OESTERREICHISCHE NATIONALBANK
EUROSYSTEM

WORKING PAPER 216

Perceptions of Inequality

Markus Knell and Helmut Stix

The *Working Paper series of the Oesterreichische Nationalbank* is designed to disseminate and to provide a platform for discussion of either work of the staff of the OeNB economists or outside contributors on topics which are of special interest to the OeNB. To ensure the high quality of their content, the contributions are subjected to an international refereeing process. The opinions are strictly those of the authors and do in no way commit the OeNB.

The Working Papers are also available on our website (http://www.oenb.at) and they are indexed in RePEc (http://repec.org/).

Publisher and editor	*Oesterreichische Nationalbank* *Otto-Wagner-Platz 3, 1090 Vienna, Austria* *PO Box 61, 1011 Vienna, Austria* *www.oenb.at* *oenb.info@oenb.at* *Phone (+43-1) 40420-6666* *Fax (+43-1) 40420-046698*
Editorial Board of the Working Papers	Doris Ritzberger-Grünwald, Ernest Gnan, Martin Summer
Coordinating editor	Martin Summer
Design	*Communications and Publications Division*

DVR 0031577

ISSN 2310-5321 (Print)
ISSN 2310-533X (Online)

Perceptions of Inequality*

Markus Knell** and Helmut Stix***
Oesterreichische Nationalbank

October 2017

Abstract

Although people's perception of (income or wealth) inequality has important effects on their decisions as economic agents or voters, little is known about how perceptions relate to measured inequality. We present a novel formal framework that is based on the assumption that people typically do not observe the entire income (wealth) distribution and that their guesses about the extent of inequality are based on reference groups. This framework predicts that perceptions of inequality will change along positions in the income distribution and that for a specific position various dimensions of inequality perception are related to each other. First, low (high) income individuals overestimate (underestimate) their own position. Second, subjective estimates of average earnings increase with the own income position. Third, high or low income people have different perceptions about the "distributional shape" of society (e.g. pyramid or diamond). Fourth, the subjective perception of inequality is lower for high-income individuals. Survey data from 40 countries provide strong support for the framework.

Keywords: Income distribution, perception of inequality, reference groups
JEL-Classification: D31, D63, D83, C81

*The views expressed in this paper do not necessarily reflect those of the Oesterreichische Nationalbank. We thank Marcel Kropp for excellent research assistance.

**OeNB, Economic Studies Division, Otto-Wagner-Platz 3, POB-61, A-1011 Vienna; Phone: (++43-1) 40420 7218, Fax: (++43-1) 40420 7299, Email: Markus.Knell@oenb.at.

***OeNB, Economic Studies Division, Otto-Wagner-Platz 3, POB-61, A-1011 Vienna; Phone: (++43-1) 40420 7211, Fax: (++43-1) 40420 7299, Email: Helmut.Stix@oenb.at.

Non-Technical Summary

The topic of inequality was high on the agenda in recent years. Academic research and the public alike are discussing the sources of inequality in income and wealth, the reasons for the apparent increase in inequality and possible consequences of these developments for society at large. For a number of potential effects it is, however, not the *measured* inequality that is important but rather the *perceived* level of inequality. This is true, e.g., for voting behavior, preferences for redistribution or the likelihood of social unrest and migration. It is therefore important to understand how people form their perceptions of inequality, to which extent these conform to measured inequality and whether there are systematic differences in perceptions that can be explained by socio-economic variables.

A natural starting point to deal with this issue is the use of surveys in which people are asked about their assessment of various dimensions of inequality. In the paper we use four types of questions about individuals' evaluation of (i) their own subjective rank, (ii) the distributional shape of society, (iii) the mean income, and (iv) the extent of income inequality. Different to the related literature we do not look at each of these questions in isolation but we investigate whether and how the answers are related. In particular, we are interested how the pattern depends on individuals' position in the (measured) income distribution.

We do not stop, however, with the analysis of empirical regularities. We also present a theoretical framework that allows us to explain the systematic differences of inequality perceptions as well as the pattern of answers to different survey questions. The framework rests on the idea that people form their perception of the income distribution by looking at reference groups that are not arbitrary subsamples of society but are typically biased towards individuals of a similar social background. When asked about their assessment of the distribution individuals base their answer on a subjective income distribution formed from this biased subsample. We analyze different specifications of possible reference group formation.

The empirical results are closely in line with the implications of the theoretical framework. First, there is a considerable degree of over- and underestimation of the subjective rank for low-income and high-income people, respectively. Second, we find that high-income people tend to give higher estimates for earnings of different professions. Third, low-income people predominately view society as a pyramid while high-income people have a stronger tendency to view it as a vase. Fourth, the subjective Gini decreases with the position in the distribution. All of these results are in line with the predictions of the theoretical model and they support the hypotheses that people have reference groups that are socially biased and centered around the own income level. This suggests that inequality misperceptions are not simply an implication of arbitrary survey responses but rather are the direct consequence of social stratification due to reference group formation.

1 Introduction

Human action depends not only on individual preferences and resource constraints but also on expectations and beliefs, for example about the unknown future, other peoples' behavior and the state of the economy. The recent years have shown a particular interest in how people assess their social and economic environment and whether their perceptions concur with reality. The discussion of a "post-factual era" and on how subjective and often severely biased views of reality influence individuals' actions as citizens provides a general case in point. The perception of the extent of income or wealth inequality has attracted particular attention given its importance for voting behavior, redistributive preferences, the likelihood of social upheaval or migration (Alesina & La Ferrara 2005) as well as on attitudes like life satisfaction and trust (Clark & D'Ambrosio 2015, Knell & Stix 2017).

The existing empirical literature has established that people have biased perceptions of the extent of inequality (Slemrod 2006, Osberg & Smeeding 2006, Norton & Ariely 2011, Kuhn 2011, Gimpelson & Treisman 2015). These studies are based on various survey questions and refer to different dimensions of inequality. What is missing so far is an encompassing and multidimensional view on this topic, both theoretically and empirically. From an empirical perspective, it is interesting to investigate how answers to different survey questions are related to each other and how this pattern itself might be associated with a respondent's socio-economic position. From a theoretical angle one would like to know the causes of perception biases. In particular, an appropriate theoretical framework should be able to explain the systematic differences of inequality perceptions as well as the pattern of answers to different survey questions in a parsimonious manner.

In this paper we extend the existing literature both along the theoretical and along the empirical dimension. We provide a unified framework that has testable implications about how agents perceive various aspects of the income distribution. The framework rests on the idea that people form their perception of the income distribution by looking at reference groups that are not arbitrary subsamples of society but are typically biased towards individuals that have a similar social and economic background. When asked about their assessment of the distribution individuals base their answer on a subjective income distribution formed from this biased subsample.[1] We look at two ways how reference groups can be formed. In the first case it is assumed that individuals only focus on incomes in a

[1] On a discussion on this topic and of a distinction between "comparative" and "normative" reference groups see Clark & D'Ambrosio (2015).

given perception span around their own position in the distribution ("limited perception span"). In the second case individuals consider the entire distribution of all income levels but put more weight on income levels close to their own position ("self-centered density function").

We document the implications of the two subjective distribution functions for various measures of distributional perceptions. Thereby we focus on measures of the perception of the income distribution that are directly related to variables that are included in existing surveys. We show how individuals at different deciles in the (objective) income distribution will (i) assess their own subjective rank (cf. "social-ladder question"), (ii) the distributional shape of society (cf. "type-of-society question"), (iii) the mean income (cf. "actual-earnings question") and (iv) the extent of income inequality (cf. the subjective Gini coefficient based on the "actual-earnings" or the "normative question"). Most importantly, our model demonstrates that these different dimensions of inequality are connected while the previous literature has treated them separately. To the best of our knowledge this is the first attempt to connect these different dimensions within a unified framework.

The framework predicts, regardless of the chosen subjective distribution function, that (i) the subjective rank is positively but not proportionally related to the objective rank, i.e. low-income (high-income) people will tend to overestimate (underestimate) their position and (ii) that an individual's estimate of the average income level increase with his income level. The assessment of the extent of inequality, however, differs with respect to how reference groups are formed. A fixed perception span implies that low income individuals will view the shape of society as a "vase" (an inverted pyramid) while high-income people will view it as a normal pyramid. The subjective Gini coefficient will be largest for the lower and upper ends of the distribution and lowest in the middle. In contrast, the assumption of a self-centered subjective density function has reverse implications: low-income people will see society as a pyramid and high-income people as a vase while the subjective Gini coefficient is negatively related to the objective position in the income distribution.

We confront the implications of the theoretical framework with data from the International Social Survey Project (ISSP) which contains information on 40 countries. While this data set provides information on all four model predictions, the key assumption about the existence of heterogeneous reference groups cannot be tested directly since the ISSP does not contain information about how individuals form reference groups.[2] The

[2]In fact, this is true for most existing surveys we know of. There is only a small set of surveys that

use of different measures of inequality perceptions allows, however, for an indirect test of the role of reference groups. In particular, if the answers to several questions about inequality perception are in line with the predictions of the theoretical framework then this provides suggestive evidence that heterogeneous reference groups are an important element to understand the pattern of these perceptions. Moreover, we can use the fact that the assumption of "limited perception span" and the "self-centered density function" yield competing implications for a sub-set of perception variables to investigate whether the data correspond better with one or the other of the two assumptions concerning the formation of reference groups.

The empirical results provide strong support for our framework. First, there is a considerable degree of over- and underestimation of the subjective rank for low-income and high-income people, respectively. On average, individuals place themselves into the range between the fourth and the sixth decile. This outcome is not only qualitatively but quantitatively similar to the results of the model using a self-centered subjective density function. Second, we find that high-income people tend to give higher estimates for earnings of characteristic professions (like unskilled worker or shop-assistant) which is again in line with the predictions of the model. Third, low-income people predominately view society as a pyramid while high-income people have a stronger tendency to view it as a vase. This again supports the assumption of a self-centered subjective density function. Fourth, the subjective Gini and the proportion of people that see income differences as being too large decreases with the position in the distribution. Under some additional assumptions discussed later, this is also compatible with the prediction of the model. Overall, the empirical results strongly support the hypothesis that individual differences in the perception of the income distribution are influenced by the presence of heterogeneous reference groups that are socially biased and centered around the own income level. This suggests that inequality misperceptions are not simply an implication of arbitrary survey responses or unsystematic biases but rather are the direct and inherently consistent consequence of social stratification due to reference group formation.

Finally, the joint look at several dimensions of inequality allows for a nuanced and cautious view on the question whether people tend to overestimate or underestimate the extent of inequality. In particular, the recent literature has produced conflicting results on this issues based on the use of different survey questions. Some authors have used the "type-of-society question" to conclude that people tend to drastically overestimate

contains information about the identity of reference groups members (see e.g. Clark & Senik 2010). For a general discussion, see Clark & D'Ambrosio (2015).

the extent of inequality (Niehues 2014) while others have employed the "social-ladder question" to arrive at the contrary conclusion (Cruces et al. 2013). We argue that the type-of-society question alone is not suitable for settling this issue conclusively since it is impossible to transform individual answers to the type-of-society question into an unequivocal measure of aggregate perceived inequality—like a subjective Gini coefficient. We suggest that the look at several dimensions of inequality perceptions gives a more informative answer. Our results imply that most people considerably *underestimate* inequality. This finding, however, has to be taken with caution as it hinges critically on various assumptions concerning the choice of reference groups.

The paper builds on several strands of the literature. A number of papers document that the perceived degree of inequality deviates considerably from the measured extent of inequality (Slemrod 2006, Osberg & Smeeding 2006, Norton & Ariely 2011, Kuhn 2011, Gimpelson & Treisman 2015). An additional strand has investigated whether people's perceptions and policy preferences change when they are provided with "informational treatments" about the reality (Cruces et al. 2013, Kuziemko et al. 2015). Other papers have used cross-sectional estimations to compare the explanatory power of measured and perceived inequality. Niehues (2014), Engelhardt & Wagener (2014) and Gimpelson & Treisman (2015) have shown, e.g., that cross-country variations in the perception of inequality are more successful in explaining variations in redistributive preferences than aggregate measures of income inequality. Clark & Senik (2010), Clark et al. (2013) and Mayraz et al. (2009) study the importance and the specific identity of reference groups. Clark & Senik (2010), e.g., use a large European survey and find, inter alia, that individuals tend to compare themselves to those with whom they interact most often.

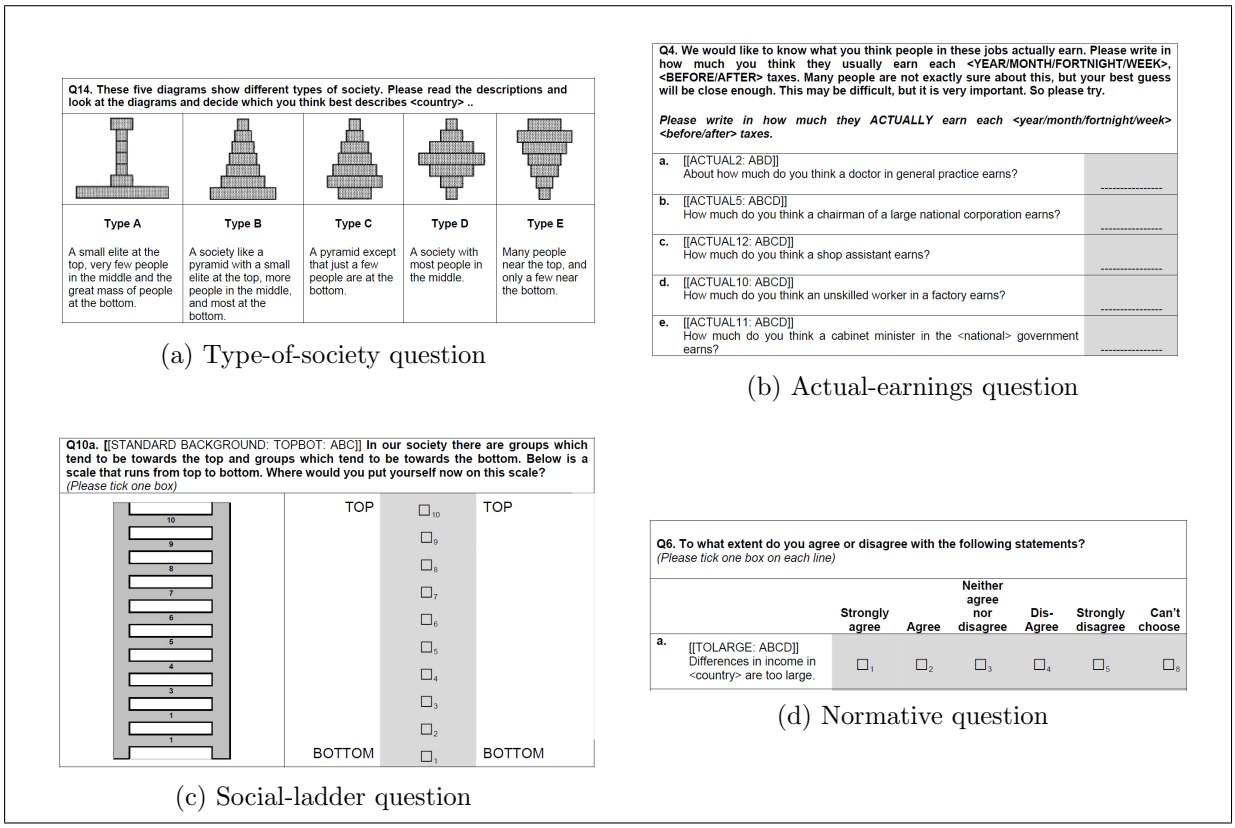
The paper is organized as follows. In the next section we discuss the main survey questions on distributional perceptions in detail and we show why the type-of-society question cannot easily be transformed into a subjective Gini coefficient. Section 3 presents the theoretical framework based on the assumption of reference groups and we discuss different possibilities for the choice of subjective distribution functions. Section 4 contains the empirical analysis and section 5 concludes.

2 Survey measures of distributional perceptions

In this section we discuss different survey measures that have been used to study individuals' perception of the income distribution and inequality.

2.1 Type-of-society question

Since 1987 the survey of the ISSP included a question in which respondents are asked to choose one among five diagrams that best describes the society in the own country (see panel (a) of figure 1). This type-of-society question has been used in Niehues (2014) and Gimpelson & Treisman (2015) to derive measures of perceived income inequality.[3] For later reference we will refer to the five types as tower (A), pyramid (B), tree (C), diamond (D) and vase (E).



(a) Type-of-society question

(b) Actual-earnings question

(c) Social-ladder question

(d) Normative question

Figure 1: Source: International Social Survey Project survey (2009) questionnaire

In contrast to these authors we argue that it is not straightforward to translate the answers to the type-of-society question into an unequivocal, quantitative measure of perceived inequality like a Gini coefficient. There are at least three obstacles to this endeavour:

[3]The latter authors also argue that although the question does not explicitly refer to income or wealth it is reasonable to assume that respondents have interpreted it in terms of income since the previous survey questions also referred to "pay" and "earnings".

1. We do not know how individuals partition the income distribution into seven classes.

2. The shape of the income distribution people have in mind might not be in line with any of the proposed types.

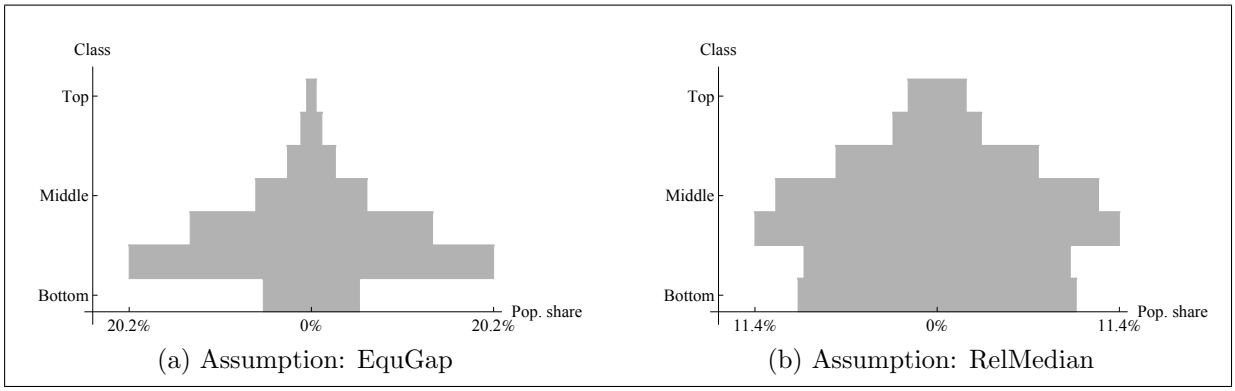
3. We do not know how an individual evaluates the *within*-class income distribution.

In appendix C we elaborate on these three issues. Here we only want to emphasize some important aspects and implications.

For the choice of the seven income levels there exist two approaches in the related literature. One approach (cf. Gimpelson & Treisman 2015) simply assumes that the income gap between all neighboring classes is identical. We refer to this assumption as EquGap. An alternative approach (cf. Niehues 2014) assumes that the seven classes correspond to fractions of median income (e.g. the second class includes all incomes between 60% and 80% of the median etc.). We call this assumption RelMedian. The chosen definitions are plausible although at the same time rather arbitrary. In figure 2 we use a numerical example to illustrate the implication of the two approaches for the shape of society. In particular, we assume that incomes are lognormally distributed with a mean (annual) income of $\overline{Y} = 50\,000$ and a standard deviation such that the Gini coefficient is 0.3.[4] There exists no a-priori reason why assumption EquGap or RelMedian should constitute a more appropriate reflection of people's views. In light of the different shapes of figures 2a and 2b we thus have a first reason to be doubtful about a straightforward matching of the five diagrams to a precise inequality measure.

In the appendix we discuss how one can use various assumptions to calculate a corresponding Gini coefficient for each of the five types A to E. One crucial assumption is that the size of each bar is interpreted as the population size of the respective class. A second crucial assumption concerns the income level that is associated with each class (and that captures the *within*-class income distribution). Depending on these assumptions (and on the choice of EquGap or RelMedian) it can be shown that the estimations of the Gini coefficient exhibit a rather wide variation. For society A, e.g., it ranges from 0.68 to 0.43 while for society E it ranges from 0.26 to 0.2. Our benchmark specification (based on EquGap) yields the following results: 0.53 (A), 0.43 (B), 0.35 (C), 0.23 (D) and 0.23 (E). It is immediately clear that this wide variation does inhibit statements about whether people under- or overestimate inequality. Despite the quantitative variation, the relative

[4]Note that for a lognormal income distribution the Gini coefficient can be calculated as $2\Phi(\frac{\sigma}{\sqrt{2}}) - 1$, where $\Phi(\cdot)$ is the CDF of the standard normal distribution. The mean of the Gini coefficient among all ISSP countries is exactly 0.3.



(a) Assumption: EquGap (b) Assumption: RelMedian

Figure 2: The figure shows the population shares of seven income classes when incomes are distributed lognormally with an (annual) mean of 50 000 and a Gini coefficient of 0.3. In panel (a) it is assumed that the income gap between two adjacent classes is the same (and given by 21 874) while in panel (b) it is assumed that the seven bars correspond to income class definitions related to the median as specified in appendix C. These two assumptions are called EquGap and RelMedian, respectively.

rank of society types with respect to the Gini coefficients is unaffected by the different assumptions. For all cases it holds that type A is perceived as the most unequal society, followed by type B and type C. The most equal societies are types D and E that are associated with rather similar Gini coefficients and which we will often lump together in the following analysis.

Summing-up, we believe that the type-of-society question provides interesting information about respondents' qualitative assessment of the income distribution. It is, however, much more problematic (and almost meaningless) to come up with precise numbers for the Gini coefficient associated with each of the five types.

2.2 Actual-earnings question

Individual perceptions of the income distribution can also be elicited by asking respondents about their assessment of various levels of earnings in characteristic jobs. In the ISSP these professions are: "doctor in general practice", "chairman of a large national corporation", "shop assistant", "unskilled factory worker", "cabinet minister in the national government" (see figure 1b). On the one hand, the resulting information can be employed to see whether the estimation of average income changes in a systematic way with the own income position.[5] On the other hand, one can use the an-

[5]It is also possible to compare the answers to the factual earnings in these professions as has been done (for Switzerland) in Kuhn (2011).

swers to form inequality measures like the perceived spread between high-income and low-income occupations or even—under some heroic assumptions—a subjective Gini coefficient (Osberg & Smeeding 2006, Kuhn 2011, Kuhn 2015, Osberg & Bechert 2016). Finally, the ISSP contains a question that asks about what people *should* earn in these professions that can then be used to compare positive and normative assessments (Osberg & Smeeding 2006, Kuhn 2011). We will say more about these questions in the empirical part of the paper.

2.3 Social-ladder question

The ISSP like a number of other survey asks individuals to locate their position in society. The exact wording of the question is: "In our society there are groups which tend to be towards the top and groups which tend to be toward the bottom. Below is a scale that runs from top to bottom (horizontal scale (10 top – 1 bottom)). Where would you put yourself now on this scale?" (see figure 1c). Again one could argue that this question does not explicitly refer to income or wealth and that people might use other categories to place themselves into the social hierarchy. There exist, however, surveys that have asked people explicitly about their guessed decile in the national income distribution (e.g. Cruces et al. 2013). The resulting pattern is qualitatively and quantitatively very similar to the one of the social-ladder question. In the related literature it is sometimes argued that the social-ladder question does not reveal anything about individuals' assessment of inequality but only about the assessment of their own position in this distribution. We believe, however, that this is an important question to gauge individual views on the shape of the income distribution and to investigate the mechanisms that are behind these views. In particular, answers to the type-of-society and the social-ladder question should not contradict each other but rather be implications of the same perception model.

2.4 Direct positive or normative questions

In a number of studies respondents have been asked directly about their perceptions of inequality. In Knell & Stix (2017), e.g., we have asked the following question: "What is your assessment about how income (the total sum of annual earnings) is distributed in <country>?" (possible answers: "extremely unequally", "very unequally", "rather unequally" and "rather equally" distributed). The ISSP has not asked positive questions like this but does contain a question on a related topic. In particular, people have been asked to which extent they agree with the statement that "differences in income in <country>

are too large" (possible answers from strongly agree to strongly disagree, see figure 1d). This is of course a normative question that can not be directly taken as a measure of perceived *actual* inequality. Empirically it has been shown that the ethical and the actual assessments are often correlated. Put differently, if one assumes that the notion of an "ethical" or "fair" income distribution is randomly distributed across members of society then people who perceive a high degree of actual inequality will on average say more often that income differences are "too large" than people who perceive a smaller degree of inequality.

3 Conceptual framework

3.1 Objective and subjective income distributions

In order to explain the pattern of answers to various distributional questions we start from the simple and quite straightforward assumption that people have different reference groups. There exists a long literature on reference groups in sociology and economics (Runciman 1966, Clark & D'Ambrosio 2015) and how they can be used to explain phenomena like relative deprivation, conspicuous consumption and "keeping up with the Joneses". For these questions it is assumed that individuals employ reference groups in order to evaluate their *own* situation in comparison to the situation of others. On the other hand, reference groups can also be used to provide an assessment of aggregate phenomena like the state of the economy or the income distribution. It is quite likely that these reference groups are not independent of the context in which they are formed. A person might, e.g., exclude the super-rich when evaluating the own economic position while including them when making an assessment about the aggregate situation. At the same time, however, it is reasonable to assume that these context-specific reference groups are not entirely different and that they will primarily differ with respect to the tails of the distribution. In general it is typically argued that reference groups are quite narrow and dominated by members with a similar social and economic background. Put differently, each individual will only use a subset of society and his or her assessment of the aggregate income distribution is based on the shape of this idiosyncratic selection. Of course there exist various channels (e.g. the influence of co-workers and in particular of the media) that provide information about the life-style and incomes of other social classes, but in general individual reference groups and their perceived income distribution will show a larger weight of the own class.

This framework can be formalized as follows.

Definition 1 (Objective Income Distribution)
Incomes $Y \in [0, \infty]$ in a country are distributed according to the probability density function $f(Y)$ with the corresponding cumulative distribution function $F(Y)$.

Due to the influence of reference groups the individual perception of the income distribution might differ from the objective one.

Definition 2 (Subjective Income Distribution)
Each individual i has a specific view of the income distribution in the country. This view is captured by a subjective probability density function $f_i(Y)$ with a support $Y \in [Y_i^L, Y_i^H]$ where Y_i^L (Y_i^H) are the lowest (highest) income in individual i's reference group. The corresponding subjective cumulative distribution function is denoted by $F_i(Y)$ where $F_i(Y_i^L) = 0$ and $F_i(Y_i^H) = 1$.

It is of course possible that the individual has a perception of the entire support of the objective function, i.e. $Y_i^L = 0$ and $Y_i^H = \infty$.

If the choice of reference groups were a random process then the heterogeneity might not have a large effect in the aggregate (due to the law of large numbers). There are, however, good reasons to believe that individual reference groups differ in a systematic and predictable manner. In particular, and as argued above, people tend to have peers that are mostly from the same or a similar class. In other words, the subjective income density function will have more weight around the own income position and less (possibly zero) weight for income levels that are far above or far below the own income. The subjective density function $f_i(Y)$ thus determines the individual assessment of the shape of society and it might well be that the latter looks like a pyramid for one group of people and like a vase for another group. The precise implications for the perceived shape will depend on the assumptions about the subjective function $f_i(Y)$ and we will give specific examples below.

Before, we want to specify a number of additional variables that can be used to capture individuals' assessment of a society's income distribution and that can be related to popular survey questions.

Definition 3 (Subjective Rank)
The subjective rank $F_i(Y_i)$ measures individual i's assessment of the position of his or her own income level Y_i in the subjective income distribution $F_i(Y)$.

For unbiased perceptions it holds that $F_i(Y) = F(Y_i)$, i.e. the objective and the subjective rank coincide. The social-ladder question can be used to get information about $F_i(Y_i)$.

Definition 4 (Subjective Mean Income)
The subjective mean income $E_i(Y)$ measures individual i's assessment of the average income in society. It is given by: $E_i(Y) \equiv \int_{Y_i^L}^{Y_i^H} Y f_i(Y)\,\mathrm{d}Y$.

For unbiased perceptions it holds that $E_i(Y) = E(Y) = \overline{Y}$. One can use questions on the guessed income of specific occupations to get information about $E_i(Y)$.

Inequality is a multi-dimensional concept that can be assessed in various manners (cf. Osberg & Bechert 2016). As discussed above, existing surveys include a number of questions that are related to the perception of inequality (e.g. the type-of-society question, the subjective Gini coefficient based on the actual-earnings question and direct positive or normative questions). In order to map the answers to these questions to the framework based on reference groups we use the subjective Gini coefficient.[6]

Definition 5 (Subjective Gini Coefficient)
The subjective Gini coefficient G_i is given by:

$$G_i = \int_{Y_i^L}^{Y_i^H} \int_{Y_i^L}^{Y_i^H} \frac{|Y_j - Y_k|}{2E_i(Y)} f_i(Y_j) f_i(Y_k) \mathrm{d}Y_j \mathrm{d}Y_k,$$

where $E_i(Y)$ is individual i's subjective expectation of mean income.

The Gini coefficient is here defined as the expected relative difference between two randomly drawn members from individual i's reference group which is a well-known formula in the literature (Yitzhaki & Schechtman 2013, p.13).

3.2 Reference groups

So far we have assumed general subjective functions $f_i(Y)$ and $F_i(Y)$. In this form the framework is highly unspecific. By choosing appropriate assumptions about the formation of reference groups and the associated subjective distribution functions it would be possible to explain almost any constellation of subjective assessments and perceptions. This element of arbitrariness could be considerably reduced if one had empirical data

[6] In a previous version of the paper we have also discussed the use of a second inequality measure: the subjective relative mean absolute difference. The implications of this alternative measure are similar and for the sake of brevity we do not report them here.

on individual reference groups. Unfortunately, our dataset does not contain this kind of detailed information. We therefore choose a second remedy against the arbitrariness and discipline ourselves in the assumptions concerning reference groups. In particular, we will assume that individuals form their reference groups according to the same rule. First, we will only consider situations where the incomes classes are viewed by all people as either EquGap or RelMedian. Second, we assume that the subjective income distribution functions are constructed in the same manner for all individuals. This means, for example, that all individuals have an identical perception span or that the share of people of the own class in the reference group is the same for everybody.

In the next two subsections (and in supplement S.1) we present a number of examples for such "rule-based" reference groups. We will look at the implications of these assumptions for various measures of individual perceptions of the income distribution that can be later compared with the empirical data. In particular, we are interested in the implications for four magnitudes: (i) how individuals at different positions in the objective income distribution $F(Y_i)$ view the shape of society, (ii) how they assess their own *subjective* position $F_i(Y_i)$ in society, (iii) how they estimate mean income $E_i(Y)$ and (iv) how they evaluate the degree of inequality in society at large (captured by the subjective Gini coefficient G_i). In the benchmark case of unbiased perceptions (i.e. where $f_i(Y_j) = f(Y_j), \forall i, j,$) one would predict that all people view the shape of society in the same manner (corresponding to the true shape), that the subjective rank is equal to the objective rank (i.e. $F_i(Y_i) = F(Y_i)$), that $E_i(Y) = \overline{Y}$ and that all have the same assessment of the Gini coefficient ($G_i = G$).

3.2.1 Limited perception span

As a first assumption concerning the subjective distribution of income we look at the case where people only consider the incomes of other people that are in the vicinity of their own position $F(Y_i)$ in the income distribution. In particular, we assume that they only look p_L percentiles downwards and p_H percentiles upwards. We define this as follows:

Assumption 1 (Limited Perception Span)
An individual with income Y_i and an objective income position $F(Y_i)$ will only consider people within the percentiles $Max(0, F(Y_i) - p_L)$ and $Min(1, F(Y_i) + p_H)$ where p_L and p_H are the downward and upward perception spans, respectively. The corresponding minimum and maximum income perception levels are given by $Y_i^L = F^{-1}(Max(0, F(Y_i) - p_L))$ and $Y_i^H = F^{-1}(Min(1, F(Y_i) + p_H))$.

For $p_L = p_H = 1$ individuals have unbiased perceptions while for $p_L = p_H = 0$ they will only look at members of the exact same income level and will thus view society as perfectly equal. For in-between values of the perception span individual reference groups will differ from each other. The subjective income density function $f_i(Y)$ corresponding to assumption 1 can be related to the objective density function $f(Y)$ in the following manner: $f_i(Y) = \frac{f(Y)}{F(Y_i^H) - F(Y_i^L)}$ for $Y \in [Y_i^L, Y_i^H]$ and 0 elsewhere.

We illustrate the implications of this assumption again for the case of a lognormal income distribution with a mean of $\overline{Y} = 50\,000$ and a Gini coefficient of 0.3. In figure 3 we show the segments of the objective income distributions that are "visible" for individuals at six specific percentiles of the income distribution if $p_L = p_H = 0.3$. As one can see, the members of the reference groups change with the position in the income distribution. For the individual at position $F(Y_i) = 0.5$, i.e. the individual with median income $Y_i = 43\,101$, the boundaries of the perception span are given by $Y_i^L = F^{-1}(0.2) = 27\,247$ and $Y_i^H = F^{-1}(0.8) = 68\,181$. The rest of the income distribution is—per assumption—invisible to individual i.

We can furthermore make the assumption of EquGap, i.e. individual i distinguishes seven income classes that are characterized by identical gaps. For the case of the median income individual the size of the gap is 5 848 and the seven income classes are thus given by $[27\,247, 33\,095]$, ... up to $[62\,335, 68\,181]$.[7]

The shape of the lognormal distribution together with the assumption EquGap and the assumption about the perception spans $p_L = p_H = 0.3$ also determine the answers to the type-of-society question. As shown in figure 4, again for six income percentiles, the shape changes for different individuals. It turns from a vase (for the first decile) to a diamond (for the third decile), to a pyramid (for the middle segments) and almost to a tower (for the ninth decile). The shape of the six diagrams shows little resemblance to the objective distribution in figure 2a.

The different shapes follow directly from the heterogeneous reference groups as depicted in figure 3. The individual in the first decile primarily looks "up-hills", i.e. the bulk of his reference group is located in the increasing part of the function. From his position class sizes increase with increasing incomes. For $F(Y_i) = 0.2$ this perspective is partly reversed since this person already looks beyond the mode of the density function. From the median person onwards the reference groups are only located along the downward-sloping part of the curve and for these individuals the class size appears to be

[7] Note that the RelMedian assumption does not make much sense here since specific classes will be empty for many individuals.

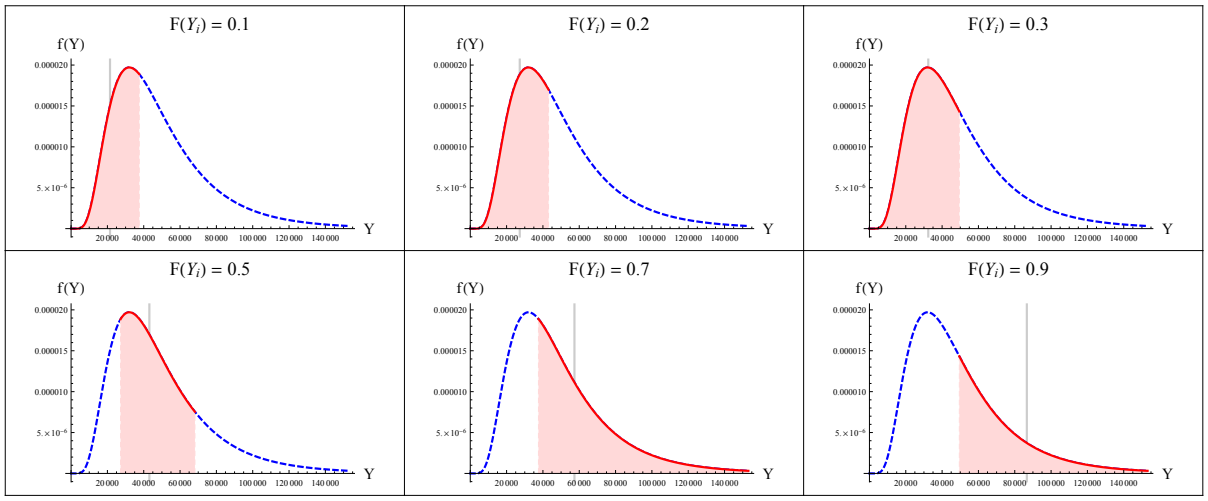


Figure 3: The pictures show the objective density function $f(Y)$ and the perception span (filled area) for six percentiles of the (objective) income distribution. The light lines show the own income. The objective income distribution is assumed to be lognormal with a mean income of 50 000 and a Gini coefficient of 0.3.

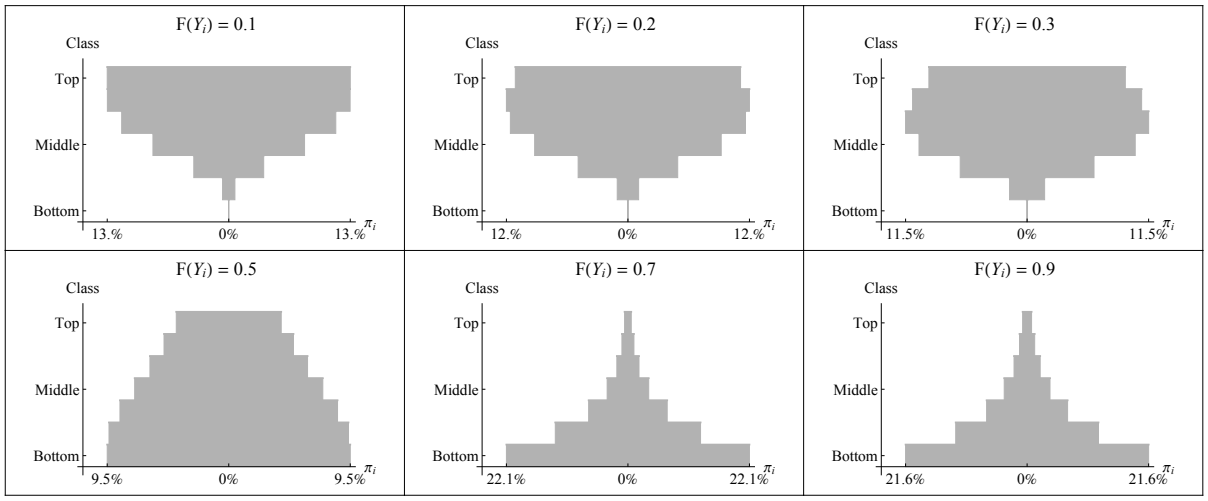


Figure 4: The pictures show the shape of society (the subjective shares π_i of the seven classes) for individuals at six deciles. It is assumed that the perception span is given by $p_L = p_H = 0.3$ and that individuals demarcate the seven classes by using assumption EquGap (equal income differences between classes). The objective income distribution is as described in figure 3.

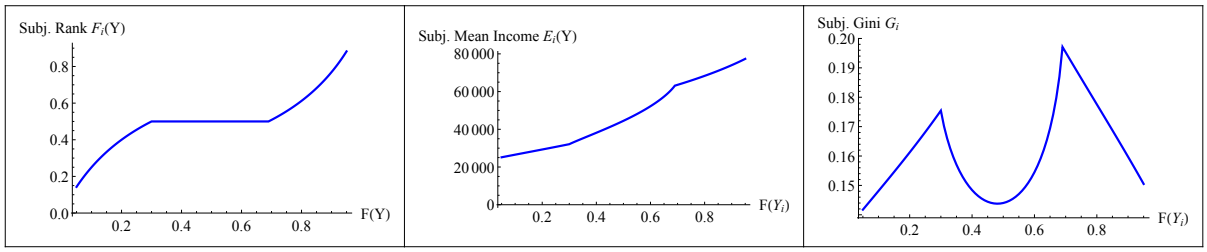


Figure 5: The pictures show three measures of distributional perceptions as stated in definitions 3 to 5. It is assumed that the perception span is given by $p_L = p_H = 0.3$.

shrinking with increasing income.

In figure 5 we show the implications of this reference group assumption for three other measures of distributional perceptions. The subjective rank $F_i(Y_i)$ shows a characteristic step function. Low-income people overestimate and high-income people underestimation their true rank. A large part of the population sees itself as being exactly in the middle which is a direct consequence of assumption 1 with a symmetric perception span $p_L = p_H$. The expectations of mean income $E_i(Y)$ increase in individual income which follows from the fact that members of the reference groups earn more with increasing own income.

The inequality measure G_i shows a non-monotonous relationship to the objective rank $F(Y_i)$, being largest for people at the percentiles p_L and $1 - p_H$. Note that all individuals in this example *underestimate* the true Gini coefficient which is given by 0.3 even though some of them view society in a rather unequal, almost "tower-like" manner (see figure 4). This reinforces the argument that we have laid out in section 2.1 that it is not meaningful to derive unequivocal subjective inequality measures from the type-of-society questions.[8]

3.2.2 Self-centered density function

Assumption 1 is rather peculiar since it assumes that individuals have narrow reference groups and they give zero weight to income levels that fall outside the boundaries of their perception span. An alternative, and arguably more plausible, approach assumes that individuals take all income levels into consideration but that they use a biased weighting function that puts more emphasis on income levels that are close to their own income (and which are therefore part of their immediate reference group). We make this more precise in the following specification.

Assumption 2 (Self-Centered Density Function)

[8]The use of asymmetric perceptions ($p_L \neq p_H$) leads to qualitatively identical results.

An individual with income Y_i and an objective income position $F(Y_i)$ has a subjective density function $f_i(Y)$ with a support over the entire range of incomes that has the mode at the own income level.

There are many possible density functions $f_i(Y)$ that are in line with assumption 2. In supplement S.1 we discuss, e.g., the case of a triangular density function with a minimum at $Y = 0$, a maximum at the $99th$ percentile ($F^{-1}(0.99)$) and the mode at $Y = Y_i$. As we illustrate there, a problem with this specific density function is that for many income levels the subjective density at the own income level is less than the objective value. This is not in line with the presumption that reference groups members are bundled around the own income level and that the density at $f_i(Y_i)$ will be larger (or at least not smaller) than the objective density $f(Y_i)$. We have experimented with a number of possible self-centered density functions that allow for such a constellation. In this section we use the most straightforward approach and assume that the subjective density function is lognormal (as is the objective density function).[9] The precise shape of the subjective density function depends on the individual position in the objective distribution function. For each individual i we choose the two parameters of the lognormal distribution in such a fashion that the mode of the function is at income Y_i (in line with assumption 2) and that, furthermore, the density at the mode is the same as the mode of the objective income density function.[10]

In figure 6 we illustrate the shape of the subjective density function together with the objective income density function for six specific income deciles. There appears a changing pattern of downward- and upward-biases that is just a consequence of assumption 2. Individuals at the lower end of the distribution tend to overweight low income levels while for the individual at the third decile the objective and subjective functions basically overlap (which is due to the fact that the mode of our lognormal income distribution is almost exactly at this position). Individuals at the upper end put a much larger weight on high incomes and their subjective density functions deviate considerably from the objective ones. In fact, one could argue that these subjective functions look implausible and that one should modify the reference group assumption (e.g. by using mixes of

[9] In supplement S.1 we illustrate the results if we use instead a two-parametric Weibull distribution.

[10] This is implemented in the following manner. For individual i the subjective probability density function is of the lognormal type and given by: $f_i(Y;\mu_i,\sigma_i) = \ln\mathcal{N}(Y;\mu_i,\sigma_i) = \frac{1}{Y\sigma_i\sqrt{2\pi}}\exp\left(-\frac{(\ln Y-\mu_i)^2}{2\sigma_i^2}\right)$ for $Y \geq 0$ and 0 for $Y < 0$, where μ_i and σ_i are the location and scale parameter, respectively. These parameters are determined by simultaneously solving two equations. First, $Y_i = e^{\mu_i-\sigma_i^2} \equiv \text{Mo}_i$ (i.e. the mode of the function $f_i(\cdot)$ is at Y_i). Second, $f_i(Y_i) = f_i(\text{Mo}_i) = f(\text{Mo}(Y))$ (i.e. the density at this mode is equal to the density of the mode of the objective density function).

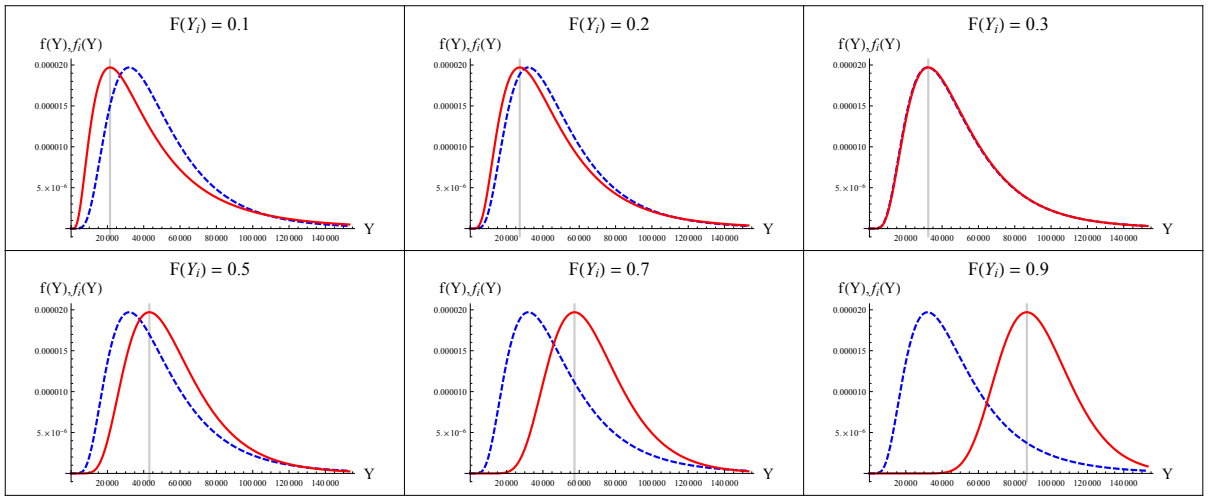


Figure 6: The pictures shows the objective density function $f(Y)$ (dashed) and the subjective density functions $f_i(Y)$ (solid) for six percentiles of the (objective) income distribution, where the vertical lines indicate the corresponding individual incomes. The subjective density function is assumed to be lognormal and constructed as described in the text. The objective income distribution is assumed to be lognormal with a mean income of 50 000 and a Gini coefficient of 0.3.

objective and subjective densities) in order to come up with less extreme perception biases. We do not follow this strategy here since the examples in sections 3.2.1 and 3.2.2 are primarily intended to illustrate the role of reference groups and the importance of specific assumptions rather than to exactly match patterns in the data.

We again assume that individuals distinguish the seven income classes by using the EquGap assumption.[11] This determines the answers to the type-of-society question as shown in figure 7 for six income percentiles. As was the case for the assumption with perception spans (see figure 4) the shape changes for different individuals. The order of the metamorphosis, however, is now almost the exact opposite to the previous case. It turns from a tower or pyramid (for the first deciles) to a diamond (for the middle deciles) and to something that vaguely resembles a "compressed diamond" or vase (for the upper deciles). Comparing the six diagrams to the shape of the objective distribution given in figure 2a one could say that the image of the individuals between the third and fifth decile are more or less in line with the objective shape. Each interviewee, however, would again has to mentally squeeze the own image into one of the five shapes provided in the type-of-society question.

[11] The results for using the RelMedian assumption are shown in supplement S.1.

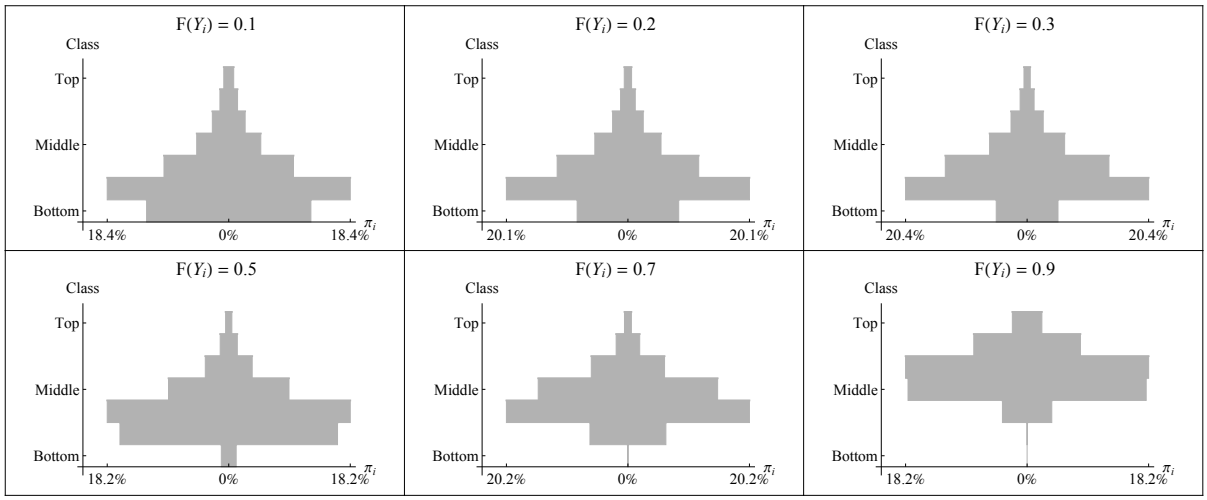


Figure 7: The pictures show the shape of society (the subjective shares π_i of the seven classes) for individuals at six deciles. It is assumed that the subjective density functions $f_i(Y)$ are lognormal as illustrated in figure 6 and that individuals demarcate the seven classes by using the EquGap assumption.

In figure 8 we show the implications of assumption 2 for the three different measures of distributional perceptions. The subjective rank $F_i(Y_i)$ increases in the objective rank $F(Y_i)$ and we can again observe a considerable degree of overestimation (underestimation) of low-income (high-income) people. Related to this, also the expectations of mean income $E_i(Y)$ increase in income. The qualitative pattern of these curves is similar to the perception span assumption presented in figure 5 although the range of subjective ranks is more narrow in this case.

For the inequality measure G_i we get different results than for the perception span assumption. While in figure 5 we could observe a non-monotonous relationship now we get a decreasing pattern. High-income individuals perceive a more equal society than low-income individuals. For the lowest deciles it is now even the case that they would overestimate the true Gini coefficient of $G = 0.3$ although most people will again tend to underestimate the true extent of income inequality.

3.3 Summary

The framework of inequality perceptions that is based on heterogeneous, self-centered reference groups and subjective income distributions has a number of implications that can be tested empirically. Some of the implications are valid for both specifications of reference groups while for others one has to distinguish between the limited-perception-

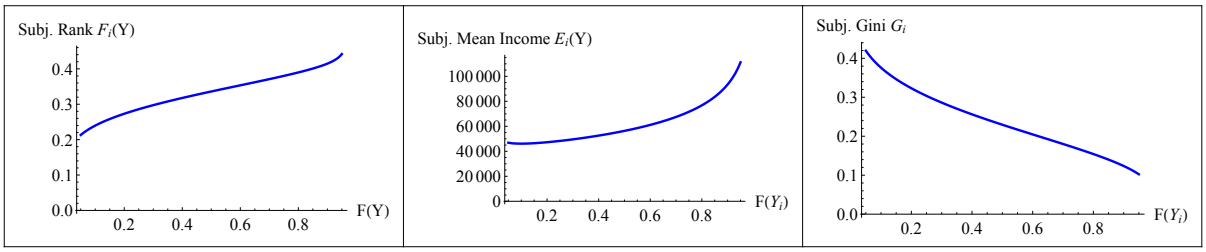


Figure 8: The pictures show three measures of distributional perceptions as stated in definitions 3 to 5. It is assumed that the subjective density functions $f_i(Y)$ are lognormal as illustrated in figure 6.

span assumption (section 3.2.1) and the self-centered-density assumption (section 3.2.2)

- **Implication 1: Subjective Rank**. For both assumption about the subjective income distribution the framework predicts an overestimation of the own income rank by low-income people and an underestimation by high-income individuals. The model with a self-centered density function is associated with a particularly compressed gradient.

- **Implication 2: Subjective Type of Society**. Both assumptions about the subjective income distribution imply that people from different income deciles will view the shape of society differently. For the limited-perception-span framework the order runs from a vase (type E) for lower deciles to a tower (type A) for the top percentiles. The self-centered density framework predicts an inverse ordering where low-income individuals view society as a tower or pyramid while the top end sees it as a vase.

- **Implication 3: Subjective Mean Income**. For both assumptions the framework predicts that the predicted mean income of society increases with the rank in the income distribution.

- **Implication 4: Subjective Income Inequality**. The limited-perception-span framework predicts a non-monotonous relation between the income percentile and the perception of inequality as measured by the subjective Gini coefficient. For the self-centered density assumption, on the other hand, the implied pattern is a downward-sloping line: higher income classes perceive less inequality than poorer households.

4 Empirical analysis

In order to test the aforementioned implications we employ data from the International Social Survey Programme's "Social Inequality" module of 2009 (ISSP Research Group 2012). The ISSP has been conducted in 40 countries and provides empirical measures for the main perception variables emphasized by the theoretical framework. The data are described in appendix A and table A.1 summarizes descriptive statistics of all variables.

In the following we test the four implications highlighted in section 3.3.

Implication 1: The subjective ranks are biased. To test implication 1, we compute the rank of the income distribution (derived from household income) for each respondent (*HH income objective rank*). This is done for each country separately. Figure 9 plots the *Objective rank* against the *Subjective rank* (based on the social-ladder question) together with a fitted polynomial line. The picture indicates that respondents at the lower end of the income distribution overestimate their rank while those at the top-end underestimate their rank. Furthermore, the shape of the line suggests a non-linear relationship between the two variables. Both the curvature and the compressed range of subjective rank estimation is closely in line with the theoretical predictions documented in figure 8. These results imply that the answers to the social-ladder question corresponds better with the assumption of a self-centered density function than with the assumption of a limited perception span. While the line refers to all respondents from all countries, we also have depicted selected countries that differ by their income level and their societal organization. In supplement S.2 we analyze some country sub-aggregates. We find that the slope of the relationship is varying as does its vertical position, nevertheless, the basic pattern that low income people overestimate and that high income people underestimate their position remains unchanged.

Implication 2: The subjective types of society depend on individuals' income rank. According to the theoretical results (see figures 4 and 8), the perceived type of society depends on the rank of the individual in the income distribution. To test implication 2 we will always look at both the objective and the subjective rank. This is done for two reasons. First, income measures elicited from surveys are known to be problematic for reasons of item non-response and measurement error. The use of an adjusted measure, like *Subjective rank*, can help to attenuate these problems. Second, the subjective rank is broader in scope and may include respondents' considerations regarding

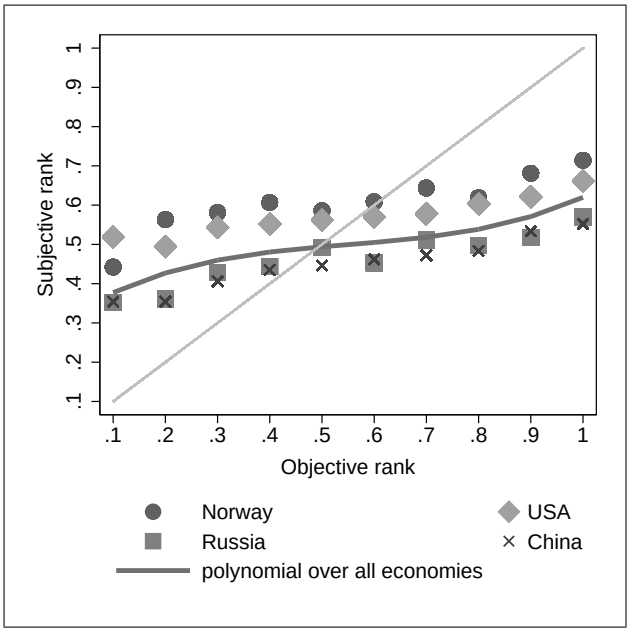


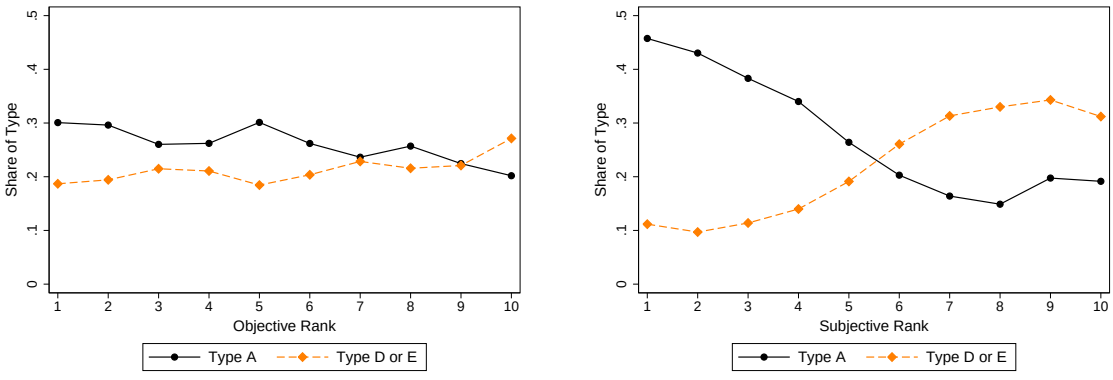
Figure 9: The figure shows country averages per *Subjective rank* in dots and a polynomial line fitted to the raw data. Variables are described in appendix A.

wealth, education, health and social background. Given that the countries in the sample differ substantially with respect to these variables, it may well be that *Subjective rank* better captures an individual's social position than the objective rank.

As a first look, figure 10 plots for each decile of the objective and subjective income distribution the average share of respondents with perceived type of society A and with perceived type of societies D and E.[12] The descriptive account shows that the share of respodents who perceive type A declines with the subjective income rank. In turn, the share of respodents who perceive type D/E is increasing in the subjective income rank. Although the survey results display the same pattern for the objective income rank, the correlations are considerably less strong.

To provide a more formal test of implication 2 we run individual-level estimations with the dependent variable *Perceived type of society* that ranges from 0 to 1. The regressions include household level controls for age, the employment status, the household size, the marital status, education as well as one variable for the interview mode. Several other household controls could be considered, including the political orientation, the size of the municipality, etc., however this comes at the cost of loosing countries because some of

[12]We have merged answers D and E into one category because the share of respondents who answer type E is very small in all countries. This, however, is inconsequential for all remaining findings.



(a) Objective Income

(b) Subjective Income

Figure 10: The picture shows the relative sample frequencies of society types for deciles of individuals' objective and subjective income.

those variables are either not available or cannot be defined in a harmonized way. In light of these obstacles we have decided to leave these variables out but run robustness regressions, some of them will be presented below, and results are not affected qualitatively by the inclusion of these additional variables. One potentially important variable which is particularly difficult to harmonize is education. Instead of a direct measure of education we use the number of books that were around when respondents' had an age between 14 and 16. In country-specific regressions, this variable is found to be a powerful predictor of respondents' degree of education. Finally, all regressions contain country fixed-effects and a series of controls for the religious affiliation of respondents.

Table 1 summarizes the results from a linear probability model.[13] Column 1 shows that *Subjective rank* exerts a significant and negative effect on the perceived society type, i.e. a higher subjective rank is associated with a more equal image. We have also experimented with a quadratic term of *Subjective rank* but since it was not significantly different from zero we have omitted it from the final estimation.[14] Column 2 confirms that the objective HH income rank exerts a (non-linear) effect on the dependent variable. The point estimates imply a monotonically decreasing effect. When *Subjective rank* and *HH income objective rank* are included jointly, both types of variables are significant (column

[13]In appendix B we document that results do not change qualitatively if an ordered probit model is used instead of the linear probability model.

[14]We follow this convention for the later regressions. We always allow for a linear and a quadratic effect for subjective and objective rank in the first specification but leave it out if it turns out to be statistically insignificant.

3). The quantitative impact is stronger for the subjective rank than for the objective rank, as suggested in the descriptive analysis of figure 10. Moving from the lowest to the highest percentile reduces the index function (which ranges from 0 to 1) by 0.22 (column 1) for *Subjective rank* and by 0.07 (column 2) for *HH income objective rank*.

The significant coefficient of the rank variables in table 1 is in line with the predictions of the theoretical framework. What is more, the empirical results help to distinguish between the two assumptions about the subjective distribution function. In particular, the results in table 1 indicate that as the rank increases, the perceived shape of society goes from a tower shape (type A) to a diamond/vase (type D/E). This provides support for the self-centered subjective density function against the limited-perception-span assumption.

An additional way to check whether non-linearities in the dependent variable affect results is to assign Gini coefficients to the various types. In table 2 we code the types according to the EquGap assumption for the seven classes as described in section 2.1. The results do not change, qualitatively.[15] Going from the lowest to the highest percentile reduces the index function (which ranges from 0.23 to 0.53) by 0.07 and by 0.02 for *Subjective rank* and for *HH income objective rank*, respectively.

Our benchmark specification contains only a limited set of household-level controls. We would like to test whether some variables which could potentially be important but for which we have no theoretical justification for their inclusion affect results. Specification 1 of Table 3 shows a specification with three additional variables and results for *Subjective rank* do not change. Columns 2 to 5 repeat estimations for sub-samples to see whether there are non-linearities in the measured effect of *Subjective rank*. Specifically, column 3 and 4 look at sub-samples of low and high educated respondents and column 5 and column 6 look at sub-samples of respondents with low and high income. We find a significant effect of *Subjective rank* which is in the range of previous results for the full sample.

Implication 3: The subjective mean incomes increase with the income rank. Table 4 tests whether the rank in the distribution leads to significantly different estimations of the income of various professions. The variable *Ln Relative earnings* measures the logarithm of the ratio of estimated earnings of a specific professions to the country average of this profession. We look at the following professions: shopkeepers, unskilled factory workers, doctors in general practice and chairmen of large national corporations. For each profession, we show a specification with only *Subjective rank* and a specification

[15]Results hold if we assign different Gini coefficients to the five societies based on the alternative assumptions described in appendix C.

Table 1: Perceived type of society and income rank

Dependent variable	Perceived type of society (Type A=1, Type B=0.66, Type C=.33, Type D/E=0)		
	(1)	(2)	(3)
Subjective rank	-0.253***		-0.251***
	(0.030)		(0.025)
HH income objective rank		0.016	0.048
		(0.035)	(0.035)
HH income objective rank squared		-0.085***	-0.068***
		(0.024)	(0.024)
Ln Number books in childhood	-0.006***	-0.008***	-0.004**
	(0.002)	(0.002)	(0.002)
Unemployed	0.012	0.020**	0.011
	(0.008)	(0.009)	(0.009)
Not in labour force	-0.005	-0.005	-0.007
	(0.005)	(0.006)	(0.006)
Age	0.002	0.002*	0.002
	(0.001)	(0.001)	(0.001)
Age squared	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
HH size 1	0.008	-0.003	0.006
	(0.008)	(0.013)	(0.011)
HH size 2	0.012	0.010	0.013
	(0.008)	(0.010)	(0.010)
HH size 3,4	0.010*	0.012	0.014*
	(0.006)	(0.007)	(0.007)
Separated	0.014**	0.015*	0.013
	(0.007)	(0.008)	(0.008)
Widowed	-0.000	0.002	0.000
	(0.011)	(0.013)	(0.013)
Married	0.005	0.005	0.008
	(0.006)	(0.006)	(0.006)
Face-to-face interview	0.041***	0.022***	0.011***
	(0.005)	(0.003)	(0.004)
Constant	0.771***	0.690***	0.790***
	(0.030)	(0.031)	(0.035)
Country fixed-effects	yes	yes	yes
Religion controls	yes	yes	yes
Adj. R-squared	0.25	0.24	0.25
Observations	35844	30659	30487

The dependent variable is *Perceived type of society* coded as 0 (Type D/E), 0.33 (Type C), 0.66 (Type B) and 1 (Type A). All models report estimates from a linear probability model and include country-fixed effects and fixed-effects for the religion of respondents. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level. Variables are defined in appendix A.

Table 2: Implied Gini coefficient and income rank

Dependent variable	Implied Gini coefficient (Type A=0.53, Type B=0.43, (Type C=.35, Type D/E=0.23)		
	(1)	(2)	(3)
Subjective rank	-0.073***		-0.072***
	(0.009)		(0.007)
HH income objective rank		0.006	0.015
		(0.010)	(0.010)
HH income objective rank squared		-0.025***	-0.020***
		(0.007)	(0.007)
Household controls	yes	yes	yes
Country fixed-effects	yes	yes	yes
Religion controls	yes	yes	yes
Adj. R-squared	0.24	0.24	0.25
Observations	35844	30659	30487

The dependent variable is the implied Gini coefficient that follow from *Perceived type of society* when using the EquGap assumption together with income levels $\{0.5, 1.5., 2.5, 3.5, 4.5, 5.5, 6.5\}$ for the seven classes as described in section 2.1. As shown there, type A (D) would imply a Gini coefficient of 0.53 (0.23). All models report estimates from a linear probability model and include country-fixed effects and fixed-effects for the religion of respondents. All models include the same household controls as in table 1. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level. Variables are defined in appendix A.

Table 3: Perceived type of society and income rank — Different subsamples

Dependent variable	*Perceived type of society* (Type A=1, Type B=0.66, Type C=.33, Type D/E=0)				
	All	Less than 50 books	More than 50 books	HH income decile less or equal than 5	HH income decile larger than 5
	(1)	(2)	(3)	(4)	(5)
Subjective rank	-0.229***	-0.200***	-0.288***	-0.239***	-0.212***
	(0.028)	(0.029)	(0.027)	(0.031)	(0.029)
Voted last election	-0.008	-0.009	-0.005	0.004	-0.015**
	(0.007)	(0.008)	(0.008)	(0.010)	(0.007)
Work important to get ahead	-0.008	-0.003	-0.011	-0.006	-0.008
	(0.005)	(0.006)	(0.009)	(0.008)	(0.007)
Preference for redistribution	0.066***	0.055***	0.078***	0.068***	0.059***
	(0.008)	(0.007)	(0.010)	(0.010)	(0.008)
Country fixed-effects	yes	yes	yes	yes	yes
Religion controls	yes	yes	yes	yes	yes
Adj. R-squared	0.26	0.20	0.34	0.21	0.29
Observations	32396	20911	11485	11785	20611

The dependent variable is *Perceived type of society* coded as 0 (Type D/E), 0.33 (Type C), 0.66 (Type B) and 1 (Type A). All models report estimates from a linear probability model as in column 1 of table 1 and include country-fixed effects and fixed-effects for the religion of respondents and the same household controls as in table 1. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level. Variables are defined in appendix A.

with only *HH income objective rank*. The control variables are the same as in the previous tables.

Table 4: Subjective earnings and income rank

Dependent variable	*Ln Relative earnings*							
	shopkeeper		unskilled		doctor		chairman	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Subjective rank	0.130***		0.159***		0.123**		-0.017	
	(0.038)		(0.036)		(0.051)		(0.105)	
HH income objective rank		0.244***		0.246***		0.315***		0.444***
		(0.034)		(0.041)		(0.067)		(0.075)
Household controls	yes	yes	yes	yes	yes	yes	yes	yes
Country fixed-effects	yes	yes	yes	yes	yes	yes	yes	yes
Adj. R-squared	0.08	0.10	0.04	0.06	0.05	0.07	0.20	0.21
Observations	34255	29515	34347	29614	33119	28704	32232	28066

The dependent variable is the log of *Relative earnings* for shopkeeper (cols 1, 2), unskilled factory workers (cols 3, 4), for doctors in general practice (cols 5, 6) and for a chairman of a large national corporation (cols 7, 8). For calculating the dependent variable we disregard the lowest and the highest 1 percent of earnings estimates per occupation and per country and express the answer *relative* to the country average. All models report estimates from a linear probability model and include country-fixed effects and fixed-effects for the religion of respondents. All models include the same household controls as in table 1. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level. Variables are defined in appendix A.

The results show that respondents' estimates of earnings increase with either the subjective or the objective rank. The estimates in column 1 suggest, for example, that persons who consider themselves in the top percentile of society assess the income of shopkeepers to be 13 percent higher than persons who consider themselves in the bottom percentile of society. For the objective rank the effect is about 24 percent. The magnitude of the effect is somewhat stronger for the other three professions. We note that jointly including the subjective rank and the objective rank leads to an insignificant subjective rank whereas the objective rank remains significant. This holds for the first three professions shown in Table 4. For the last profession, chairmen, the subjective rank enters negatively but the overall effect (subjective and objective rank) is still positive.

The positive effect of the income position on the estimated income levels is in line with the prediction of the theoretical framework (see figure 5)—with the caveat that the framework refers to mean income while for the empirical results we can only use the limited number of profession to proxy for individuals' assessment of the aggregate variable.

Osberg & Bechert (2016) and Kuhn (2015) propose two variants to deduce a subjective estimate of mean earnings from answers to the perceived income of the five professions. As both approaches require strong assumptions (see the discussion in the next paragraph), we do not show the corresponding results but note that the objective rank enters positively

and significantly if these measures are employed as dependent variables which corroborates the results of Table 4.

Implication 4: The subjective extent of inequality decreases with the income rank. For implications 1 to 3 the ISSP data provide a (more or less) direct survey measure. For the fourth implication such a measure is not available. We approach this issue from two angles. First, we employ respondents' subjective estimates of the Gini coefficient as proposed in Osberg & Bechert (2016) and Kuhn (2015), called GiniOB and GiniK in the following. Both approaches employ the earnings questions to derive individual specific subjective estimates of the Gini coefficient.[16]

The first four columns of Table 5 summarize respective results. The structure of the table and the control variables mimics those of the previous tables. With the exception of the first column, the results reveal a significant effect of the subjective rank. The point estimate of *Subjective rank* in column 4, for example, implies that a person in the top of the subjective income ladder estimates the Gini to be 4.2 percentage points lower than a person in the bottom of the subjective income ladder (the sample mean of GiniOB is 0.47, the sample mean of GiniK is 0.50). For the objective rank, the result is not unambiguous: In columns 2 the overall effect is slightly positive and in column 4, the effect is statistically not different from zero.

While the use of GiniOB or GiniK provides interesting insights, neither of these two measures is without problems (which is admitted in the respective papers). For example, GiniK assumes that each individual has the same view about the share of the bottom income group and that this share can be estimated from survey information about respondents' occupations. In practice this results in very different estimates of the size of the bottom income group and of a countries' average income level for otherwise rather similar economies. The GiniOB, on the other hand, assumes that all professions have an equal share in the economy. Both measures potentially suffer from a wide variability induced by respondents' estimates of the salary of a chairman of a large national corporation. In light of these issues, we propose to use a variable that indicates whether income differences are considered too large. Under the assumption that the "ethically optimal" subjective

[16]The construction of these variables is described in the appendix. In a nutshell, Kuhn (2015) uses the spread between "bottom" incomes and "top" incomes along with the shares of the population that fall in each category (varying across countries but assumed to be fixed across individuals in a given country). Osberg & Bechert (2016) calculate the Gini from respondents estimates of the wage of four professions (assuming that the economy consist of only four individuals — one unskilled, one shop assistant, one doctor and one chairman).

Table 5: Subjective Gini — Income differences too large

Dependent variable	GiniOB		GiniK		Income differences too large	
	(1)	(2)	(3)	(4)	(5)	(6)
Subjective rank	-0.021	-0.043***	-0.046***	-0.046***	-0.226***	-0.198***
	(0.013)	(0.015)	(0.015)	(0.015)	(0.037)	(0.033)
HH income objective rank		0.023		-0.007		0.174***
		(0.016)		(0.015)		(0.049)
HH income objective rank squared		0.012		0.001		-0.215***
		(0.012)		(0.015)		(0.057)
Country fixed-effects	yes	yes	yes	yes	yes	yes
Religion controls	yes	yes	yes	yes	yes	yes
Adj. R-squared	0.30	0.32	0.43	0.44	0.13	0.13
Observations	31367	27217	31058	26865	37023	31331

In columns 1 and 2 the dependent variable is an individuals specific estimate of the Gini following Osberg & Bechert (2016). In columns 3 and 4 the dependent variable is an individuals specific estimate of the Gini following Kuhn (2015). For each of these dependent variables we disregard the lowest and the highest 1 percent of responses per country. In columns 5 and 6 the dependent variable is the dummy variable *Income differences too large*. All models report estimates from a linear probability model and include country-fixed effects and fixed-effects for the religion of respondents. All models include the same household controls as in table 1. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level. Variables are defined in appendix A.

degree of inequality is randomly distributed over the population, which admittedly is also strong, this variable can be considered as a proxy for the individual-level subjective Gini.[17] The point estimates of *Subjective rank* in columns five and six reveal that the likelihood that income differences are viewed as too large decreases with the subjective rank in the income distribution. In column 6, the objective rank and its squared term are significant but exert only a minor quantitative effect compared to *Subjective rank*.

Given that we do not have direct survey information on the subjective Gini of individuals and that we have to rely on proxy variables, we take from these regressions that results are by and large in line with the predictions of the theoretical framework based on the use of subjective self-centered density functions (as illustrated in the lower left panel of figure 8).

Robustness: We have discussed that our results are qualitatively unaffected if we use an ordered probit model instead of a linear probability model. Moreover, results are robust to various subsamples and to different specifications. In supplement S.2 we conduct

[17]This assumption is difficult to validate. One possibility is to calculate an ethical Gini in the line of Osberg & Bechert (2016) and Kuhn (2015), by using answers on how much typical professions should earn. Once controlling for country fixed effects and individual-level control variables, we find that the ethical Gini according to Osberg & Bechert (2016), for example, essentially does not depend on the objective or the subjective income rank, which provides at least some support for the assumption.

additional robustness tests. Given the vast differences of countries in our sample—with respect to dimensions like economic development, institutions, the existence of a welfare state—we take a closer look at subgroups of countries and at individual countries to make sure that results are not driven by single (influential) countries. For example, we repeat estimations for selected specifications country-by-country thus obtaining 40 point estimates which we graphically plot against their t-values. Overall, these robustness tests provide strong support for the implications of our framework. Results do not seem to be driven by influential observations, the patterns of results can be found in various economic or regional sub-groups of countries and the per-country regressions yield the predicted effects for the majority of economies.

5 Conclusions

In this paper, we argue that people perceive inequality of the income distribution by looking at reference groups that are not arbitrary subsamples of the society but are typically biased towards individuals that have a similar social and economic background. Individuals base their assessment of the distribution on a subjective income distribution formed from this subsample of the population. This mechanism generates three implications. First, perceptions can deviate substantially from objective reality. Second, perception biases follow a systematic pattern, i.e., their size and direction depends on an indivual's position in the income distribution. Third, perceptions of different dimensions of inequality are related to each other.

We confront these implications with survey data from 40 countries and find strong and robust support for this framework. Persons with low income: (i) underestimate their true position in society while persons with higher income overestimate their true rank; (ii) see the income distribution as a "tower" or "pyramid" while persons with higher income see it as a "vase"; (iii) have a lower estimate of average earnings than persons with higher income; (iv) have a higher estimate of the Gini coefficient than persons with higher income. These are novel results, first, because the previous literature has not focused on the relation between the objective income rank and perceptions and, second, because it has treated the different perception measures in isolation. With regard to the specific form of reference group formations, the empirical findings are in line with the framework in which people do consider the entire income distribution but place too much weight on incomes close to their own position (self-centered perception biases).

We think that these results reveal an important mechanism that is behind the existence of perception biases. We do not claim, however, that the presence of heterogeneous reference groups is the only channel of influence. Certainly there are other factors that are likely to play an important role, in particular the influence of media consumption and the presence of specific psychological traits. On the one hand, media consumption could potentially be a powerful transmitting factor of "fake stories" and "alternative facts" thereby causing perception biases. On the other hand, one could revert to psychological explanations and argue, e.g., that there is a conformity bias which induces people to associate with the middle income group. Without doubt these additional factors are likely to have an impact on inequality perceptions and more research (and data) are needed to quantify their importance. It is, however, not immediately obvious how and whether they will be able to fully explain the within- and across-country pattern in perceptions found in our data. For example, the existence of a conformity bias is likely to play a role in respondents' choice of middle categories to the social-ladder question. At the same time this bias does not explain why people have different estimates of average earnings or why they form different perceptions about the distributional shape of incomes.

Finally, we want to emphasize that there are a number of open issues that should be dealt with in future research. In particular, this concerns survey information on reference group formation. We have stipulated the existence of reference groups and shown indirectly that they are important. Direct survey information about reference groups and whether there are heterogeneities across people would greatly foster our understanding of how perception biases relate to reality.

References

Alesina, A. & La Ferrara, E. (2005), 'Preferences for Redistribution in the Land of Opportunities', *Journal of Public Economics* **89**(5), 897–931.

Clark, A. E. & D'Ambrosio, C. (2015), Chapter 13 - Attitudes to Income Inequality: Experimental and Survey Evidence, *in* A. B. Atkinson & F. Bourguignon, eds, 'Handbook of Income Distribution', Vol. 2 of *Handbook of Income Distribution*, Elsevier, pp. 1147 – 1208.

Clark, A. E. & Senik, C. (2010), 'Who Compares to Whom? The Anatomy of Income Comparisons in Europe', *The Economic Journal* **120**(544), 573–594.

Clark, A., Senik, C. & Yamada, K. (2013), The Joneses in Japan: Income Comparisons and Financial Satisfaction, ISER Discussion Paper 0866, Institute of Social and Economic Research, Osaka University.

Cruces, G., Perez-Truglia, R. & Tetaz, M. (2013), 'Biased Perceptions of Income Distribution and Preferences for Redistribution: Evidence from a Survey Experiment', *Journal of Public Economics* **98**(C), 100–112.

Engelhardt, C. & Wagener, A. (2014), Biased Perceptions of Income Inequality and Redistribution, Hannover Economic Papers (HEP) dp-526, Leibniz Universitt Hannover, Wirtschaftswissenschaftliche Fakultt.

Gimpelson, V. & Treisman, D. (2015), Misperceiving Inequality, NBER Working Papers 21174, National Bureau of Economic Research.

ISSP Research Group (2012), International Social Survey Programme 2009: Social Inequality IV, Technical report, GESIS Data Archive, Cologne. ZA5400 Data file Version 3.0.0 (doi:10.4232/1.11506).

Knell, M. & Stix, H. (2017), Inequality, Perception Biases and Trust. Oesterreichische Nationalbank Working Paper No. 211.

Kuhn, A. (2011), 'In the Eye of the Beholder: Subjective Inequality Measures and Individuals' Assessment of Market Justice', *European Journal of Political Economy* **27**(4), 625–641.

Kuhn, A. (2015), The Individual Perception of Wage Inequality: A Measurement Framework and Some Empirical Evidence, IZA Discussion Papers 9579, Institute for the Study of Labor (IZA).

Kuziemko, I., Norton, M. I., Saez, E. & Stantcheva, S. (2015), 'How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments', *American Economic Review* **105**(4), 1478–1508.

Mayraz, G., Schupp, J. & Wagner, G. G. (2009), Life Satisfaction and Relative Income: Perceptions and Evidence, CEP Discussion Papers dp0938, Centre for Economic Performance, LSE.

Niehues, J. (2014), Subjective Perceptions of Inequality and Redistributive Preferences: An International Comparison. Discussion Paper, Cologne Institute for Economic Research.

Norton, M. I. & Ariely, D. (2011), 'Building a Better America - One Wealth Quintile at a Time', *Perspectives on Psychological Science* **6**(1), 9–12.

Osberg, L. & Bechert, I. (2016), Social Values for Equality and Preferences for StateIntervention: Is the USA "Exceptional"?, Working Paper 2016-04, Department of Economics, Dalhouse University.

Osberg, L. & Smeeding, T. (2006), '"Fair" Inequality? Attitudes Toward Pay Differentials: the United States in Comparative Perspective', *American Sociological Review* **71**(3), 450–473.

Runciman, W. G. (1966), *Relative Deprivation and Social Justice: a Study of Attitudes to Social Inequality in Twentieth-Century Britain* , Routledge & Kegan Paul, London.

Slemrod, J. (2006), 'The Role of Misconceptions in Support for Regressive Tax Reform', *National Tax Journal* **59**(1), 57–75.

Van Ourti, T. & Clarke, P. (2011), 'A Simple Correction to Remove the Bias of the Gini Coefficient Due to Grouping', *Review of Economics and Statistics* **93**(3), 982–994.

Yitzhaki, S. & Schechtman, E. (2013), More Than a Dozen Alternative Ways of Spelling Gini, *in* 'The Gini Methodology', Vol. 272 of *Springer Series in Statistics*, Springer New York, pp. 11–31.

Appendices

A Data description

The data are drawn from the International Social Survey Programme's (ISSP) "Social Inequality" module of 2009 (ISSP Research Group 2012). The ISSP has been conducted in 40 countries.

All surveys have been carried out by independent institutions in each country between February 2008 and Jan 2012, with the majority of interviews in 2009. Details about the sampling universe, the sampling design and the survey mode are summarized in ISSP Research Group (2012). The sample sizes comprises between 880 (Finland) and 3 300 (South Africa) respondents.

For the purpose of this study we will not use sampling weights (which are unavailable) and we have not imputed missing observations (e.g. for income questions). Also, we have eliminated all respondents below the age of 18 years, persons in education and persons with a monthly household income of 30 USD or less.

Aggregate information is taken from the World Bank Financial Development Database. We separate countries in "rich countries" and "poor countries" based on whether they are above or below the sample median for per capita Gross National Income (27 735 PPP-USD).

Geographical coverage: The sample covers 25 countries from Europe, 2 from Central Asia (Russia, Turkey), 7 from East Asia and Pacific (Australia, Japan, New Zealand, China, Taiwan, Korea, Philippines), 3 from Latin America (Argentina, Chile, Venezuela), 1 from Middle East (Israel), 1 from North America (USA) and 1 from Sub-Saharan Africa (South-Africa).

Dependent variables:

Perceived type of society: Based on "V54": "What type of society is < respondent's country > today - which diagram comes closest?" Categorical variable = 1 if "Type A" (A small elite at the top, very few people in the middle and the great mass of people at the bottom.), = 0.66 if "Type B" (A society like a pyramid with a small elite at the top, more people in the middle and the most at the bottom.), = 0.33 if "Type C" (A pyramid except that just a few people are at the bottom.), = 0 if "Type D" (A society with most people in the middle.), = 0 if "Type E" (Many people near the top, and only a few near the bottom.).
Based on the variable *Perceived type of society* the dummy variables *Perceived type of society A*, *Perceived type of society B*, *Perceived type of society C* and *Perceived type of society D, E* are generated which are 1 for the respective categories, 0 otherwise.

Implied Gini: The Gini that is associated with the 4 types of society if calculated as in Gimpelson & Treisman (2015). The EquGap assumption is used, with income levels $\{0.5, 1.5., 2.5, 3.5, 4.5, 5.5, 6.5\}$ for the seven classes as described in section 2.1. The implied Gini coefficients are 0.53 (type A), 0.43 (type B), 0.35 (type C) and 0.23 (type D/E).

GiniOB: Osberg & Bechert (2016) calculate the Gini from respondents estimates of the wage of four professions (unskilled factory workers, shop assistants, doctors and chairmen). The individual specific estimate assumes that the economy consist of four people, i.e. that all four professions have the same share.

GiniK: In the calculation we deviate from Kuhn (2015) in that we define the bottom income group as consiting of unskilled factory workers and of shop assistants (whereas Kuhn uses only the former). The top income group consist of doctors and ministers (and we omit chairmen). As in Kuhn (2015) the share of the bottom income group (which varies across countries but not across individuals) is taken from sample averages. Individuals for whom the estimate of *Gini Kuhn* is negative are omitted.

Income differences too large: Constructed from *V32*: "Differences in income in <respondents country> are too large". Dummy variable = 1 if "Strongly agree" or "Agree", = 0 if "Neither agree nor disagree" or "Disagree" or "Strongly disagree".

Relative earnings shopkeeper, unskilled, minister, doctor, chairman: Constructed from *V24* to *V28*: Respondents were asked to give their estimate of the earnings of the respective profession. *Relative earnings* is calculated by dividing individual-level estimates about how much the respective profession earns by the country average of all respondents. For each profession, we winsorize the lowest and highest 1% of earning per country.

Household-level control variables:

Subjective rank: Based on "V44": "In our society there are groups which tend to be towards the top and groups which tend to be towards the bottom. Below is a scale that runs from top to bottom. Where would you put yourself now on this scale?" Categorical variable = 0.1 if "Bottom, Lowest, 01", = 0.2 if "02", = 0.3 if "03", = 0.4 if "04", = 0.5 if "05", = 0.6 if "06", = 0.7 if "07", = 0.8 if "08", = 0.9 if "09", = 1 if "Top, Highest, 10".

HH income objective rank: Percentile of household income calculated from *country_inc_15*. *country_inc_15* is an income-variable that is generated out of the original country-income-variables *country_INC*.

Books in childhood: Constructed from *V61*: "About how many books were there around your familys house when you were <14-15-16> years old?" Categorical variable = 1 if "0", = 2 if "1.5", = 3 if "10", = 4 if "20", = 5 if "50", = 6 if "100", = 7 if "200", = 8 if "500", = 9 if "1000 or more". *Ln Number books in childhood* is the natural logarithm of *(books + 1)*.

Employed, Unemployed, Not in labour force: Dummy variables constructed from *WRKST* which asks the respondent for his/her current employment status. *Employed* = 1 if "Employed, full-time" or "Employed, part-time", 0 otherwise; *Unemployed* = 1 if "Unemployed", 0 otherwise; *Not in labour force* = 1 if "Housewife, -man, home duties" or "Other, not in labour force", 0 otherwise.

Age: Age of respondent in years.

HH size1, HH size 2, HH size 3/4, HH size 5+: Dummy variables constructed from *HOMPOP*: "How many persons in household (including respondent)?" *HH size 1* = 1 if "1", 0 otherwise; *HH size 2* = 2 if "2", 0 otherwise; *HH size 3/4* = 1 if "3" or "4", 0 otherwise; *HH size 5+* = 1 if "5" or more, 0 otherwise.

Separated, Widowed, Married: Dummy variables constructed from *MARITAL*: "What is your marital status?" *Separated* = 1 if "Divorced" or "Separated (married but sep./not living with legal spouse)", 0 otherwise; *Widowed* = 1 if "Widowed", 0 otherwise; *Married* = 1 if "Married", 0 otherwise.

Face-to-face interview: Constructed from *MODE*: "Administrative mode of data-collection". The variable *MODE* has 18 categories that we do not report here. Ten out of these categories are categories for face-to-face-interviews with various additional infos which are combined to the dummy variable *Face-to-face interview*.

Religion: *RELIGGRP*, the respondents' main religious groups.

Macroeconomic variables:

Gross National Income per capita in PPP USD: Source: World Bank Financial Development Database.

Gini Eurostat: Gini coefficient of equivalized disposable income. This variable is only available for European countries. Source: EU-SILC.

Table A.1: Descriptive statistics

	N	mean	sd	min	max
Dependent variables					
Perceived type of society (1/0.66/0.33/0)	38024	0.56	0.36	0.00	1.00
Perceived type of society A	38024	0.27	0.44	0.00	1.00
Perceived type of society B	38024	0.34	0.47	0.00	1.00
Perceived type of society C	38024	0.18	0.38	0.00	1.00
Perceived type of society D/E	38024	0.21	0.41	0.00	1.00
Implied Gini coefficient	38024	0.40	0.11	0.23	0.53
GiniOB	33789	0.47	0.14	0.00	0.75
GiniK	33455	0.50	0.20	0.00	1.00
Income differences too large	39463	0.84	0.36	0.00	1.00
Relative earnings shopkeeper	36311	1.00	0.51	0.00	11.50
Relative earnings unskilled	36406	1.00	0.42	0.00	8.54
Relative earnings minister	34297	1.00	1.19	0.00	46.25
Relative earnings doctor	35034	1.00	0.78	0.00	26.95
Relative earnings chairman	34091	1.00	1.91	0.00	99.92
Household-level control variables					
Subjective rank	39680	0.50	0.18	0.10	1.00
HH income objective rank	33474	0.59	0.28	0.10	1.00
Ln Number books in childhood	39166	3.55	1.69	0.00	6.91
Unemployed	39866	0.10	0.30	0.00	1.00
Not in labour force	39866	0.16	0.37	0.00	1.00
Age	40155	42.90	13.42	18.00	98.00
Age squared	40155	2020.34	1229.17	324.00	9604.00
HH size 1	39967	0.11	0.32	0.00	1.00
HH size 2	39967	0.24	0.43	0.00	1.00
HH size 3,4	39967	0.45	0.50	0.00	1.00
Separated	39933	0.09	0.29	0.00	1.00
Widowed	39933	0.04	0.21	0.00	1.00
Married	39933	0.60	0.49	0.00	1.00
Face-to-face interview	40223	0.70	0.46	0.00	1.00
	N	median	sd	min	max
Macroeconomic information					
GNI per capita in PPP USD	40	27735.00	12198.56	6830.00	56500.00
Population (in mio)	40	13.89	211.92	0.32	1331.26
Gini Eurostat	23	0.29	0.04	0.23	0.38

B Additional estimation results

In section 4 we have used linear probability models. Table B.1 employs an ordered probit model that can be compared to table 1 based on the linear probability model. The results do not change qualitatively. The marginal effects are computed for the probability of type A. A one unit increase in the *subjective rank* increases the probability of type A by 2 percentage points. Likewise, the probability of type A is 18 percentage point higher for someone in the top percentile relative to someone in the bottom percentile. This effect is large relative to the sample frequency of type A of 26%.

Table B.1: Perceived type of society and income rank: OLS vs. ordered probit

Dependent variable	Perceived type of society					
	OLS	Ord. Probit	OLS	Ord. Probit	OLS	Ord. Probit
	(1)	(2)	(3)	(4)	(5)	(6)
Subjective rank	-0.253***	-0.254***			-0.251***	-0.245***
	(0.030)	(0.027)			(0.025)	(0.023)
HH income objective rank			0.016	0.011	0.048	0.043
			(0.035)	(0.032)	(0.035)	(0.031)
HH income objective rank squared			-0.085***	-0.079***	-0.068***	-0.063***
			(0.024)	(0.021)	(0.024)	(0.021)
Ln Number books in childhood	-0.006***	-0.005***	-0.008***	-0.007***	-0.004**	-0.004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Unemployed	0.012	0.014*	0.020**	0.021**	0.011	0.012
	(0.008)	(0.008)	(0.009)	(0.008)	(0.009)	(0.008)
Not in labour force	-0.005	-0.006	-0.005	-0.005	-0.007	-0.007
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.005)
Age	0.002	0.002*	0.002*	0.002*	0.002	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age squared	-0.000	-0.000	-0.000	-0.000*	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HH size 1	0.008	0.009	-0.003	-0.002	0.006	0.006
	(0.008)	(0.008)	(0.013)	(0.012)	(0.011)	(0.011)
HH size 2	0.012	0.013	0.010	0.010	0.013	0.013
	(0.008)	(0.008)	(0.010)	(0.010)	(0.010)	(0.009)
HH size 3,4	0.010*	0.011*	0.012	0.011	0.014*	0.014**
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Separated	0.014**	0.013*	0.015*	0.013*	0.013	0.011
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
Widowed	-0.000	-0.002	0.002	0.002	0.000	-0.001
	(0.011)	(0.011)	(0.013)	(0.013)	(0.013)	(0.012)
Married	0.005	0.005	0.005	0.005	0.008	0.008
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)
Face-to-face interview	0.041***	0.035***	0.022***	0.019***	0.011***	0.009**
	(0.005)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Constant	0.771***		0.690***		0.790***	
	(0.030)		(0.031)		(0.035)	
Country fixed-effects	yes	yes	yes	yes	yes	yes
Religion controls	yes	yes	yes	yes	yes	yes
Sample Frequency Type A		0.27		0.26		0.26
Log-Likelihood	-9140.46	-43387.42	-7943.92	-37357.72	-7665.45	-36913.81
Observations	35844	35844	30659	30659	30487	30487

The dependent variable is *Perceived type of society*. The table compares OLS estimates with marginal effects from ordered probit estimations. The marginal effects refer to the probability of the outcome of Type A. The specifications correspond to those in table 1 (cols 1 to 3). Standard errors in parentheses are adjusted for clustering at the country level. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level. Variables are defined in appendix A.

C Why it is impossible to use the type-of-society question to derive a unique measure of subjective inequality

In section 2.1 we have discussed the difficulties to associate each of the five type-of-society diagrams with an unequivocal and robust inequality measure like a Gini coefficient. As a consequence, it is almost impossible to infer from the answers to this question whether an individual under- or overestimates the degree of inequality in society. As stated in the text this difficulty is due to at least three factors:

1. It is unknown how individuals split the income distribution into seven classes.
2. It might be the case that none of the five diagrams corresponds to an individual's view of the income distribution in which case they have to choose the "least evil".
3. It is unknown how individuals assess the *within*-class income distribution.

As briefly described in the text the related literature has dealt with these difficulties by using various assumptions with respect to each of these three issues. As far as the first point (the choice of the income levels associated with the seven bars) is concerned a straightforward approach is to assume that the income gap between all neighboring classes is identical and we call this assumption EquGap. This approach has been illustrated in Figure 2a for a numerical example in which we assumed that incomes are lognormally distributed with a mean (annual) income of $\overline{Y} = 50\,000$ and a standard deviation such that the Gini coefficient is 0.3. In order to be able to form reasonable categories we cap the income distribution at the 99th percentile where the income is 153 122 and add the remaining 1% to the highest class. Under the assumption of equal spaces between the classes the seven groups correspond to the following income intervals: $[0, 21\,875]$, $[21\,875, 43\,749]$, $[43\,749, 65\,624]$, $[65\,624, 87\,499]$, $[87\,499, 109\,373]$, $[109\,373, 131\,248]$ and $[131\,248, \infty]$ (after adding the richest percent).

Beside EquGap the related literature has used other possibilities to demarcate the seven income levels. Niehues (2014), for example, has proposed a definition of social classes that relate to median income. In particular, she has suggested the following boundaries. If the median income is normalized to 100 then the lowest class is assumed to have income levels that are lower than 60, in the second lowest class one can observe incomes between 60 and 80, while the following classes are given by incomes between 80

and 110, 110 and 150, 150 and 200, 200 and 250 and the richest class has incomes of more than 250. For our example the income intervals of this alternative definition are given by : $[0, 25\,861]$, $[25\,861, 34\,481]$,$[34\,481, 47\,411]$, $[47\,411, 64\,652]$, $[64\,652, 86\,202]$,$[86\,202, 107\,753]$ and $[107\,753, \infty]$ if we add again the richest percent. We have called this the RelMedian assumption and Figure 2b illustrates the resulting shape.

In order to calculate the Gini coefficients that correspond to each of the types A to E in figure 1a for both the EquGap and the RelMedian assumption we take the following steps. First, we measure the size of each bar and interpret this as the population share of the corresponding classes. Then we calculate the Gini coefficient using the formula:

$$G = 2\frac{\sum_{i=1}^{N} Y_i \frac{i-1/2}{N}}{\sum_{i=1}^{N} Y_i} - 1, \tag{1}$$

where N is the number of observations (chosen to be large) and Y_i is the income of individual i (which is assumed to be identical within each of the seven classes). As suggested by Gimpelson & Treisman (2015) we also use the correction for the bias associated with the use of grouped data (see Van Ourti & Clarke 2011). For the assumption of equal spacing between income classes (EquGap) we assign the income levels $\{0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5\}$ to the seven classes while for the assumption of median-related income classes (RelMedian) we use the income levels $\{0.3, 0.7, 0.95, 1.3, 1.75, 2.25, 3\}$. The resulting Gini coefficients for the five types of society are 0.53 (A), 0.43 (B), 0.35 (C), 0.23 (D) and 0.23 (E) for assumption EquGap and 0.49 (A), 0.39 (B), 0.33 (C), 0.24 (D) and 0.25 (E) for assumption RelMedian. As stated in the text, the qualitative pattern of the Gini coefficients is similar for both assumptions (i.e. it always decreases from type A to type E) while the quantitative differences are considerable.

The second difficulty in relating the type-of-society question to an unambiguous Gini coefficient stems from the potential challenge to match a subjective image of society to one of the five presented diagrams. This issue is clearly visible by comparing figures 2a and 2b to the five diagrams in the type-of-society question (figure 1a). Neither of the two figures shows a high degree of similarity to any of the five types. The shape of the EquGap assumption resembles a tree, although with wider branches at the lower end and a thinner top than in diagram C of the type-of-society question. This is due to the fact that incomes above 100 000 are already rather rare and thus the top two classes are tiny. For the RelMedian assumption, on the other hand, the closest fit seems to be the diamond although with less symmetry than in type D of the question. It might thus well be that

someone who chooses type D is just selecting the "least evil" while his subjective image of society is characterized by a considerably larger or smaller Gini coefficient.

A third factor exacerbates the difficulties for clear-cut calculations of Gini coefficients—the necessary assumption about the within-class distribution. For the previous calculations we have associated each bar with the midpoint of the interval. This, however, is not the only possibility. If we assume, e.g., that the income level is given by the lower or upper boundary then this results in considerably different Gini coefficients. In particular, for the case of the EquGap assumption the use of income levels $\{1, 2, 3, 4, 5, 6, 7\}$ leads to Gini coefficients that are much lower than before: 0.43 (A), 0.35 (B), 0.30 (C), 0.21 (D) and 0.20 (E). The use of the lower boundary, on the other hand, i.e. of income levels $\{0, 1, 2, 3, 4, 5, 6\}$ implies the following (much higher) Gini coefficients: 0.68 (A), 0.54 (B), 0.43 (C), 0.27 (D) and 0.26 (E). The ranking of types A to D/E is again unaffected by these choices but the precise numbers are quite different.

In reality one would assume that in order to calculate their subjective Gini coefficients individuals will use not only one income level for each of the seven bars but they will also resort to their a-priori knowledge about the income distribution to make some implicit assumptions about the with-in class distribution.

S Supplementary Appendix for "Perceptions of Inequality" (not for publication)

Appendix S.1 provides additional results for the theoretical part of the paper. In particular, we look at additional assumptions concerning the subjective density functions. Appendix S.2, on the other hand, provides extensions (country-specific estimations) for the empirical part.

S.1 Additional assumptions about reference groups

In this appendix we present illustrations for subjective density functions that differ from the assumptions made in section 3.2.2 of the paper. We show the figures for the decile-specific reference groups, for the decile-specific shapes of societies and for the three distributional perception measures that correspond to figures 6, 7 and 8 in the text.

S.1.1 Triangular subjective density function with assumption EquGap

We assume here that the subjective distribution function is given by a triangular distribution with a minimum at 0, a maximum at the 99th percentile of the objective distribution and—in line with assumption 2—the mode at the own income Y_i. In figure S.1 we show the reference groups that are associated with this assumption. One can observe a problematic and counter-intuitive property of this assumption: up to the fifth decile the subjective weight of the own income level is *lower* than the objective frequency. This is in contraction to the presumption that members the own class are overrepresented in the reference group. Despite this objection the implications of the assumption are more or less in line with the one presented in section 3.2.2. Individuals' assessment of the shape of society turns from a tower (or a "tree") to a vase, there is overestimation (underestimation) of the subjective rank for low income (high income) individuals (though the compression is less than before) and the subjective mean income increases and the subjective Gini coefficient decreases with the position in the objective income distribution.

S.1.2 Weibull subjective density function with assumption EquGap

As an alternative subjective density function one could also assume a two-parametric Weibull density function instead of the lognormal function used in section 3.2.2. We use the same assumptions that the mode of the density function is at the own income Y_i and

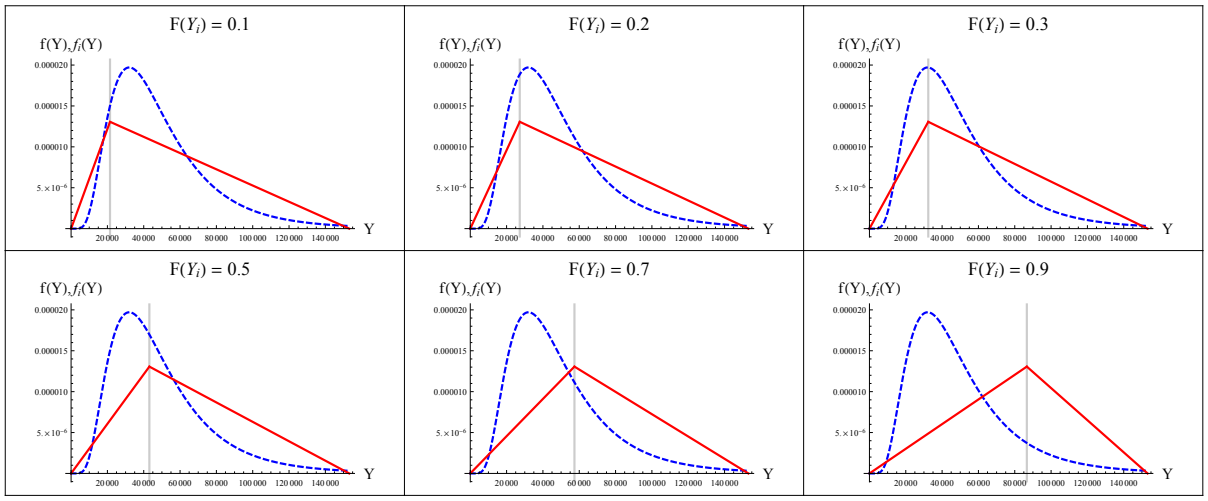


Figure S.1: The pictures shows the objective density function $f(Y)$ (dashed) and the subjective density functions $f_i(Y)$ (solid) for six percentiles of the (objective) income distribution, where the vertical lines indicate the corresponding own incomes. The subjective density function is assumed to be triangular. The objective income distribution is assumed to be lognormal with a mean income of 50 000 and a Gini coefficient of 0.3.

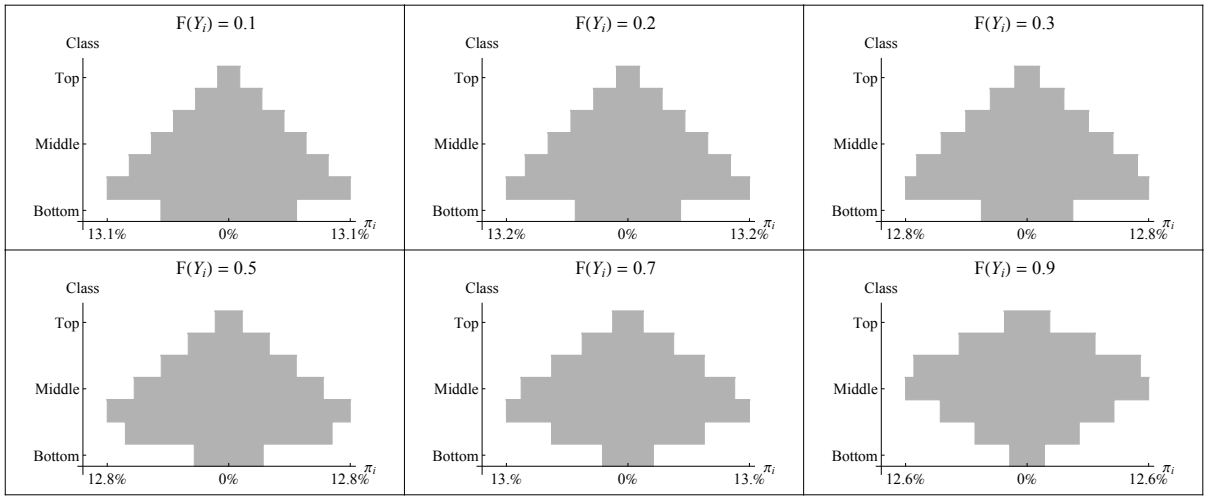


Figure S.2: The pictures show the shape of society (the subjective shares π_i of the seven classes) for individuals at six deciles. It is assumed that the subjective density functions $f_i(Y)$ are triangular as illustrated in figure S.1 and that individuals demarcate the seven classes by using the assumption EquGap.

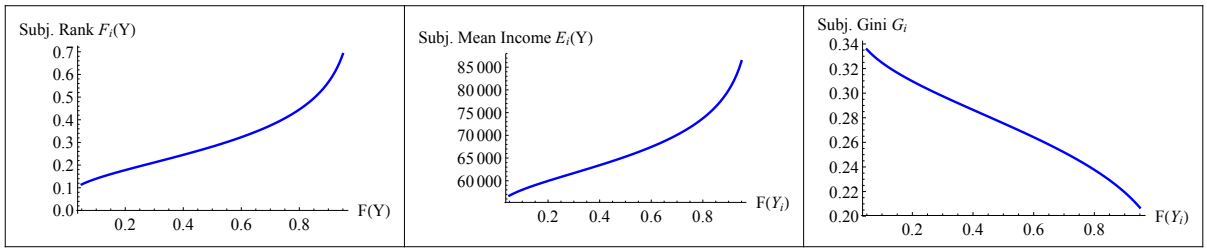


Figure S.3: The pictures show three measures of distributional perceptions as stated in definitions 3 to 5. It is assumed that the subjective density functions $f_i(Y)$ are triangular as illustrated in figure S.1.

that the density at this mode is the same as the density at the mode of the objective distribution function.[18] The results are completely parallel to the one presented for the lognormal function. In figure S.4 one can see that now for $F(Y_i) = 0.3$ the subjective and objective distribution function are still very close, while for lower percentiles there is a downward bias while for upper percentiles there is an upward bias. The subjective shape of society again turns from a tower to a vase and also the properties of the four subjective measures in figure S.6 are completely analogous to figure 8.

S.1.3 Lognormal subjective density function with assumption RelMedianp

So far (in section 3.2.2 and in the sections above) we have used the EquGap assumption. The use of the assumption RelMedian does not change the results in an important manner, in fact it makes them stronger. The subjective shapes of society (see figure S.7) are also more pronounced (now starting with a clear tower and going up to a more vase-like figure). Again it is the case that the subjective degree of inequality is lower for higher incomes. The three perception measures are unchanged.

S.2 Country-Specific Estimations

The 40 countries included in the sample differ in many dimensions. For example, the Gross National Income per capita ranges from 6 830 to 56 500 PPP-adjusted US dollars.

[18] This is implemented in the following manner. For individual i the subjective probability density function is of the Weibull type and given by: $f_i(Y;\lambda_i,k_i) = \frac{k_i}{\lambda_i}\left(\frac{Y}{\lambda_i}\right)^{k_i-1} e^{-(Y/\lambda_i)^{k_i}}$ for $Y \geq 0$ and 0 for $Y < 0$, where $k_i > 1$ and $\lambda_i > 0$ are the individual-specific shape and scale parameter, respectively. These parameters are determined by simultaneously solving two equations. First, $Y_i = \lambda_i \left(\frac{k_i-1}{k_i}\right)^{\frac{1}{k_i}} \equiv \text{Mo}_i$ (i.e. the mode of the function $f_i(\cdot)$ is at Y_i). Second, $f_i(Y_i) = f_i(\text{Mo}_i) = f(\text{Mo}(Y))$ (i.e. the density at this mode is equal to the density of the mode of the objective density function).

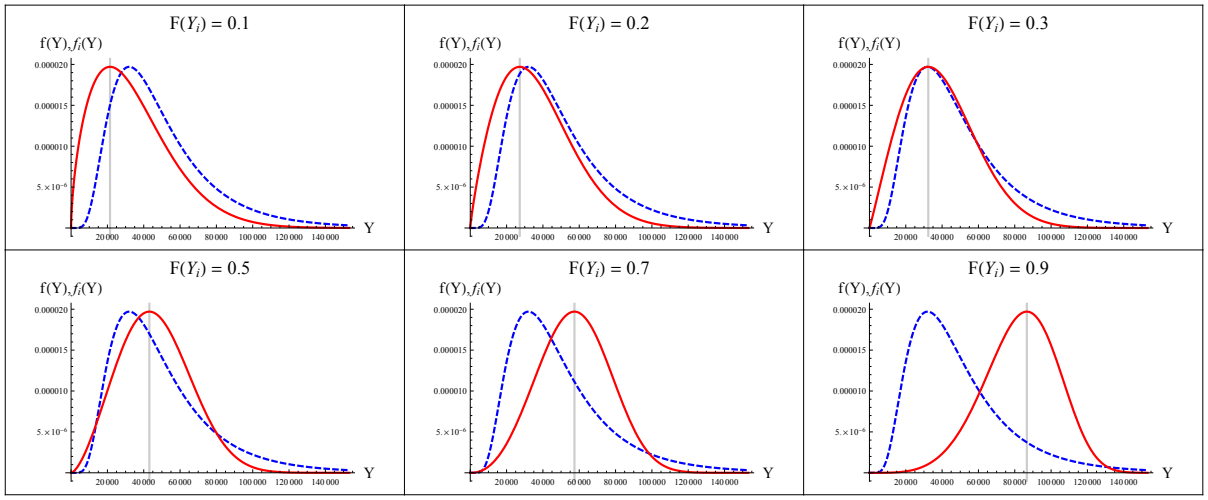


Figure S.4: The pictures shows the objective density function $f(Y)$ (dashed) and the subjective density functions $f_i(Y)$ (solid) for six percentiles of the (objective) income distribution, where the vertical lines indicate the corresponding individual incomes. The subjective density function is of the Weibull type and it constructed for each individual as described in the text. The objective income distribution is assumed to be lognormal with a mean income of 50 000 and a Gini coefficient of 0.3.

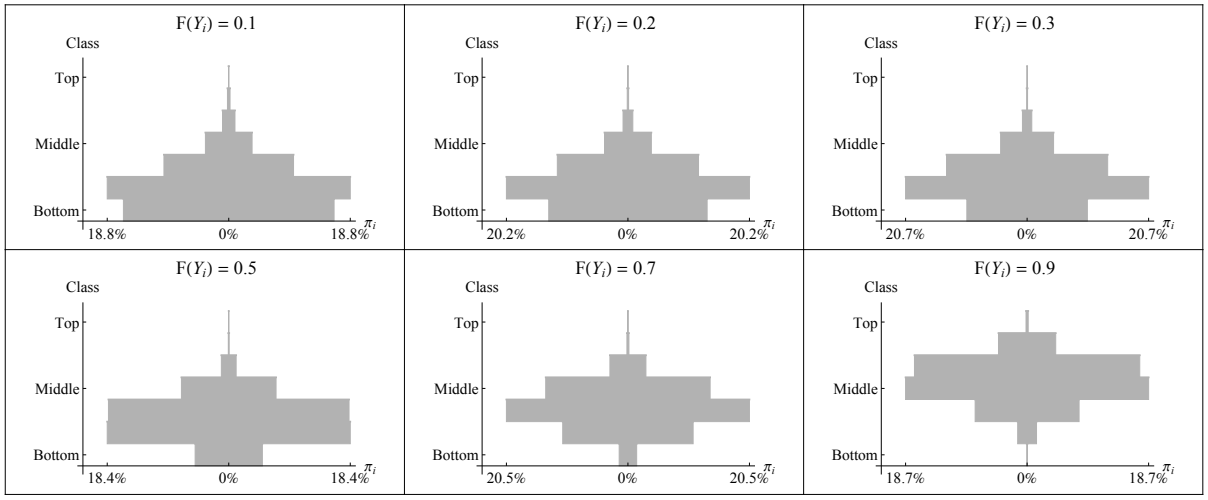


Figure S.5: The pictures show the shape of society (the subjective shares π_i of the seven classes) for individuals at six deciles. It is assumed that the subjective density functions $f_i(Y)$ are of the Weibull type as illustrated in figure S.4 and that individuals demarcate the seven classes by using the assumption EquGap.

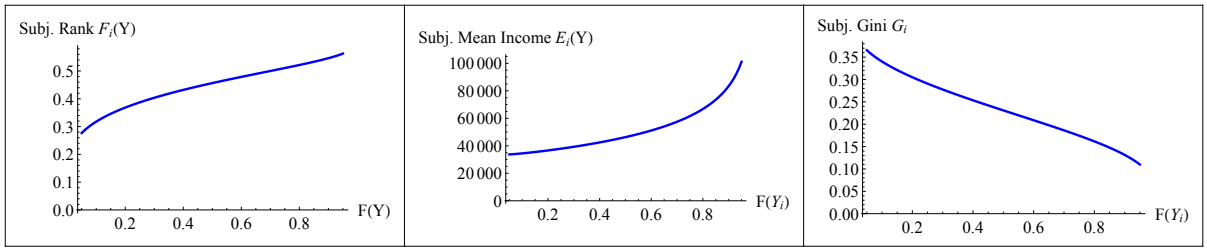


Figure S.6: The pictures show three measures of distributional perceptions as stated in definitions 3 to 5. It is assumed that the subjective density functions $f_i(Y)$ are of the Weibull type as illustrated in figure S.4.

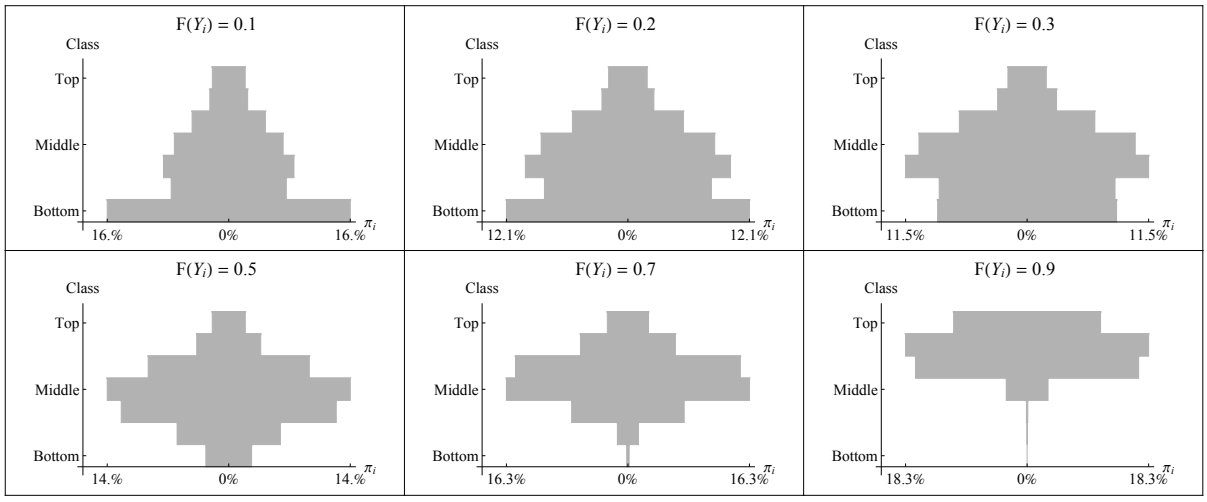


Figure S.7: The pictures show the shape of society (the subjective shares π_i of the seven classes) for individuals at six deciles. It is assumed that the subjective density functions $f_i(Y)$ are of the lognormal type as illustrated in figure 6 and that individuals demarcate the seven classes by using the relative-to-median assumption RelMedian.

The societal or political characteristics differ widely in countries like China, Venezuela or the US. Subjective views on the relative position and views about overall inequality may depend on the size of the welfare state, etc. As a case in point, figure S.8 shows answers to the type-of-society question for rich countries (top panel) and poor countries (bottom panel)—rich and poor is defined by the cross-country median of Gross National Income.[19] It is evident that the share of responses for type A is much higher for poor than for rich countries. Scandinavian countries have the highest share of type D answers. Heterogeneity is also apparent for *Subjective rank*, although it is somewhat smaller (figure S.9).

Up until now, we have controlled for country differences by including country-fixed effects. In this section, we look at subgroups of countries and at individual countries to make sure that results are not driven by single (influential) countries.

Implication 1. Figure S.10 shows the relationship of the objective rank and the subjective rank for various country groups. Respondents in rich countries perceive themselves in a somewhat higher subjective rank than respondents from poor countries (top left panel). Despite this difference, the shape of the relationship is very similar between rich and poor countries. A similar observation holds for rich European countries and the US relative to not rich non-European countries (top right panel). Perceptions of respondents in Eastern European countries have been found to differ from respondents from other countries with respect to many aspects, i.e. views about their societies. Potentially, this could also be of relevance in our context, however the bottom left panel shows that the perception of the subjective rank is comparable to those from other countries. Finally, we also find that the extent of objective inequality does not exert strong influence over our overall findings (bottom right panel).[20]

Implication 2. Table S.1 confirms that the negative relationship between the type of society and the subjective rank does not depend on specific groups of countries. In all six country groups that are analyzed in table S.1, the coefficient of *Subjective rank* is significant, negative and of about similar size.

[19]The country codes are: Argentina (AR), Austria (AT), Australia (AU), Belgium-Flanders (BE), Bulgaria (BG), Switzerland (CH), Chile (CL), China (CN), Czech Republic (CZ), Germany (DE), Denmark (DK), Estonia (E), Spain (ES), Finland (FI), France (FR), Great Britain (GB), Croatia (HR), Hungary (HU), Israel (IL), Iceland (IS), Italy (IT), Japan (JP), South Korea (KR), Latvia (LV), Norway (NO), New Zealand (NZ), Philippines (PH), Poland (PL), Portugal (PT), Russia (RU), Sweden (SE), Slovenia (SI), Slovakia (SK),Turkey (TR), Taiwan (TW), Ukraine (UA), United States (US), Uruguay (UY), Venezuela (VE), South Africa (ZA).

[20]The objective Gini is only available for European countries.

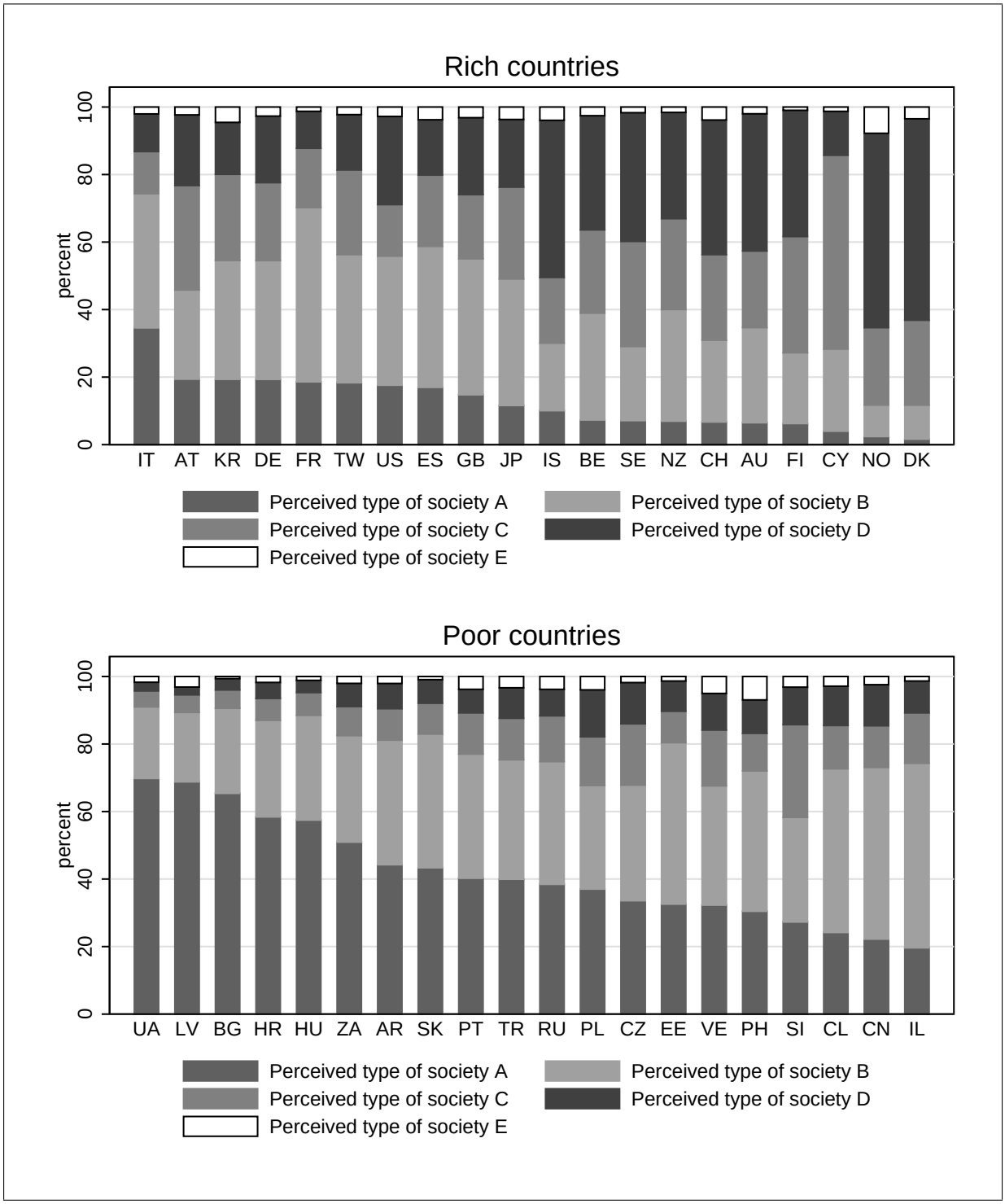


Figure S.8: The figures shows the relative sample frequencies of the types of society that respondents perceive. Countries are ordered by the frequency of type A.

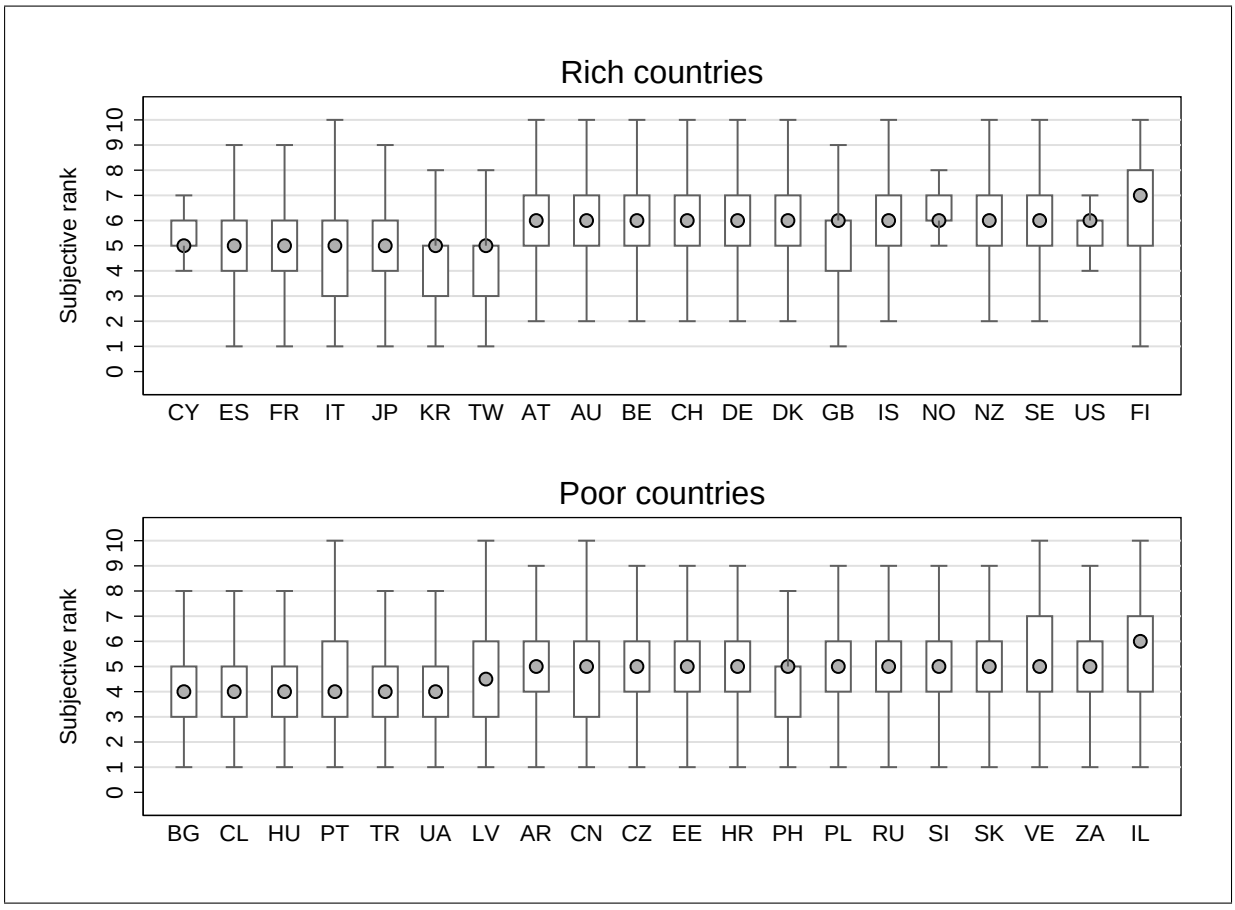


Figure S.9: The figures shows box plot for *Subjective rank*. Circles denotes median values and boxes the 25-th to 75-th percentiles. Countries are ordered by their median.

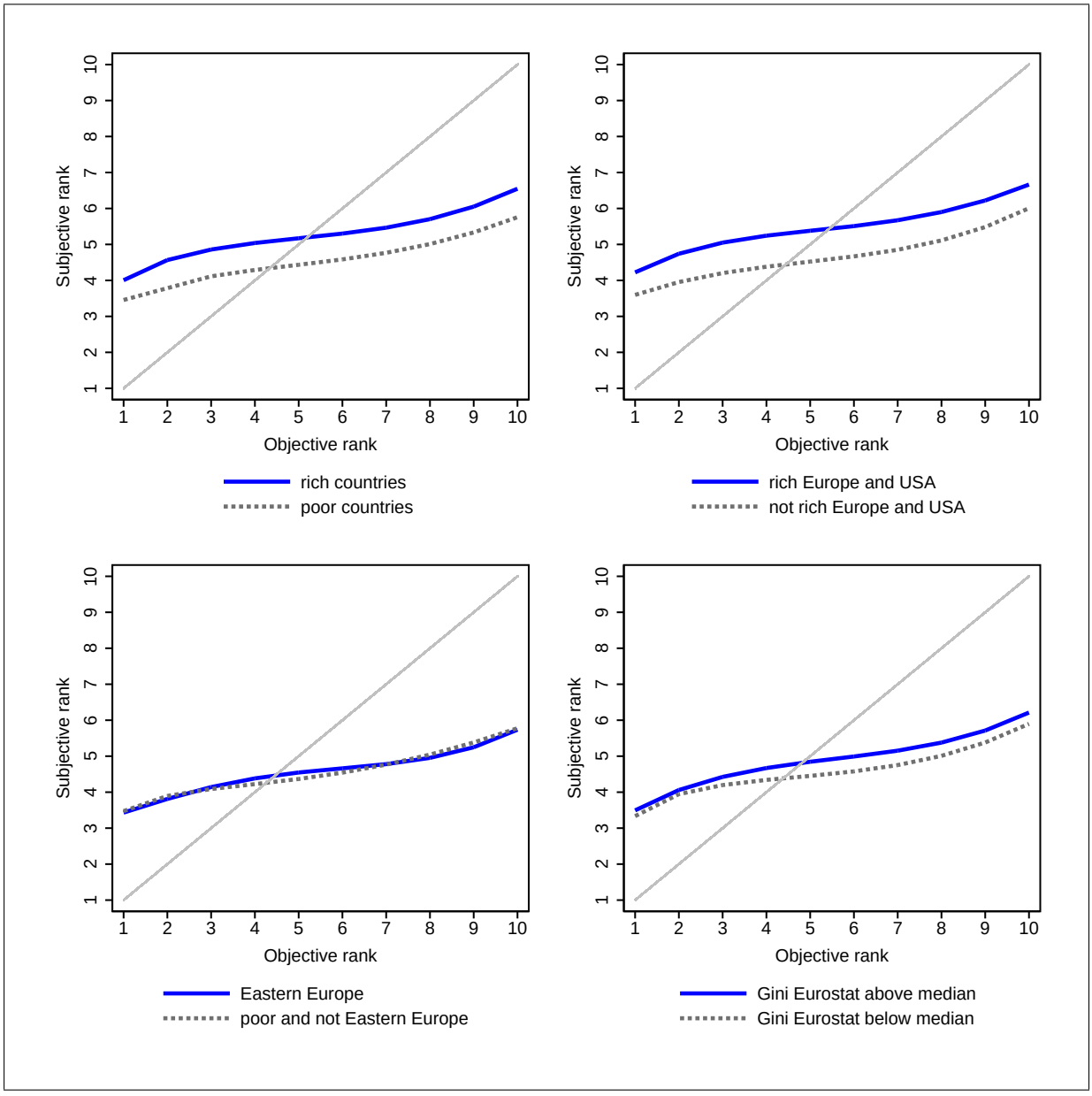


Figure S.10: The figures show polynomial lines fitted to the raw data for four country groups. Rich and poor countries are defined by being above or below the median GNI. Eastern Europe denotes the group of all Central and Eastern European countries including Russia. "Gini Eurostat above (below) median" refers to European countries that have an above (below) median Gini coefficient for disposable income as reported by Eurostat. Variables are described in appendix A.

Table S.1: Perceived type of society and income rank - Different country groups

Dependent variable	*Perceived type of society* (Type A=1, Type B=0.66, Type C=.33, Type D/E=0)					
	Rich countries (1)	Poor countries (2)	Rich Europe & USA (3)	non Euro. & USA (4)	Eastern Europe (5)	Above median Gini (6)
Subjective rank	-0.356***	-0.182***	-0.346***	-0.215***	-0.234***	-0.310***
	(0.032)	(0.031)	(0.047)	(0.031)	(0.036)	(0.055)
Country fixed-effects	yes	yes	yes	yes	yes	yes
Religion controls	yes	yes	yes	yes	yes	yes
Adj. R-squared	0.18	0.08	0.24	0.17	0.11	0.24
Observations	17236	18608	11699	24145	7589	8020

The dependent variable is *Perceived type of society* coded as 0 (Type D/E), 0.33 (Type C), 0.66 (Type B) and 1 (Type A). All models report estimates from a linear probability model as in column 1 of table 1 and include country-fixed effects and fixed-effects for the religion of respondents and the same household controls as in table 1. The sample is constrained to rich countries in column 1, to poor countries in column 2, to rich European & the USA in column 3, to non-rich European & non-USA in column 4, to Central and Eastern European countries in column 5 and to countries with an above median Gini Eurostat in column 6. ***, **, * denote significance at the 0.01, 0.05 and 0.10-level. Variables are defined in appendix A.

Given that we have a sufficient number of observations for each country, we can also run each regression country-by-country. Figure S.11 visualizes the findings from this exercise. In each regression we have applied the same specification as in column 1 of table 1. The horizontal axis denotes the point estimates for *Subjective rank* and the vertical axis the corresponding t-values. Point estimates lower than zero are in line with the prediction of our theoretical framework. If point estimates are also below the horizontal line (-1.96) they are statistically significant at the 5% level. The figure reveals that the predicted effect can be found in a large share of countries.

Implication 3. The same exercises is repeated for implications 3 and 4. In both cases we present the coefficients for *Subjective rank* in the left panel and for the objective rank in the right panel. Since the objective rank enters quadratically, we display the overall average marginal effect.

Figure S.12 reveals that not many countries display a significantly positive effect between the subjective estimated income of unskilled factory workers and the *Subjective rank* (left panel). For the objective rank (right panel), however, we find the predicted pattern for many countries.

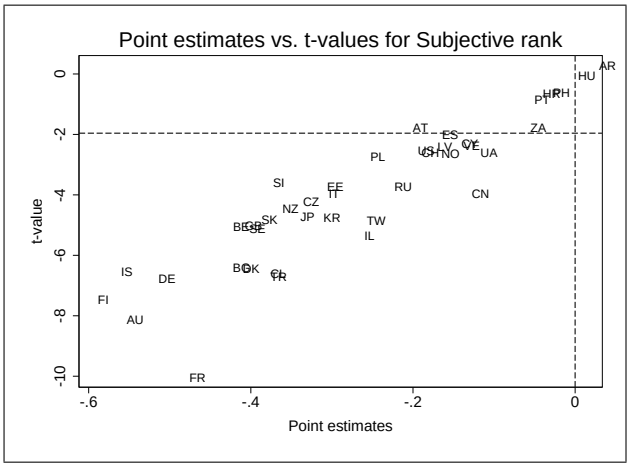


Figure S.11: The figure summarizes results obtained when specification 1 of table 1 is estimated separately for each country. The horizontal axis denotes the point estimates for *Subjective rank* and the vertical axis the corresponding t-values. Point estimates below the horizontal line (-1.96) are significant at the 5% level. Variables are described in appendix A.

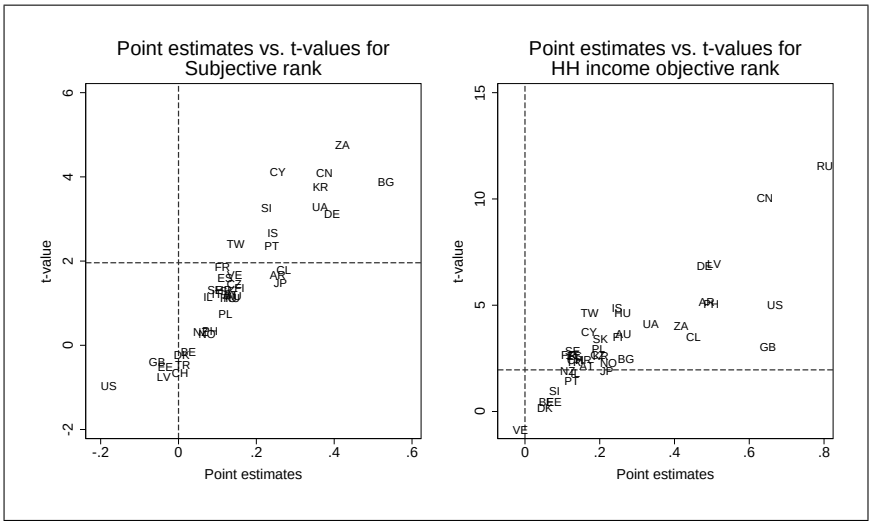


Figure S.12: The figures summarize results obtained when specification 3 (left panel) and 4 (right panel) of table 4 is estimated separately for each country. The horizontal axis denote the point estimates for *Subjective rank* and *Objective rank*, respectively, and the vertical axis the corresponding t-values. Point estimates larger than zero and above the horizontal line (1.96) are significant at the 5% level. Variables are described in appendix A.

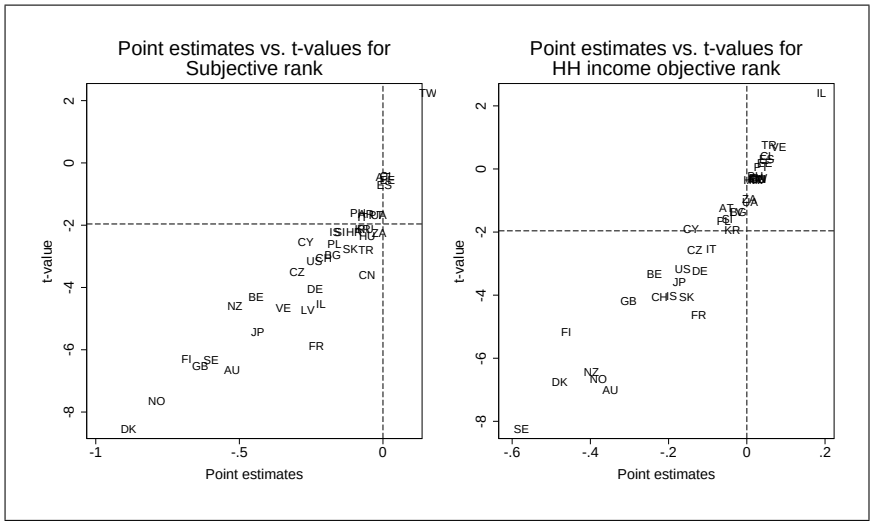


Figure S.13: The figures summarize results obtained when specification 5 (left panel) and 6 (right panel) of table 5 is estimated separately for each country. The horizontal axis denote the point estimates for *Subjective rank* and *Objective rank* respectively and the vertical axis the corresponding t-values. Point estimates smaller than zero and below the horizontal line (-1.96) are significant at the 5% level. Variables are described in appendix A.

Implication 4. The subjective rank has a negative effect on the subjective Gini, as predicted, in the large majority of countries (Figure S.13, left panel). For the objective rank, the pattern is less clear-cut.

Index of Working Papers:

March 5, 2015	Jonas Dovern, Martin Feldkircher, Florian Huber	200	Does Joint Modelling of the World Economy Pay Off? Evaluating Global Forecasts from a Bayesian GVAR
May 19, 2015	Markus Knell	201	The Return on Social Security with Increasing Longevity
June 15, 2015	Anil Ari	202	Sovereign Risk and Bank Risk-Taking
June 15, 2015	Matteo Crosignani	203	Why Are Banks Not Recapitalized During Crises?
February 19, 2016	Burkhard Raunig	204	Background Indicators
February 22, 2016	Jesús Crespo Cuaresma, Gernot Doppelhofer, Martin Feldkircher, Florian Huber	205	US Monetary Policy in a Globalized World
March 4, 2016	Helmut Elsinger, Philipp Schmidt-Dengler, Christine Zulehner	206	Competition in Treasury Auctions
May 14, 2016	Apostolos Thomadakis	207	Determinants of Credit Constrained Firms: Evidence from Central and Eastern Europe Region
July 1, 2016	Martin Feldkircher, Florian Huber	208	Unconventional US Monetary Policy: New Tools Same Channels?
November 24, 2016	François de Soyres	209	Value Added and Productivity Linkages Across Countries
November 25, 2016	Maria Coelho	210	Fiscal Stimulus in a Monetary Union: Evidence from Eurozone Regions
January 9, 2017	Markus Knell, Helmut Stix	211	Inequality, Perception Biases and Trust

January 31, 2017	Steve Ambler, Fabio Rumler	212	The Effectiveness of Unconventional Monetary Policy Announcements in the Euro Area: An Event and Econometric Study
May 29, 2017	Filippo De Marco	213	Bank Lending and the European Sovereign Debt Crisis
June 1, 2017	Jean-Marie Meier	214	Regulatory Integration of International Capital Markets
October 13, 2017	Markus Knell	215	Actuarial Deductions for Early Retirement
October 16, 2017	Markus Knell, Helmut Stix	216	Perceptions of Inequality