummy variables are 0. As a methodological side note, the outcome of the SVMSES would *not* be affected by such adjustments, i.e. the SVMSES would assume the identical value when applied to the re-coded dummy variables, as the underlying information in the data remains unchanged.

The aggregation procedures of the traditional techniques are applied as described in Section (3.2). We employ the traditional techniques based on $\omega_1, \ldots, \omega_m = 1/m$, which is common in the social science literature. Moreover, we have no theoretical guideline for a different setting of the weight scheme.

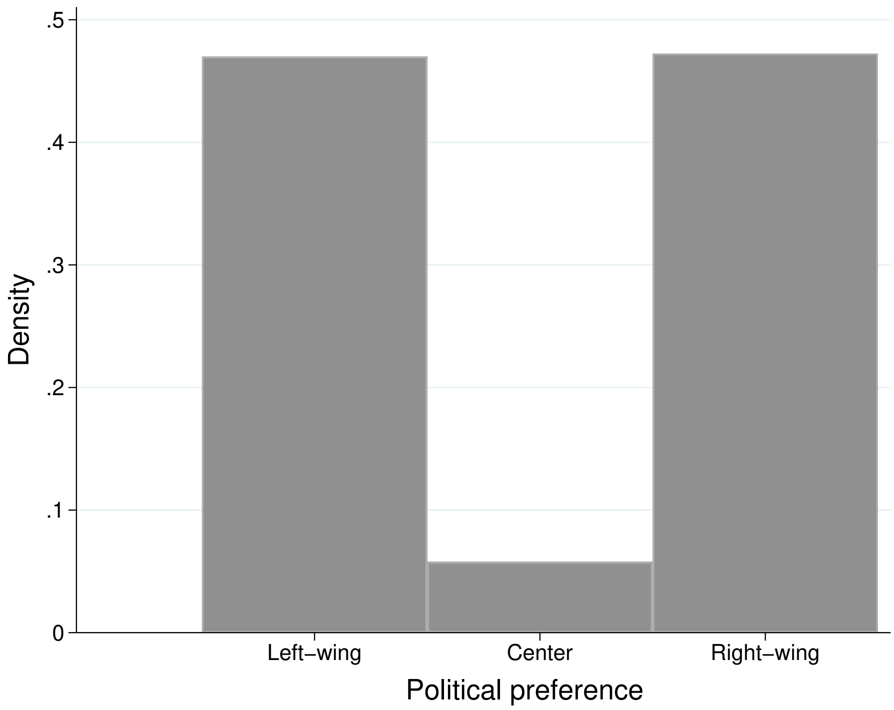
Table (6) reveals important differences between the SVM algorithm and traditional aggregation schemes. We observe little correlation between the SVMSES and the additive aggregation scheme. The correlation is moderate (0.56) with respect to the multiplicative approach and stronger (0.77) with respect to the PCA-based index. Also, the PCA-index is related much more to the SVMSES than to the additive or the multiplicative index. These patterns suggest that a substantial part of the differences is due to the weighting scheme: while the additive and the multiplicative index by construction require ex ante assumptions about the weighting scheme, PCA and the SVM-algorithm are flexible in the sense that they recover $\omega_1, \ldots, \omega_m$ from the data. However, the differences between the PCA index and the SVMSES are still substantial and result from the (demanding) linearity assumption of the PCA.

## A.4 Notes on the measurements of redistribution preferences

We compute redistribution preferences $\text{Pref}_{ij}$ based on question E035 of the WVS, which asks respondents to place their view about redistribution on a scale running from 1 to 10. A value of 1 means complete agreement with the statement "incomes should be made more equal", while a number of 10 refers to "we need larger income differences as incentives". The exact wording of question E035 in the WVS is: *"Now I'd like you to tell me your views on various issues. How would you place your views on this scale? 1 means you agree completely with the statement on the left; 10 means you agree completely with the statement on the right; and if your views fall somewhere in between, you can choose any number in between. (Code one number for each issue): Incomes should be made more equal vs. We need larger income differences as incentives."* We recode this question so that larger number reflect greater preferences for redistribution.

Redistribution preferences via question E035 can be computed for 97 countries, 25 of them from the OECD ($N = 264,528$). This preference variable is available for a larger set of countries than our status variable, as some of the status characteristics have not been asked in each country. For example, for the OECD, status variables are missing for France. At the time of this study, the WVS includes 6 waves, which were conducted at different time periods: Wave 1 (1981–1984), Wave 2 (1990–1994), Wave 3 (1995–1998), Wave 4 (1999-2004), Wave 5 (2005–2009), and Wave 6 (2010–2014). Each wave includes information on preferences for redistribution, but demographic factors were only collected since Wave 2.

**Figure 9** PARTY PREFERENCES OF INDIVIDUALS IN THE OECD




*Notes*: Distribution of party preferences among individuals living in OECD countries. The figure shows the political ideology of parties for which individuals would vote for in the next election. Data refers to question E179 of the WVS.

## A.5 Notes on the coding of party preference and summary statistics

To code revealed political preferences of individuals, we use data from question E179 of the WVS, which asks respondents about the political party they will vote in the next election. We use information from the Database of Political Institutions to code their ideology on a three-level scale, including 1 (left-wing ideology), 2 (center party), and 3 (right-wing ideology). For some parties (about 20% of all parties), there is no information available in the DPI. This mainly affects small parties with few voters. In this case, we use the official party programs to code their ideology.

We then construct a dummy variable for left-wing party preference, which is 1 if the party has left-wing ideology, and zero otherwise. Figure (9) shows the distribution of party preference among households in the OECD. The figure shows that the distribution between left-wing and right-wing parties is balanced.

There is strong heterogeneity in party preferences across countries. While left-wing parties are popular among voters in Sweden (86.8%) and Canada (67.2%), they receive less support in the Czech Republic (25.,%), Chile (27.8%), and Slovenia (31.0%).

## A.6 Notes on the pre-post measure of redistribution on the macroeconomic level

To measure the generosity of the welfare state, we use the "pre-post-approach", which gauges governmental intervention in the income distribution via the difference of inequality before and after taxes and transfers (Lupu and Pontusson, 2011) based on Gini coefficients, i.e.

$$\mathrm{R}_{jt} = \mathrm{Gini(M)}_{jt} - \mathrm{Gini(N)}_{jt} \tag{34}$$

where Gini(M) and Gini(N) denote market and net Ginis in country $j = 1, \ldots, N$ at time $t = 1, \ldots, T$. The main challenge to compute Equation (34) is to acquire comparable data on inequality pre and post taxes and transfers that is based on the same uniform set of definitions and assumptions. The problem is that conceptualizations of inequality series often vary substantially across countries, which results in incomparability problems when using secondary datasets (Atkinson and Brandolini, 2001). Based on a harmonized method that is comparable across nations, the LIS Cross-National Data Center provides micro data of unparalleled comparability and quality, but it only includes 232 observations from 41 countries. Seven of these countries are only included in a single period. This limitation is particularly serious for our analysis, as the included countries are biased towards OECD countries, and inequality series for most of the countries in our large sample are not available.

Including many country-years, however, comes at the cost of sacrificing the benefits of harmonization, which imposes an inevitable trade-off between coverage and comparability. Two datasets have been particularly sucessful in tackling this trade-off, namely the "World Income Inequality Database" (WIID) of UNU-WIDER and the "Standardized World Income Inequality Database" (SWIID) compiled by Solt (2016). Both datasets are regularly updated and include a broad sample of country-years based on the highest possible degree of comparability. For our estimation, we concentrate on data from the SWIID. The reason is that the scope of included country-years for which data on incomes pre and post taxed and transfer is availble is considerably smaller compared to the SWIID. This is especially the case for developing economies, where only a few country-year observations include market *and* net Ginis.[12] The SWIID further provides a sub-set of country-years with superior data quality that only includes observations for which micro data exist. This subsample includes 2,030 country-years. All country-year observation in our OECD sample is entirely based on micro data.

[12]The WIID also features substantial differences in the definitions of taxable income and the tax unit as well as the degree of evasion and tax avoidance across incomes. Thus, the WIID is suitable when comparing trends over time across countries, but not levels. For a detailed discussion of this argument, see Solt (2016).

# B. Supplementary Tables

**Table B-1** CHARACTERISTICS OF THE WORLD VALUE SURVEY (WVS) USED FOR THE OPERATIONALIZATION OF THE SOCIO-ECONOMIC STATUS

| **Characteristic** | **Description** | **WVS** |
|---|---|---|
| Education | Highest level of education attained on an internationally comparable scale running from 1 to 8. | X025 |
| Income | Income level of the household on an internationally comparable scale running from 1 to 10. | X047 |
| Age | Age of the respondents in years at the time the questionnaire was answered | X003 |
| Employment status | Indicates whether a person is employed in a full-time position | X028 |
| Unemployment | Dummy variable that indicates whether an individual is unemployed | X028 |
| Employer | Dummy variable that indicates whether a person is owner or manager of a firm and employs others | X036 |
| Student | Dummy variable that indicates whether a person is enrolled as a full-time student | X028 |
| Self-employed | Dummy variable that indicates whether a person in self-employed | X028 |
| Friendship | Measures the agreement to the question *"how important are friends in your life"* on a scale running from 1 to 4 | A002 |

*Notes*: The table lists and describes the variables used to compute the social status of individuals. The column labeled "WVS" displays the number of the question in the longitudinal database of the WVS, which includes all data from Waves 1–6. Note that the numbering scheme differs across waves. The 6 waves have been conducted in: Wave 1 (1981–1984), Wave 2 (1990–1994), Wave 3 (1995–1998), Wave 4 (1999-2004), Wave 5 (2005–2009), and Wave 6 (2010–2014). Due to coverage of the questionnaires, we use data from Waves 2–6.

**Table B-2** MISPERCEPTIONS IN THE OECD, SUMMARY STATISTICS

| **Country** | **Mean** | **Std. Dev.** | **Observations** | **Minimum** | **Maximum** |
|---|---|---|---|---|---|
| Australia | 0.055 | 1.049 | 3,950 | -2.885 | 3.484 |
| Canada | 0.172 | 1.047 | 3,398 | -2.895 | 3.619 |
| Chile | -0.046 | 0.878 | 3,869 | -3.613 | 2.833 |
| Czech Republic | -0.450 | 0.883 | 894 | -2.732 | 2.221 |
| Estonia | -0.286 | 0.934 | 2,408 | -3.521 | 2.945 |
| Finland | -0.012 | 1.023 | 1,703 | -3.026 | 3.829 |
| Germany | 0.393 | 0.910 | 2,501 | -2.940 | 3.711 |
| Hungary | -0.387 | 0.779 | 956 | -2.754 | 1.798 |
| Italy | -0.354 | 0.886 | 647 | -2.910 | 2.273 |
| Japan | -0.021 | 1.042 | 3,856 | -3.913 | 3.753 |
| Latvia | -0.344 | 0.916 | 1,130 | -3.518 | 2.340 |
| Lithuania | -0.448 | 0.867 | 891 | -3.192 | 2.102 |
| Mexico | -0.036 | 1.052 | 5,046 | -3.833 | 3.892 |
| New Zealand | 0.164 | 1.081 | 2,307 | -2.862 | 3.544 |
| Norway | 0.093 | 1.051 | 1,842 | -2.868 | 3.860 |
| Poland | -0.453 | 0.939 | 2,701 | -3.619 | 3.853 |
| Slovakia | -0.439 | 0.913 | 895 | -2.643 | 1.909 |
| Slovenia | -0.048 | 0.839 | 1,936 | -2.839 | 3.409 |
| South Korea | -0.099 | 0.943 | 3,604 | -3.806 | 3.768 |
| Spain | -0.128 | 1.008 | 3,808 | -3.256 | 3.199 |
| Sweden | 0.316 | 1.011 | 2,738 | -3.158 | 3.708 |
| Switzerland | 0.343 | 0.894 | 1,996 | -2.336 | 3.708 |
| Turkey | 0.024 | 1.216 | 7,463 | -3.539 | 3.873 |
| United States | 0.320 | 0.991 | 5,728 | -2.990 | 3.912 |

*Notes*: Table displays the mean of $\text{Mispercept}_{ij}$, along with the standard deviation (Std. Dev.), the number of households included in the WVS for which misperceptions can be calculated (Observations), the minimum (Minimum) and the maximum (Maximum) level of misperception in the corresponding country.

**Table B-3** SELF-POSITIONING BIAS BY INCOME DECILES IN GERMANY AND THE UNITED STATES

| | Germany | | United States | |
|---|---|---|---|---|
| Objective decile | Mean bias | Proportion with positive bias | Mean bias | Proportion with positive bias |
| 1 | 0.879 | 0.723 | 0.780 | 0.581 |
| 2 | 1.290 | 0.829 | 1.174 | 0.807 |
| 3 | 1.256 | 0.862 | 1.222 | 0.882 |
| 4 | 1.300 | 0.810 | 1.340 | 0.783 |
| 5 | 0.953 | 0.783 | 1.044 | 0.766 |
| 6 | 0.321 | 0.658 | 0.758 | 0.638 |
| 7 | -0.520 | 0.415 | -0.351 | 0.467 |
| 8 | -1.122 | 0.374 | -2.077 | 0.307 |
| 9 | -1.281 | 0.386 | -2.712 | 0.267 |
| 10 | -1.400 | 0.323 | -1.442 | 0.319 |

*Notes*: 'Mean Bias' reports the average level of $\text{Mispercept}_{ij}$ for different income deciles. 'Proportion with positive Bias' reports the fraction of repondents with $\text{Mispercept}_{ij} > 0$.

**Table B-4** REDISTRIBUTION PREFERENCES—SOCIAL STATUS VERSUS INCOME

| Dependent variable: Subjective preferences for income redistribution, $\text{Pref}_{ij}$ | | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Income | -0.157*** (0.0026) | | -0.0688*** (0.00412) | -0.0678*** (0.00420) |
| Social Status ($\text{SVMSES}^{\mathcal{R}}_{ij}$) | | -0.355*** (0.0053) | -0.240*** (0.0086) | -0.235*** (0.0090) |
| Observations | 264,528 | 264,528 | 264,528 | 264,388 |
| Countries | 97 | 97 | 97 | 97 |
| R-Squared | 0.092 | 0.094 | 0.095 | 0.096 |
| F Stat | 1223.2 | 1354.7 | 1170.6 | 518.1 |
| Country-fixed effects | Yes | Yes | Yes | Yes |
| Wave-fixed effects | Yes | Yes | Yes | Yes |
| Controls | No | No | No | Yes |

*Notes*: Robust standard errors adjusted for clustering by countries are in parentheses. The dependent variable is $\text{Pref}_{ij}$, which gives the subjective preference for income redistribution on a scale running from 1 (low) to 10 (high). $\text{SVMSES}^{\mathcal{R}}_{ij}$ denotes the social status measured by the SVM-algorithm. All models include fixed effects for countries and waves inclkuded in the WVS. The list of control variables includes age (measured in years), sex, and dummy variables for marital status, retirement and dummy variables indicating whether there are children living in the household and whether individuals are divorced or widowed. We also include a dummy variable indicating whether respondents are chief wage earners.

*** Significant at the 1 percent level,
** Significant at the 5 percent level,
* Significant at the 10 percent level

**Table B-5** STATUS MISPERCEPTION AND REDISTRIBUTION PREFERENCES—BASELINE ESTIMATES, FULL SAMPLE OF 97 COUNTRIES

Dependent variable: Subjective preferences for income redistribution, $\text{Pref}_{ij}$

| | (1) Perceptions | | (2) Misperceptions | | (3) Misperceptions and Controls | |
|---|---|---|---|---|---|---|
| | Coefficient | SE | Coefficient | SE | Coefficient | SE |
| $\text{Subjective}_{ij}$ | -0.278*** | (0.0224) | | | | |
| $\text{Mispercept}_{ij}$ | | | -0.145*** | (0.0248) | -0.149*** | (0.0252) |
| $\text{SVMSES}^{\mathcal{R}}_{ij}$ | | | -0.409*** | (0.0278) | -0.402*** | (0.0284) |
| Age | | | | | 0.002** | (0.0009) |
| Female | | | | | 0.102*** | (0.0214) |
| Divorced | | | | | 0.073 | (0.0494) |
| Married | | | | | -0.022 | (0.0494) |
| Widowed | | | | | 0.109** | (0.0493) |
| Child in household | | | | | -0.101*** | (0.0352) |
| Chief wage earner | | | | | -0.062 | (0.0467) |
| Retired | | | | | 0.089** | (0.0391) |
| Observations | 241,757 | | 241,757 | | 241,622 | |
| Countries | 97 | | 97 | | 97 | |
| Country-fixed effects | Yes | | Yes | | Yes | |
| Wave-fixed effects | Yes | | Yes | | Yes | |
| R-squared | 0.093 | | 0.099 | | 0.100 | |
| F-Stat | 272.4*** | | 293.4*** | | 273.6*** | |

*Notes*: Robust standard errors adjusted for clustering by countries are in parentheses. The dependent variable is $\text{Pref}_{ij}$, which gives the subjective preference for income redistribution on a scale running from 1 (low) to 10 (high). $\text{SVMSES}^{\mathcal{R}}_{ij}$ denotes the SES measured with the SVM-algorithm. In order to rule out issues with multi-collinearity, the table does not include further measures of socio-economic status, such as income, education, and occupational status. "Age" gives the age of respondents in years, "Female" is a dummy variable that equals 1 if the respondent is female, "Divorced", "Married", "Widowed", and "Retired" are dummy variables that control for the corresponding status, "Child in household" equals 1 if a child lives in the household of the respondent, and "Chief wage earner" denotes whether the respondent is the chief wage earner of the household. The data refers to questions X001, X011, X007, X041, C006, and S002 of the consolidated dataset of the WVS and are re-coded accordingly. All models include fixed effects for countries and waves.

*** Significant at the 1 percent level,

** Significant at the 5 percent level,

* Significant at the 10 percent level

**Table B-6** STATUS MISPERCEPTION AND REDISTRIBUTION PREFERENCES—ACCOUNTING FOR MEASUREMENT UNCERTAINTY, MI-ESTIMATES ON THE INDIVIDUAL-LEVEL DISTRIBUTION OF MISPERCEPTIONS

| Dependent variable: Subjective preferences for income redistribution, $\text{Pref}_{ij}$ | | | | |
|---|---|---|---|---|
| Specification | Coefficient of Misperception | SE | $M$ | $N$ |
| 1. Baseline (OECD) | -0.251*** | (0.0293) | – | 65,345 |
| 2. Baseline (OECD), MI-estimates | -0.252*** | (0.0292) | 200 | 65,345 |
| 3. Reduced model (OECD), MI-estimates | -0.297*** | (0.0329) | 200 | 65,809 |
| 4. Baseline (all) | -0.126*** | (0.0241) | – | 240,278 |
| 5. Baseline (all), MI-estimates | -0.128*** | (0.0240) | 200 | 240,278 |
| 6. Reduced (all), MI-estimates | -0.158*** | (0.0241) | 200 | 241,757 |

*Notes*: Robust standard errors adjusted for clustering by countries are in parentheses. The dependent variable is $\text{Pref}_{ij}$, which describes the subjective preference for income redistribution on a scale running from 1 (low) to 10 (high). All specifications include country-fixed effects. The specifications entitled "1. Baseline (OECD)" and "3. Baseline (all)" refer to the fully specified models reported in Tables (1)and (B-5). Specifications 2. and 5. estimate the same models using multiple-estimations based on the 200 SVM-iterations of $\text{Mispercept}_{ij}$ and $\text{SVMSES}^{\mathcal{R}}_{ij}$ that constitute their empirical distribution. Each specification shows the results of 200 estimations that are combined following the MI rules introduced by Rubin (1987). Specifications 3. and 6. show the effect of measurement uncertainty in the reduced specifications. The baseline variant is the fully specified model in Table (B-5).

*** Significant at the 1 percent level,
** Significant at the 5 percent level,
* Significant at the 10 percent level

**Table B-7** STATUS MISPERCEPTION AND REDISTRIBUTION PREFERENCES—ESTIMATES USING HOUSEHOLD-LEVEL DATA FROM THE LIFE IN TRANSITION SURVEY (LITS)

Dependent variable: Subjective preferences for income redistribution, $\text{Pref}_{ij}$

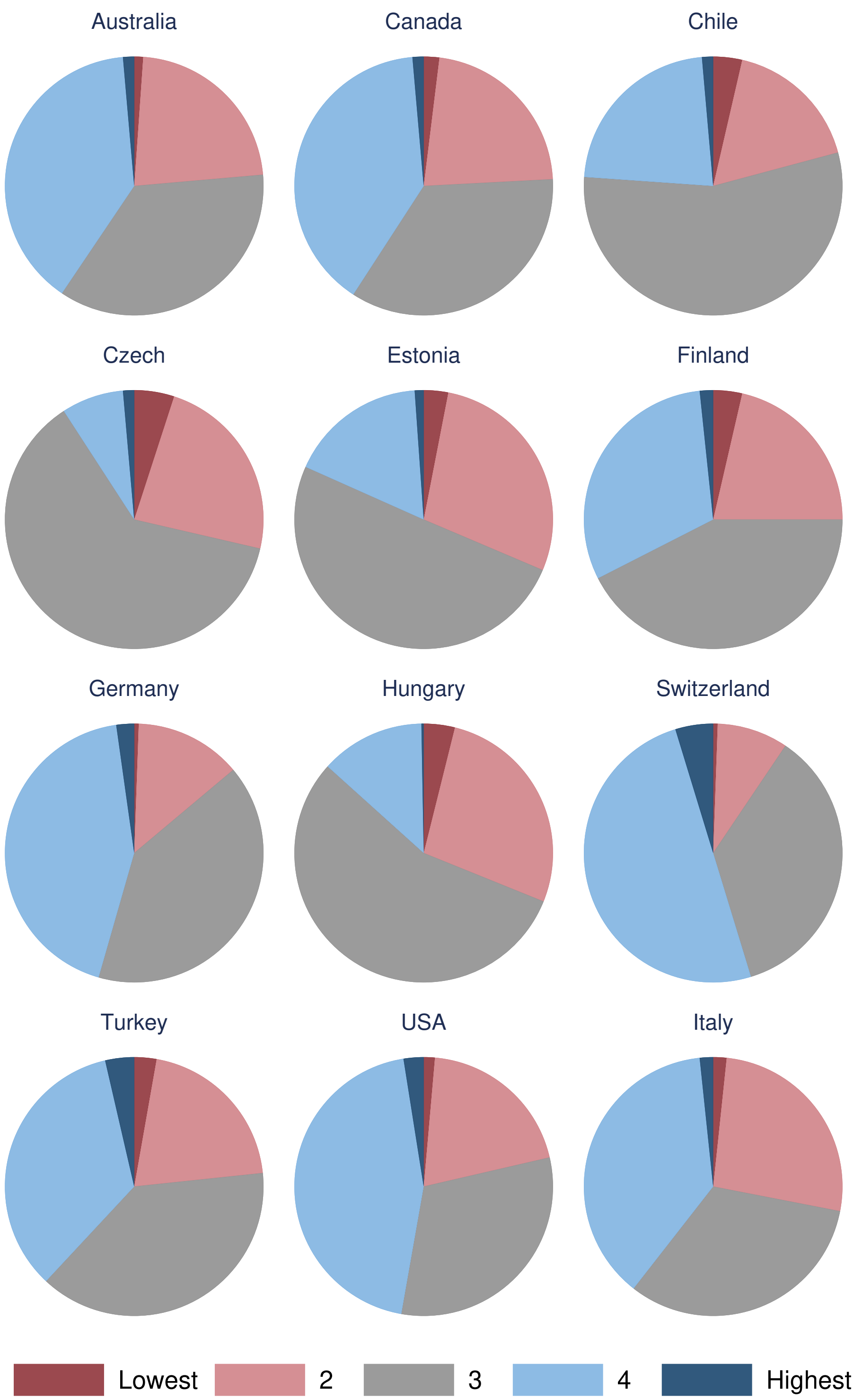
| | (1) Perceptions | | (2) Misperceptions | | (3) Misperceptions and Controls | |
|---|---|---|---|---|---|---|
| | Coefficient | SE | Coefficient | SE | Coefficient | SE |
| $\text{Subjective}_{ij}$ | -0.217*** | (0.0063) | | | | |
| $\text{Mispercept}_{ij}$ | | | -0.170*** | (0.0104) | -0.155*** | (0.0105) |
| Status $\text{LITS}^{\mathcal{R}}_{ij}$ | | | -0.248*** | (0.0098) | -0.207*** | (0.0104) |
| Age | | | | | 0.003** | (0.0010) |
| Female | | | | | 0.092*** | (0.0308) |
| Married | | | | | 0.0201 | (0.0464) |
| Widowed | | | | | -0.122 | (0.0727) |
| Child in household | | | | | 0.0289 | (0.0194) |
| Student | | | | | 0.0428 | (0.1037) |
| Observations | 84,486 | | 38,314 | | 38,314 | |
| Countries | 34 | | 34 | | 34 | |
| Country-fixed effects | Yes | | Yes | | Yes | |
| Wave-fixed effects | Yes | | Yes | | Yes | |
| R-squared | 0.084 | | 0.129 | | 0.133 | |
| F-Stat | 207.9*** | | 170.1*** | | 153.1*** | |

*Notes*: Robust standard errors adjusted for clustering by countries are in parentheses. The dependent variable is $\text{Pref}_{ij}$, which gives the subjective preference for income redistribution on a scale running from 1 (low) to 10 (high). Status $\text{LITS}^{\mathcal{R}}_{ij}$ denotes the socio-economic status based upon the LITS data. In order to rule out issues with multi-collinearity, the table does not include further measures of socio-economic status, such as income, education, and occupational status. "Age" gives the age of respondents in years, "Female" is a dummy variable that equals 1 if the respondent is female, "Married", "Student", and "Retired" are dummy variables that control for the corresponding status, "Child in household" equals 1 if a child lives in the household of the respondent. The data refer to the Life in Transition Survey I, II, and III. Due to data availability, Column (1) uses all available data in the LITS (Waves I, II, and III), and Columns (2) and (3) only use the most recent wave. All models include fixed effects for countries and waves.

*** Significant at the 1 percent level,
** Significant at the 5 percent level,
* Significant at the 10 percent level

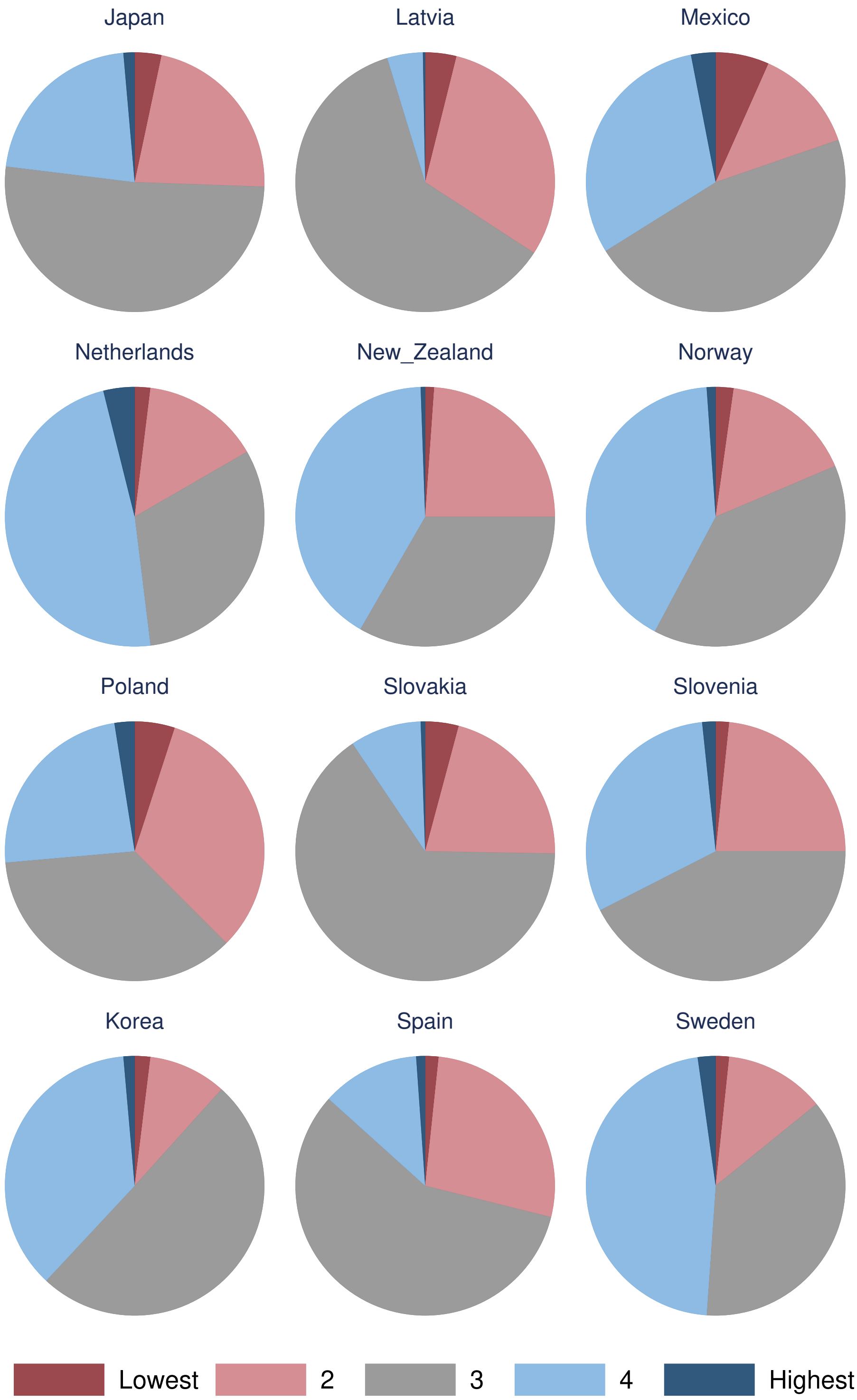
## C. Supplementary Figures

**Figure C-1** SUBJECTIVE PERCEPTIONS ABOUT SOCIAL STATUS IN OECD COUNTIES (1/2)



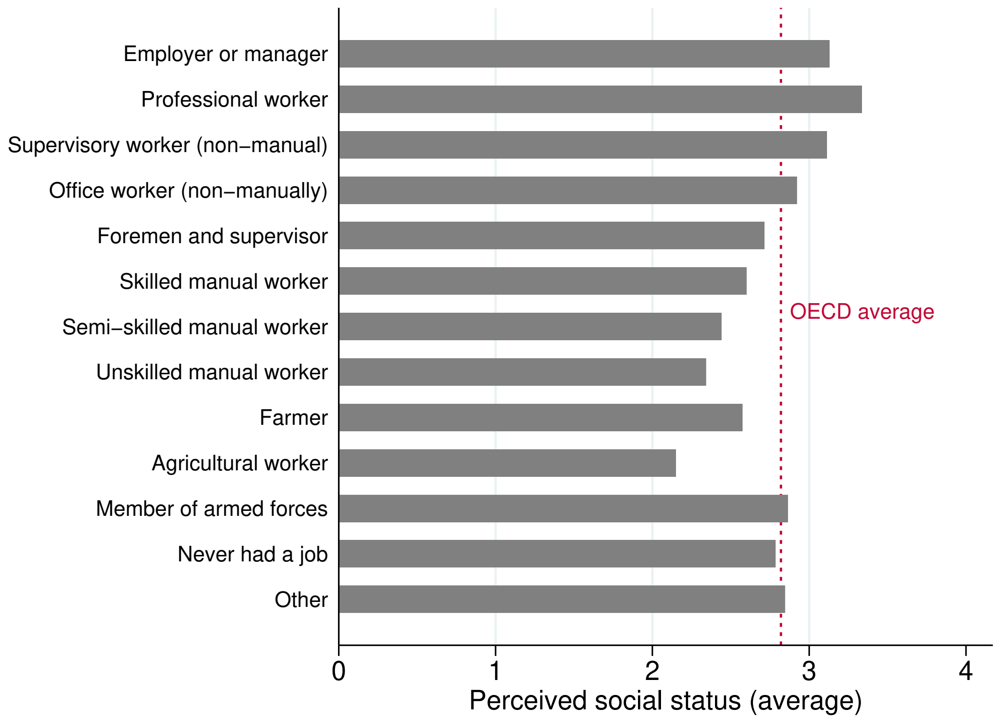

*Notes*: The figure shows self-assessment of social status by individuals on a five-scale ladder (running from 1, the lowest possible level, to 5, the highest possible level). The graph uses information on all individuals living in the respective OECD country.

**Figure C-2** SUBJECTIVE PERCEPTIONS ABOUT SOCIAL STATUS IN OECD COUNTIES (2/2)



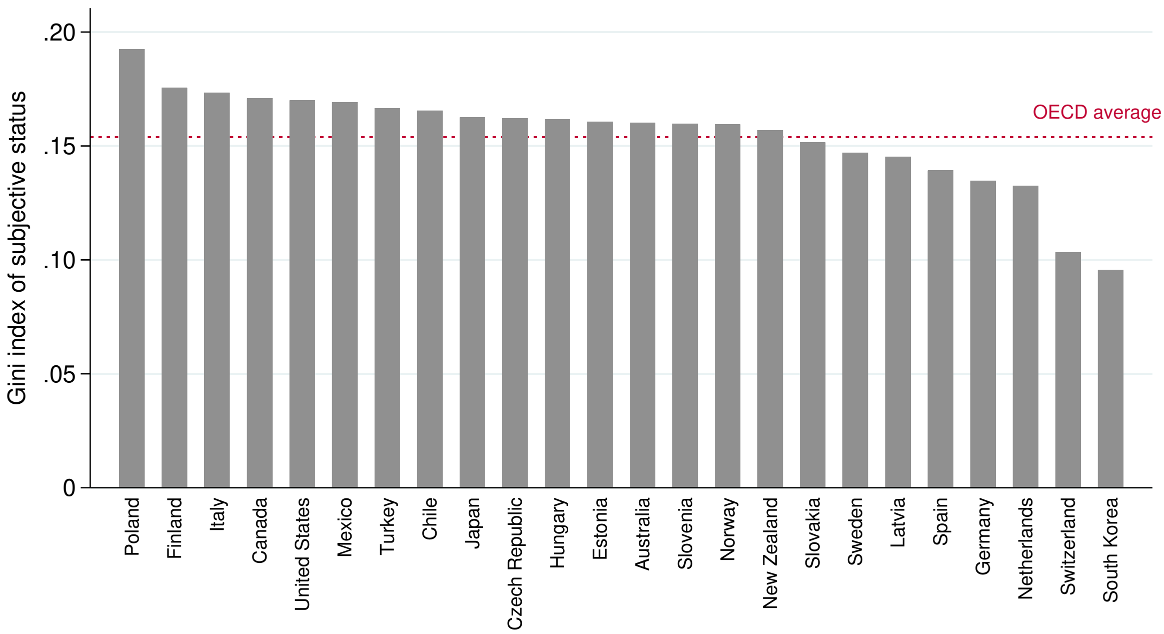

*Notes*: The figure shows self-assessment of social status by individuals on a five-scale ladder (running from 1, the lowest possible level, to 5, the highest possible level). The graph uses information on all individuals living in the respective OECD country.

**Figure C-3** SUBJECTIVE PERCEPTIONS ABOUT SOCIAL STATUS, OECD, OCCUPATION TYPES



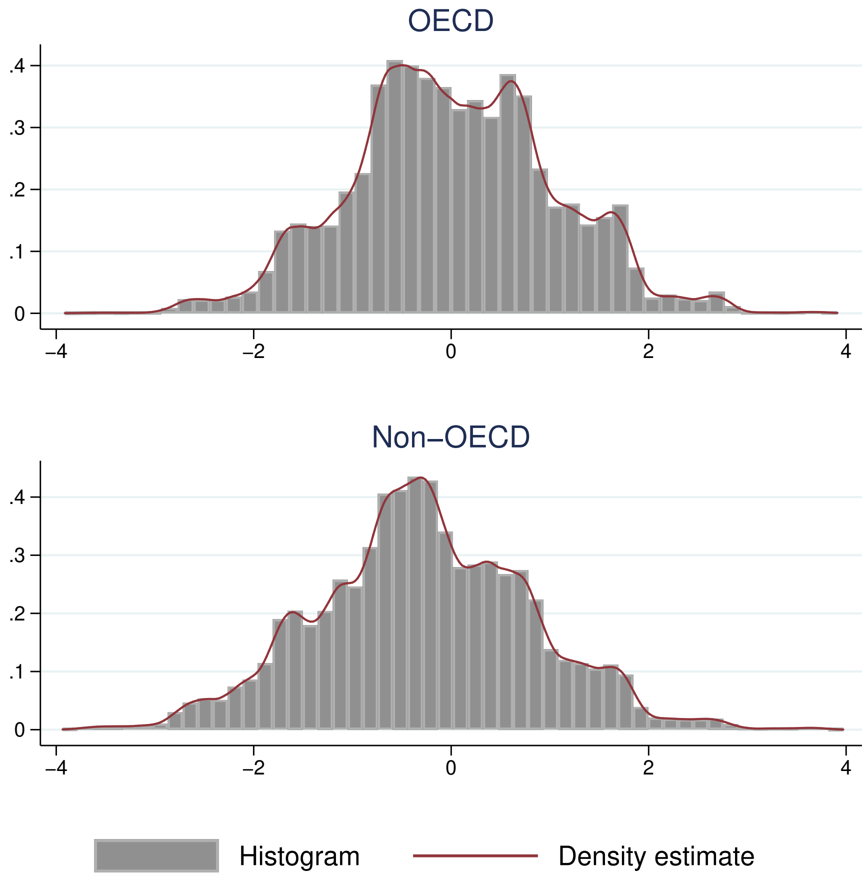

*Notes*: Subjective perceptions about social status in the OECD for different occupation types. The figure graphs the average level of subjectively perceived social status per occupation across individuals of all OECD countries included in the WVS.

**Figure C-4** GINI INDICES OF SUBJECTIVELY PERCEIVED SOCIAL STATUS



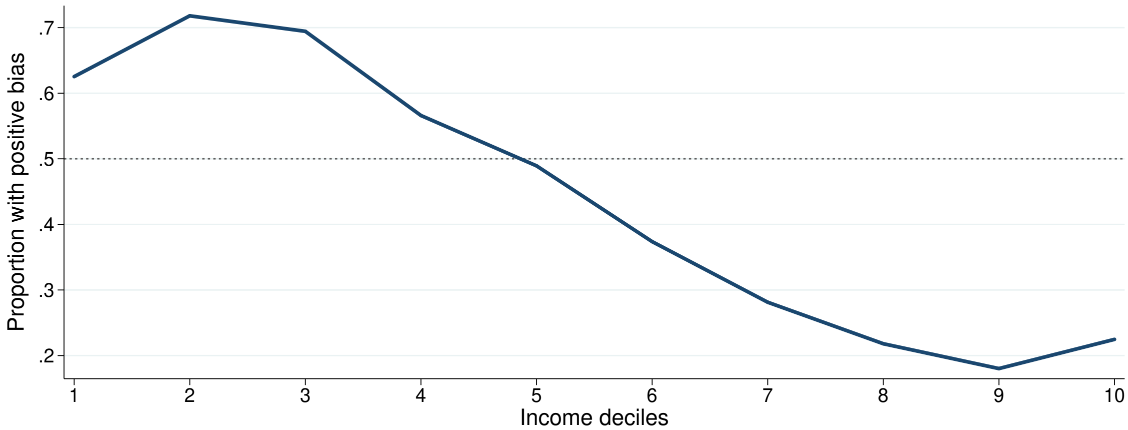

*Notes*: Gini indices of subjectively perceived social status. OECD countries. The figure shows Gini indices of self-reported social status for each of the OECD countries. Greater values reflect greater inequality in self-assessment. Gini indices are based on all available information for each OECD country.

**Figure C-5** EMPIRICAL DISTRIBUTION OF MISPERCEPTIONS, OECD AND NON-OECD



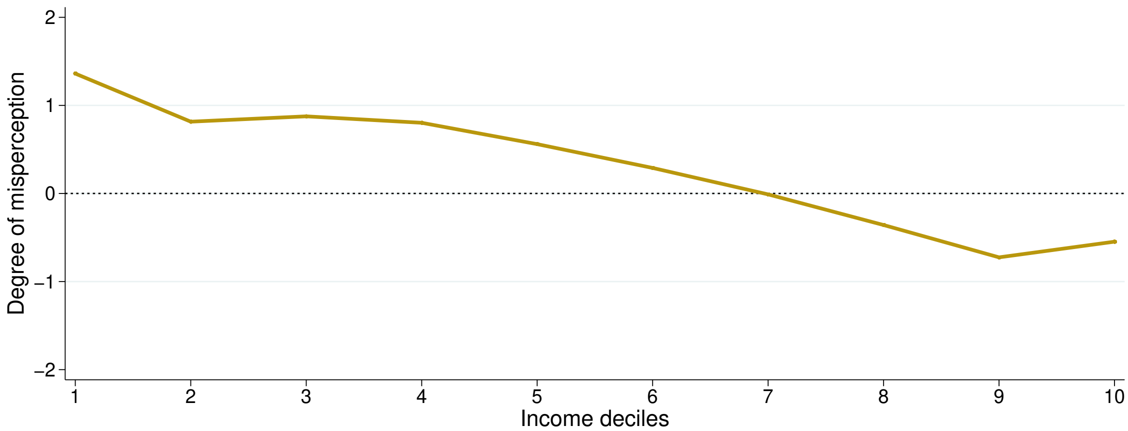

*Notes*: The empirical distribution of Mispercept$_{ij}$, OECD and Non-OECD. The figure displays the frequency (bars) and the kernel density (line) of the degree of misperception. The Epanechnikov kernel is used for the estimation of the empirical density.

**Figure C-6** MISPERCEPTIONS AT INCOME DECILES, OECD AVERAGE



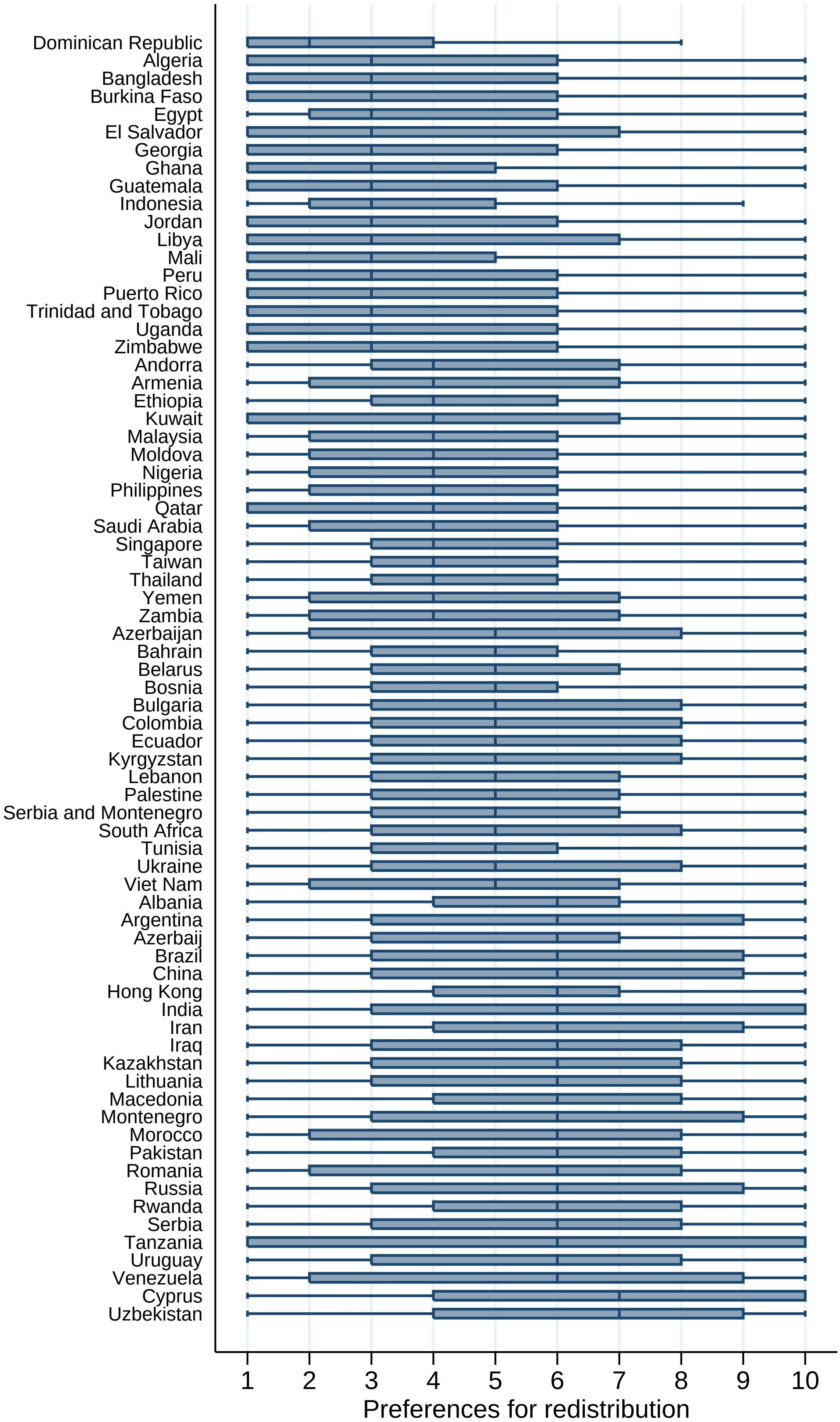

*Notes*: Misperceptions at different income deciles in the OECD. The figure shows the proportion of households with a positive bias for different income deciles.

**Figure C-7** MISPERCEPTIONS ACROSS INCOME DECILES, SWEDEN



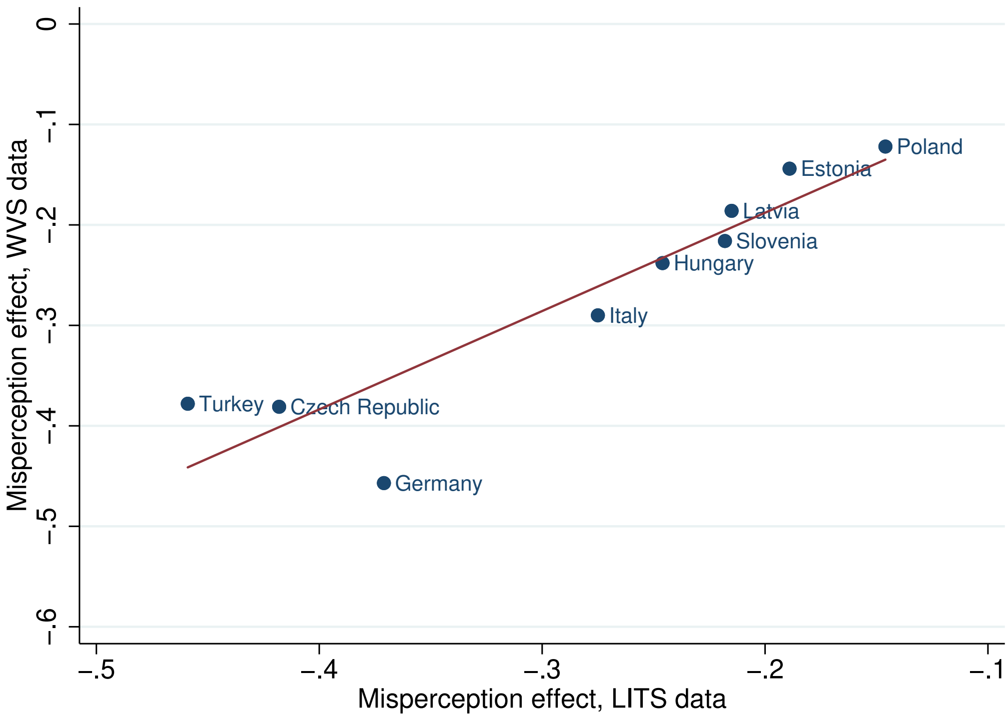

*Notes*: Misperceptions at different income deciles in Sweden. The figure shows the average degree of Mispercept$_{ij}$ for different income deciles. The numbers shown in the figure strongly resemble the patterns found in the experimental study of Karadja et al. (2017).

**Figure C-8** PREFERENCES FOR REDISTRIBUTION, NON-OECD COUNTRIES




*Notes*: The graph displays the median (shown by the vertical marker) and the distribution (shown by the boxes; 25th and 75th percentile) of the distribution of redistribution preferences per country. The "whiskers" (horizontal lines) show the minimum and the maximum values. Redistribution preferences are measured on a scale between 1 and 10; higher values reflect higher support for redistribution policies. Countries are ranked by the median.

**Figure C-9** CROSS-VALIDATION: MISPERCEPTION EFFECT, WVS-DATA AND LITS-DATA




*Notes*: The effect of misperceptions on redistribution preferences across OECD countries, WVS results versus LITS results. The figure reports results of equation (19) for OECD countries, estimated using WVS and LITS data. The empirical specifications are identical to the model reported in Column (2) of Table (1).

# References


Abe, S. (2010). *Support Vector Machines for pattern classification*. Springer, London, 2 edition. 14

Alesina, A. and Angeletos, G.-M. (2005). Fairness and redistribution. *American Economic Review*, 95(4):960–980. 12

Alesina, A. and Giuliano, P. (2011). Preferences for redistribution. In Benhabib, J., Bisin, A., and Jackson, M. O., editors, *Handbook of Social Economics, Volume 1*, pages 93–131. North Holland, Amsterdam. 3, 6, 8, 10, 23

Alesina, A. and La Ferarra, E. (2005). Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89(5-6):897–931. 11

Alesina, A., Miano, A., and Stantcheva, S. (2018a). Immigration and redistribution. *NBER Working Paper No.24733*. 2, 6

Alesina, A., Stantcheva, S., and Teso, E. (2018b). Intergenerational mobility and preferences for redistribution. *American Economic Review*, 108(2):521–554. 6

APA (2007). *Report of the APA task force on socioeconomic status*. American Psychological Association, Washington, DC. 14

Aronsson, T. and Johansson-Stenman (2008). When the Joneses' consumption hurts: Optimal public good provision and nonlinear income taxation. *Journal of Public Economics*, 92:986–997. 10

Atkinson, A. and Brandolini, A. (2001). Promise and pitfalls in the use of "secondary" data-sets: Income inequality in OECD countries as a case study. *Journal of Economic Literature*, 39(3):771–799. 48

Ballard, C. and Gupta, S. (2018). Perceptions and realities of average tax rates in the federal income tax: Evidence from Michigan. *National Tax Journal*, 71(2):263–294. 5

Becker, G. S. (1974). A theory of social interactions. *Journal of Political Economy*, 82(6):1063–1093. 9, 11

Bénabou, R. and Ok, E. A. (2001). Social mobility and the demand for redistribution: The POUM-hypothesis. *Quarterly Journal of Economics*, 116(2):447–487. 3, 6, 10

Bilancini, E. and Boncinelli, L. (2012). Redistribution and the notion of social status. *Journal of Public Economics*, 96(9–10):651–657. 4

Blinder, A. and Krueger, A. (2004). What does the public know about economic policy, and how does it know it? *Brooking Papers on Econmic Activity*, 35(1):327–397. 5

Böhlmark, A. and Lindquist, M. (2006). Life-cycle variations in the association between current and lifetime income: Replication and extension for Sweden. *Journal of Labor Economics*, 24(4):879–896. 10

Boser, S., Guyon, I., and Vapnik, V. (1992). A training algorithm for optimal margin classifiers. In Haussler, D., editor, *Proceedings of the Annual Conference on Computational Learning Theory*, pages 144–152. ACM Press, Pittsburg (US). 40

Boskin, M. and Sheshinski, E. (1978). Optimal redistributive taxation when individual welfare depends upon relative income. *Quarterly Journal of Economics*, 92(4):589–601. 10

Bound, J., Brown, C., and Mathiowetz, N. (2001). Measurement erros in survey data. *Handbook of Econometrics*, 5:3705–3843. 3

Bradley, R. H. and Corwyn, R. F. (2002). Socioeconomic status and child development. *American Journal of Sociology*, 53(1):371–399. 12

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, 16(3):199–231. 14, 39

Brown-Iannuzzi, J., Lundberg, K. B., Kay, A. C., and Payne, B. K. (2014). Subjective status shapes political preferences. *Psychological Science*, 26(1):15–26. 2

Carlsson, F., Johansson-Stenman, O., and Martinsson, P. (2007). Do you enjoy having more than others? survey evidence on positional goods. *Economica*, 74(296):486–598. 9

Chan, T. and Goldthorpe, J. (2007). Class and status: The conceptual distinction and its empirical relevance. *American Sociological Review*, 72(4):512–532. 12, 41

Chetty, R., Friedman, J., and Saez, E. (2013). Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings. *American Economic Review*, 103(7):2683–2721. 6

Clark, A. E., Frijters, P., and Shields, M. A. (2008). Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic Literature*, 46(1):95–144. 2, 9

Corneo, G. and Grüner, H. (2000). Social limits to redistribution. *American Economic Review*, 90(5):1491–1507. 2

Corneo, G. and Grüner, H.-P. (2002). Individual preferences for political redistribution. *Journal of Public Economics*, 83(1):1491–1507. 2, 23

Cox, D. (1987). Motives for private income transfers. *Journal of Political Economics*, 95(3):508–546. 3, 11

Cox, D., Hansen, B., and Jimenez, E. (2004). How responsive are private transfers to income? Evidence from a laissez-faire economy. *Journal of Public Economics*, 88(9–10):2193–2219. 3, 11

Cruces, G., Perez-Truglia, R., and Tetaz, M. (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics*, 98:100–112. 2, 5, 8, 19, 37

De Bartolome, C. (1995). Which tax rate do people use: Average or marginal? *Journal of Public Economics*, 56(1):79–96. 5

Decanq, K. and Lugo, M. A. (2013). Weights in multidimensional indices of wellbeing: An overview. *Econometric Review*, 32(1):7–34. 13

Duesenberry, J. S. (1949). *Income, Savings and the Theory of Consumer Behavior*. Harvard University Press, Cambridge, MA. 12

Duncan, O. D. (1961). A socioeconomic index for all occupations. In Reiss, A. J., editor, *Occupations and Social Status*. Free Press, New York. 12

Durante, R., Putterman, L., and van der Weele, J. (2018). Preferences for redistribution and perception of fairness: An experimental study. *Journal of the European Economic Association*, 71(2):263–294. 6

Ellison, N., Steinfield, C., and Lampe, C. (2007). The benefits of facebook "friends": Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168. 41

Engelhardt, C. and Wagener, A. (2017). What do Germans think and know about income inequality? A survey experiment. *Socio-Economic Review*, pages 1–18. 2, 5, 19, 20

Falk, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. *Quarterly Journal of Economics*, 133(4):1645–1692. 29

Falk, A., Kosse, F., Pinger, P., Schildberg-Hörisch, H., and Deckers, T. (2020). Socio-economic status and inequalites in children's IQ and economic preferences. *Journal of Political Economy*, forthcoming. 3, 15, 43

Feldman, N., Katuščák, P., and Kawano, L. (2016). Taxpayer confusion: Evidence from the child tax credit. *American Economic Review*, 106(3):807–835. 6

Fong, C. (2001). Social preferences, self-interest, and the demand for redistribution. *Journal of Public Economics*, 82(2):225–246. 23

Gründler, K. and Köllner, S. (2017). Determinants of governmental redistribution: Income distribution, development levels, and the role of perceptions. *Journal of Comparative Economics*, 45(4):930–962. 6, 12

Gründler, K. and Köllner, S. (2020). Culture, diversity, and the welfare state. *Journal of Comparative Economics*, 48(4):913–932. 12

Gründler, K. and Krieger, T. (2016). Democracy and growth: Evidence from a machine learning indicator. *European Journal of Political Economy*, 45(S):85–107. 14

Gründler, K. and Krieger, T. (2021a). Should we care (more) about data agregation? *European Economic Review*, forthcoming. 3, 14

Gründler, K. and Krieger, T. (2021b). Using machine learning for measuring democracy: A practitioners' guide and a new updated dataset for 186 countries from 1919 to 2019. *European Journal of Political Economy*, page 102047. 14, 35, 42

Guiliano, P. and Spilimbergo, A. (2014). Growing up in a recession. *Review of Economic Studies*, 81(2):787–817. 6, 24

Haider, S. and Solon, G. (2006). Life-cycle variation in the association between current and lifetime earnings. *American Economic Review*, 96(4):1308–1320. 3, 10

Hobbes, T. (1651). *Leviathan or the Matter, Forme and Power of a Commonwealth Ecclesiasticall and Civil*. reprint edited by Noel Malcolm (2012), Oxford University Press, Oxford. 3, 12

Hollingshead, A. B. (1971). Commentary on the "indiscriminate state of social class measurement". *Social Forces*, 49(4):563–567. 12

Hollingshead, A. B. and Redlich, F. (1958). *Social class and mental illness: Community study*. Wiley, Hoboken, NJ. 12

Hvidberg, K. B., Kreiner, C., and Stantcheva, S. (2020). Social position and fairness views. *NBER Working Paper No.28099*. 2, 3, 6

Ioannidis, J. P. A., Stanley, T. D., and Doucouliagos, H. (2017). The power of bias in economics research. *Economic Journal*, 127(605):F236–F265. 6

Jackman, M. R. and Jackman, R. W. (1973). An interpretation of the relation between objective and subjective social status. *American Sociological Review*, pages 569–582. 16

Kahneman, D., Slovic, P., and Tversky, A. (1982). *Judment under Uncertainty: Heuristics and Biases*. Cambridge University Press. 8

Kahneman, D. and Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3(3):430–454. 8

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291. 9

Karadja, M., Mollerstrom, J., and Seim, D. (2017). Richer (and holier) than thou? The effect of relative income improvements on demand for redistribution. *Review of Economics and Statistics*, 99(2):201–212. 2, 5, 19, 20, 60

Keely, L. and Tan, C. M. (2008). Understanding preferences for income redistribution. *Journal of Public Economics*, 92(5-6):944–961. 2, 5

Kuziemko, I., Buell, R., Reich, T., and Norton, M. (2014). "Last-place aversion": Evidence and redistributive implications. *Quarterly Journal of Economics*, 129(1):105–149. 6

Kuziemko, I., Norton, M., Saez, E., and Stantcheva, S. (2015). How elastic are preferences for redistribution? Evidence from randomized survey experiments. *American Economic Review*, 105(4):1478–1508. 2, 6

Lefgren, L. J., Sims, D. P., and Stoddard, O. B. (2016). Effort, luck, and voting for redistribution. *Journal of Public Economics*, 143:89–97. 12

Ljungqvist, L. and Uhlig, H. (2000). Tax policy and aggregate demand management under catching up with the Joneses. *American Economic Review*, 90(3):356–366. 10

Lockwood, B. and Weinzierl, M. (2015). De gustibus non est taxandum: Heterogeneity in preferences and optimal redistribution. *Journal of Public Economics*, 124:74–80. 6

Lupu, N. and Pontusson, J. (2011). The structure of inequality and the politics of redistribution. *American Political Science Review*, 105(2):316–336. 34, 48

Luttmer, E. F. and Singhal, M. (2011). Culture, context, and the taste for redistribution. *American Economic Journal: Economic Policy*, 3(1):157–179. 6, 25, 32

Luttmer, E. F. P. (2005). Neighbors as negatives: Relative earnings and well-being. *Quarterly Journal of Economics*, 120(3):963–1002. 3, 4, 9

Marks, G. N. (2011). Issues in the conceptualisation and measurement of socioeconomic background: Do different measures generate different conclusions? *Social Indicators Research*, 104(2):225–251. 3, 13

Marx, K. (1867). *Das Kapital: Kritik der politischen Oekonomie. Volume 1: Der Produktionsprozess des Kapitals*. Hamburg, 1 edition. 12

Mattera, D. and Haykin, S. (1999). Support Vector Machines for dynamic reconstruction of a chaotic system. In Burges, C., Schölkopf, B., and Smola, A., editors, *Advances in Kernel Methods. Suppoert Vector Learning.*, pages 211–241. MIT-Press, Cambridge, MA. 42

Mayer, B. D., Mok, W. K., and Sullivan, J. X. (2015). Household surveys in crisis. *Journal of Economic Perspectives*, 29(4):199–226. 3, 11

Meltzer, A. H. and Richard, S. F. (1981). A rational theory of the size of government. *Journal of Political Economy*, 89(5):914–927. 2, 6, 7

Mills, C. W. (1651). *The Power Elite*. Oxford University Press, London. 12

Mittag, N. (2019). Correcting for misreporting of government benefits. *American Economic Journal: Economic Policy*, 11(2):142–64. 3, 11

Moreno-Maldonado, C., Rivera, F., Ramos, P., and Moreno, C. (2018). Measuring the socioeconomic position of adolescents: A proposal for a composite index. *Social Indicators Research*, 136:517–538. 3, 12

Nam, C. B. and Boyd, M. (2004). Occupational status in 2000. Over a century of census-based measurement. *Population Research and Policy Review*, 23:327–358. 12

Oakes, M. and Rossi, P. H. (2003). The measurement of ses in health research: Current practice and steps toward a new approach. *Social Science and Medicine*, 56(4):769–784. 3, 13

Oswald, A. (1983). Altruism, jealousy and the theory of optimal non-linear taxation. *Journal of Public Economics*, 20:77–87. 9

Reardon, S. F. and Bischoff, K. (2011). Income inequality and income segregation. *American Journal of Sociology*, 116(4):1092–1153. 9

Rees-Jones, A. and Taubinsky, D. (2020). Measuring "schmeduling". *Review of Economic Studies*, 87:2399–2438. 5

Romer, T. (1975). Individual welfare, majority voting, and the properties of a linear income tax. *Journal of Public Economics*, 4(2):163–185. 2, 6, 7

Roth, C. and Wohlfahrt, J. (2018). Experienced inequality and preferences for redistribution. *Journal of Public Economics*, 167:251–262. 24

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York, NY. 29, 53

Rueda, D. and Stegmueller, D. (2019). *Who Wants What?: Redistribution Preferences in Comparative Perspective*. Cambridge University Press. 3, 11

Saenger, G. H. (1945). Social status and political behavior. *American Journal of Sociology*, 51(2):103–113. 12

Sapienza, P. and Zingales, L. (2013). Economic experts versus average Americans. *American Economic Review*, 103(3):636–642. 5

Scartascini, C., Cruz, C., and Keefer, P. (2017). *The Database of Political Institutions 2017*. Inter-American Development Bank, Washington D.C. 34

Schölkopf, B. and Smola, A. J. (2001). *Learning with Kernels*. MIT-Press, Cambridge (US) and London (UK). 39

Smola, A. J. and Schölkopf, B. (2004). A tutorial on Support Vector Regression. *Statistics and Computing*, 14(3):199–222. 14

Solnick, S. J. and Hemenway, D. (2005). Are positional concerns stronger in some domains than others? *American Economic Review: Papers and Proceedings*, 95(2):147–151. 3

Solt, F. (2016). The Standardized World Income Inequality Database. *Social Science Quarterly*, 97(5):1267–1281. 34, 48

Stantcheva, S. (2020a). How people understand and reason about health insurance in the U.S.? Harvard University Working Paper. 5

Stantcheva, S. (2020b). Understanding economic policies: What do people know and learn? Harvard University Working Paper. 5

Stantcheva, S. (2020c). Understanding tax policy: How do people reason. NBER Working Paper No.27699. 2, 5

Steinwart, I. and Christmann, A. (2008). *Support Vector Machines*. Springer Verlag, New York (US). 39

Stevens, G. and Featherman, D. L. (1981). A revised socioeconomic index of occupational status. *Social Science Research*, 10(4):364–395. 12

Ursprung, H. and Breyer, F. (1998). Are the rich too rich to be expropriated? Economic power and the feasibility of constitutional limits to redistribution. *Public Choice*, 94(1-2):135–156. 6

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer, New York (US). 14, 39, 40

Vapnik, V. N. (1998). *Statistical Learning Theory*. John Wiley & Sons, New York (US). 14, 39, 40

Vyas, S. and Kumaranayake, L. (2006). Constructing socio-economic indices: How to use Principal Component Analysis. *Health Policy and planning*, 21(1):459–468. 13

Weber, M. (1922). *Economy and Society*. University of California Press, reprint 1978, Berkeley and Los Angeles, CA. 12

Weiss, Y. and Fershtman, C. (1998). Social status and economic performance: A survey. *European Economic Review*, 42(3–5):801–820. 5, 11