Check for
updates

Measuring Perceptions of Economic Inequality and Justice: An Empirical Assessment

Nicholas Heiserman[1] · **Brent Simpson[1]**

Abstract

How should we measure people's perceptions of—and attitudes about—economic inequality? A recent literature seeks to quantify the level of inequality that people, especially Americans, perceive and prefer in society. These findings have garnered much attention from both social scientists and the public. But many of the methods used in this literature are either known to have methodological issues or have not been thoroughly compared against other methods. Thus it is not clear which, if any, are valid and reliable measures of perceived, or preferred, inequality. To assess these measures, we conducted a large web-based study ($N=831$) to compare key methods for measuring perceived inequality and their related justice attitudes. In addition to comparing the resultant summary statistics, we assess how well the different measures correlate with each other and with Likert scale measures of perceived inequality. Our analysis reveals a range of issues with these measures, including failure to provide logical responses, large method effects on point estimates of inequality, and low correlations between methods and with criteria measures. We conclude our analysis with three recommendations for researchers aiming to measure inequality perceptions and preferences.

Keywords Perceived inequality · Just inequality · Income inequality · Inequality estimation

Introduction

What do people know about inequality? Do they believe inequality is high? And do they think there is more inequality than there "ought" to be? The past decade has seen an upsurge in interest in research that quantifies perceptions and preferences about inequality in the USA (Chambers et al., 2014; Dawtry et al., 2015; Eriksson

✉ Nicholas Heiserman
    heisermn@email.sc.edu

[1] Department of Sociology, University of South Carolina, Columbia, SC 29205, USA

& Simpson, 2012, 2013; Hauser & Norton, 2017; McCall, 2013; Norton & Ariely, 2011; Pedersen & Mutz, 2018) and elsewhere (Arsenio, 2018; Gimpelson & Treisman, 2018; Kiatpongsan & Norton, 2014; Kuhn, 2019). If subjective estimates can be mapped onto the objective metrics used by inequality researchers, then researchers can assess whether people over- or underestimate the level of inequality. Correcting these misperceptions could then inform and alter public opinion about inequality and relevant social policies (Hauser & Norton, 2017; Heiserman et al., 2020; McCall et al., 2017).

The most widely publicized study in this literature found that Americans underestimate the level of income inequality in the USA and prefer even less inequality than they erroneously believe exists (Norton & Ariely, 2011). But other research suggests that these findings are methodological artifacts (Chambers et al., 2014; Eriksson & Simpson, 2012, 2013). Similar criticisms have been leveled at other techniques for measuring perceived inequality and related attitudes (Markovsky & Eriksson, 2012; Pedersen & Mutz, 2018). Perhaps as a result, measures of perceived (and preferred) inequality have proliferated.

But differences between measures make it unclear whether conclusions about lay beliefs and preferences about inequality are "real," and whether they would have been obtained using different measures intended to tap into the same underlying perception. Some inconsistencies may stem from researchers' focus on different aspects of inequality (e.g., wealth vs. income). But even when researchers work with the same definition, different measurement methods can yield different findings (Norton & Ariely, 2011; Eriksson & Simpson, 2012; Pedersen & Mutz, 2018). Since many researchers assume that these studies are all assessing the same (or highly related) perceptions of the extent of economic inequality (Hauser & Norton, 2017), it is important to understand the extent to which conclusions about inequality perceptions or preferences rest on the particular measures employed.

We take stock of this rapidly growing literature to assess the extent to which different measures converge on similar conclusions, and to provide recommendations for future studies of inequality perceptions and attitudes. This is important not only for academic reasons. Since these questions are of broad interest in political and economic debates, it is particularly important to promote best practices.

To address these issues, we conducted a study in which we asked participants to provide their perceptions of—and preferences for—inequality using a wide range of recently used methods. After reviewing these methods and the results, we outline three key guidelines to help researchers improve future studies on inequality and justice perceptions.

The Relationship Between Perceived Inequality/Justice and Attitudes About Inequality

Much of the recent literature addresses both how much inequality people believe exists, and how much inequality they think ought to exist. These are distinct but related questions. For instance, rational-choice models predict that when inequality is higher (or perceived to be higher), more citizens should come to see that

inequality as unacceptable and against their own interests, and will thus prefer policies aimed at redressing inequality (Gimpelson & Treisman, 2018; Meltzer & Richard, 1981; McCall, 2013; McCall et al., 2017; Niehues, 2014). Indeed, while studies often show that areas with more income inequality can actually become more tolerant of inequality (Schröder, 2017; Trump, 2017), experiments show that when people perceive inequality as higher they tend to evaluate existing economic arrangements more negatively (Heiserman et al., 2020; McCall et al., 2017).

Theories of distributive justice also specify how people react to inequality (Hegtvedt & Isom, 2014; Jasso & Wegener, 1997). These theories typically assume that individuals compare actual states of affairs against what they consider an "ideal" allocation of rewards or resources. In the case of economic inequality, that means that people compare the amount of inequality they perceive against an 'ideal' level of inequality. If the actual amount of inequality is higher than the ideal, they should judge inequality to be less acceptable and more unfair (Hegtvedt & Isom, 2014; Kiatpongsan & Norton, 2014; Osberg & Smeeding, 2006). Indeed, the link between justice evaluations and inequality attitudes is often assumed to be consistent enough that many researchers study quantitative measures of perceived justice as a direct measure of attitudes about inequality in themselves (Kiatpongsan & Norton, 2014; Osberg & Smeeding, 2006; Schneider & Castillo, 2015). The link between justice perceptions and attitudes toward inequality may not be so unambiguous, however. A variety of factors can suppress individuals' negative emotional reactions to deviations from the outcomes they see as ideal or 'just', including denigration of the work ethic and merit of the poor (Heiserman & Simpson, 2017; Hunt & Bullock, 2016; Schneider & Castillo, 2015) and belief in economic meritocracy (Cech, 2017; Heiserman et al 2020; Reynolds & Xian, 2014). Nevertheless, justice perspectives predict that when the wealthy are seen as having more (and the poor as having less) than they fairly deserve, people should come to see inequality as less acceptable and should want that inequality reduced.

Perceived Inequality Measures

Although the study of inequality perceptions has a long history in the social sciences (Kluegal & Smith, 1986; Evans et al., 1992), the aforementioned research by Norton and Ariely (2011) led to upsurge of interest in quantifying perceptions of inequality. Norton and Ariely asked American participants to provide the percentage of all wealth held by each of the five income quintiles in the USA, and found that they underestimated the level of inequality. We refer to this as the *Norton–Ariely* measure. Supplementary Material (SM) Section 5 provides the full text of all measures and information treatments.

Eriksson and Simpson (2012) introduced an alternative to the Norton–Ariely measure in which participants provide estimates of the average wealth, rather than the percentage, in each quintile. These averages are then converted into shares by dividing each average by the sum of all five averages. We refer to this as the *Eriksson–Simpson* measure. Though these two measures assessed perceived wealth
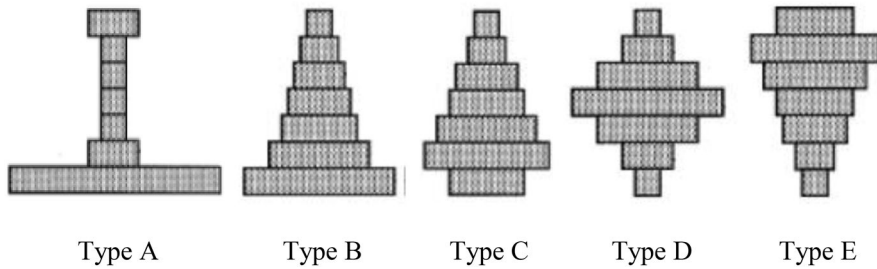
Fig. 1 Stratification belief diagrams

inequality, subsequent research has typically focused on income. We thus operationalize both measures in terms of income.

A third measure asks participants to provide the proportion of Americans who fall into different income ranges (Chambers et al., 2014; Dawtry et al., 2015). Chambers et al. (2014) used three income ranges, while Dawtry et al. (2015) used eleven. Since this task is not generally attributed to specific scholars, we refer to it as the *income brackets* measure.

An alternative to direct quantification is to present several possible inequality levels and ask participants which comes closest to reality. The International Social Survey Programme (ISSP) included in its Social Inequality Modules a graphical measure of 'collective stratification beliefs' which has been used in studies of international variation in perceived inequality (Arsenio, 2018; Gimpelson & Treisman, 2018; Hertel & Schöneck, 2019; Niehues, 2014). Participants choose which of five 'types' of societies comes closest to the actual U.S. distribution. Inequality is represented by histograms of the proportion of people at the top, middle, and bottom. Diagrams are ranked by the location of the bulk of society: most at the bottom in Type A to most near the top in Type E. We refer to this as the *stratification belief diagrams* measure (Fig. 1).

Researchers also use pie (or bar) charts to manipulate or measure perceptions of inequality (Chambers et al., 2014; McCall et al., 2017). These measures are useful since pie charts are both visually intuitive and mirror popular metaphors about inequality (e.g., how big a 'slice of the pie' different groups have). We refer to this as the *pie charts* measure.

We also included a measure (Davidai & Gilovich, 2018) that asks participants to rank the countries of the G7 (the USA, Canada, the UK, France, Germany, Italy, and Japan) in order of their level of inequality. This measure assesses perceived inequality relative to what participants' perceptions of other countries' levels of inequality. We refer to this as the *international ranking* measure.

Some work expresses perceived inequality in terms of the ratio of the perceived income of CEOs and unskilled factory workers (e.g., Kiatpongsan & Norton, 2014; McCall, 2013; Osberg & Smeeding, 2006). These two occupations anchor either end of the socioeconomic spectrum. Thus, the difference between participants' perceptions of their incomes is typically taken as a measure of perceived income inequality more broadly. We refer to this as the *perceived pay ratio* measure. As Jasso and

Wegener (1997) recommend, we log-transform all the pay and justice ratios in this study to make them more normally distributed. We describe the items we used to make this measure in more detail in the next sections.

$$\text{Percevied pay ratio} = \ln\left(\frac{\text{Perceived actual CEO income}}{\text{Perceived actual factory worker income}}\right)$$

Finally, much research uses *Likert scales* to measure perceptions of, and attitudes about, inequality. This includes national and international survey analyses (McCall, 2013; Schröder, 2017), as well as experiments (Clay-Warner et al., 2016; Melamed et al., 2018). Rather than review this voluminous literature, we instead note a critical issue with them. Most measures of inequality attitudes assess some evaluative or normative attitude (e.g., the GSS asks if income differences are *too* large, or if inequality exists to benefit the wealthy) or policy preferences (e.g., redistributive preferences) rather than simply having respondents indicate whether they perceive inequality as high or low. Thus, while there is extensive research on attitudes about inequality using Likert items (e.g., Cech, 2017; McCall, 2013), these measures focus much more on evaluations of inequality than the question of how much inequality respondents perceive.

Attitudes About Inequality

As noted above, much research assesses normative attitudes about whether inequality is *too high* (or *low*), compared to some ideal. If people think inequality is higher than they prefer, they may view it as unfair, which should alter their preferences for social action, e.g., income redistributive policies (Hegtvedt & Isom, 2014). A major approach to capturing inequality attitudes entails measuring people's attitudes about the justice or fairness of the incomes received by people in different occupations.

First, as discussed above, researchers frequently investigate attitudes about inequality by measuring whether respondents think people in different occupations earn more or less than they deserve. We refer to this approach as *occupational income justice ratio* measures. When people in high-status occupations (e.g., CEOs) are seen as over-rewarded while those in low-status occupations (e.g., unskilled factory workers) are seen as under-rewarded, researchers infer that the respondent thinks income inequality is too large (Evans & Kelley, 1993; Osberg & Smeeding, 2006; Kiatpongsan & Norton, 2014; Kuhn, 2019; Pedersen & Mutz, 2018).

Surveys implement this method by asking first how much the typical person in an occupation earns, then how much that person *ought* to earn. The (log-transformed) ratio of these quantities represents how fair or just the participant sees that income as being for that occupation.

$$\text{Occupational justice ratio} = \ln\left(\frac{\text{Perceived actual income}}{\text{Ideal income}}\right)$$

We can combine perceptions of ideal and actual CEO and unskilled factory worker incomes to signify whether a participant thinks inequality is too high, since

these two occupations represent either end of the socioeconomic scale. This and similar ratios (transformed or not) have been used in research on justice attitudes (Kelley & Evans, 1993; Kiatpongsan & Norton, 2014; McCall, 2013; Osberg & Smeeding, 2006; Schneider & Castillo, 2015):

$$\text{Just inequality ratio} = \ln \left(\frac{\text{Perceived actual CEO income} - \text{perceived actual factory worker income}}{\text{Ideal CEO income} - \text{ideal factory worker income}} \right)$$

CEOs and unskilled workers are the most frequently studied occupations using this measure, but surveys typically also include other occupations spread across the socioeconomic spectrum. The same principle applies to these other occupations: if participants believe that inequality, in general, is too high, they should be likely to report that low- and middle-status occupations, in general, have lower incomes than they ought to, and/or that high-status occupations have higher incomes than they ought to.

We can also measure perceptions of over- and under-reward using Likert items. We refer to this approach as the *occupational justice Likert* measure. Both survey and laboratory work on justice perceptions uses Likert scales that ask respondents how fair or just they think an income or reward level is (Auspurg et al., 2017; Clay-Warner et al., 2016; Melamed et al., 2018; Sauer & Valet, 2013; Schneider & Valet, 2017). Because ratio and Likert methods should measure the same construct, we include both in our study. Finally, as noted in the previous section, researchers also frequently use *Likert scales* to measure attitudes about whether inequality is too high, too low, or just right.

Method

We recruited a sample of 831 participants from Mechanical Turk to complete a short (mean = 14.5 min) study between March and April of 2019.[1] The study was described as a survey about how people perceive social issues. Participants who clicked on that task were then provided more description and a link to the study. We improved data quality in a variety of ways using the Turkprime (recently rebranded as CloudResearch) platform for mTurk (Litman et al., 2017; Litman & Robinson, 2021), including blocking suspicious geolocations and duplicate IP addresses, excluding the most active 10% of mTurk users, and including a randomly placed attention check which asked participants to select the response option "strongly disagree" to show they were paying attention.

---

[1] To recruit participants, we posted a task on mTurk's online marketplace, where 'Workers' (people from the general public with an mTurk worker account) can perform short tasks posted by 'Requesters' (including academic researchers, but also many commercial and industrial groups, e.g., companies crowdsourcing human responses for machine learning). Litman and Robinson (2021) provide full detail on Mechanical Turk. Overall, 54% of Workers who clicked on the description of our study then participated in the survey.

Table 1 Analytic sample demographics

Demographic	Mean (SD), median, or %
Age	35.2 (11.6)
Male[a]	48.7%
Non-White	19.4%
College degree	52.8%
Family income	$50 k-$60 k
Subjective SES	5.2 (1.7)
Conservatism	3.8 (1.7)
ANES party affiliation	3.8 (2.2)
Clinton voters	35.3%
Trump voters	32.4%

[a]Two participants identified as nonbinary. We include these participants in the reference category with those identifying as female

As explained below, a critical question in the literature is the extent to which there is a partisan divide in perceptions of—and preferences for—inequality. Given that the population of mTurk leans more liberal (Berinsky et al., 2012; Litman & Robinson, 2021), but is otherwise politically similar to the general population (Clifford et al., 2015), we slightly oversampled conservatives to create a more politically balanced sample.

Fifty (6.0%) participants failed the attention check and were excluded at the outset. We restricted our sample further after determining that 111 participants failed to provide rank-ordered responses to measures that require it. We describe this restriction and its implications further in our analysis. The final N was 680. Table 1 shows the demographics of our sample.

Before beginning the study, participants were told they would complete questions to assess their beliefs about inequality and social mobility.[2] In addition to providing participants with explanations of inequality,[3] social mobility, and income quintiles. We randomly assigned participants to complete three (out of eight) measures. We presented only a subset of measures to minimize participant exhaustion, a key concern given the cognitive demands of some of these measures (Eriksson & Simpson,

---

[2] We will report the results for social mobility measures in a different manuscript.

[3] We defined income inequality as "the size of the difference in income between people with the highest incomes in a society and people with the lowest incomes." While this definition does not satisfy all principles of inequality measurement, e.g., scale invariance (Hao & Naiman, 2010), and does not specify the nature of the population the participant should calculate inequality within, e.g., only full-time workers, similar phrasing used in the GSS variable eqwlth about "income differences between rich and poor" is typically understood as referring to attitudes about income inequality in general (McCall, 2013; McCall et al., 2017). We were also concerned with defining the concept in a way that parallels definitions in related work (e.g., Chambers et al., 2014; Eriksson & Simpson, 2012; McCall, 2013; Norton & Ariely, 2011), but which would not further create unnecessary cognitive difficulty. We therefore used this definition to activate the concept of income inequality.

Table 2 Descriptive statistics for inequality and inequality attitudes measures

Perceived inequality measures		Inequality attitudes measures		
Measure	Mean (SD)	Measure	Exp. mean	Mean (SD)
Norton–Ariely Gini	40.45 (13.77)	Perceived CEO/worker Ratio	36.58	3.60 (1.88)
Eriksson–Simpson Gini	46.39 (13.02)	Inequality justice ratio	2.52	.93 (1.43)
Income brackets Gini	54.24 (8.15)	Occupational justice ratios		
Stratification belief Diagrams Gini	32.73 (7.44)	CEO	2.39	.87 (1.39)
International Ranking[a]	5.82 (1.93)	Doctor	1.01	.01 (.66)
Perceived inequality scale	5.55 (1.05)	Manager	.94	− .06 (.42)
		Teacher	.64	− .44 (.47)
		Cashier	.69	− .37 (.51)
		Unskilled worker	.72	− .33 (.43)
		P's occupation	.69	− .37 (.88)
	%			Mean (SD)
Stratification belief diagrams		Occupational Justice Likerts		
Type A	23.7	CEO		1.65 (1.21)
Type B	36.7	Doctor		.46 (1.02)
Type C	19.8	Manager		.14 (.84)
Type D	16.9	Teacher		− 1.55 (1.20)
Type E	2.9	Cashier		− 1.02 (1.02)
Pie Charts		Unskilled Worker		− .89 (1.04)
Low	25.9	P's Occupation		− .90 (1.06)
Medium	17.9	Inequality attitudes scale		5.32 (1.28)
High	56.3			
International ranking				
US has highest inequality	63.4			

[a]Inequality rankings are reversed so that higher values represent more inequality

2012, 2013; Pedersen & Mutz, 2018). Because participants are randomly assigned to measures, we are able to estimate correlations between the measures using the sub-sample which completed each pair of measures. The N for each pair of measure is provided in the Supplementary Material (Section 1).

The eight measures were: Norton–Ariely, Eriksson–Simpson, Income Brackets, Stratification Belief Diagrams, Inequality Pie Charts, International Ranking, Occupational Justice Ratios, and Occupational Justice Likerts. Table 2 presents descriptive statistics for all measures. For several measures, we focus on the Gini coefficient to summarize perceived inequality. This is because the Gini correlated very highly with other summary statistics that we generated from the same data (e.g., quintile ratios; Supplementary Material Section 2), and because the Gini is commonly used and understood by social scientists (Hao & Naiman, 2010). To be clear, we do not argue that lay people intuitively understand the Gini, a point we elaborate on later.

Instead, we simply use these Ginis as an efficient way of presenting the tendencies of the other measures of variation we calculated. All Gini coefficients are scaled to range from 0 to 100. For justice ratios, we also report exponentiated means.

Since some measures provide ranked (or non-normal) data, our analyses primarily use Spearman rank order correlations (ρ). This is a nonparametric, nonlinear measure of association interpreted as the correlation between rank scores rather than the scores themselves. This makes ρ robust to curvilinearity so long as the relationship between two variables is non-decreasing (for positive relationships) or non-increasing (for negative relationships). The more continuous two variables are, the more closely ρ approximates the Pearson correlation, r. Given our relatively large sample size and the number of pair-wise comparisons, we focus our analysis on trends in the magnitude of these correlations. The Supplementary Material file provides statistical significance for all correlations.

Inequality Measures

In the Norton–Ariely measure, participants were asked to provide the percentage of all income that goes to each of five income quintiles in the USA. We specifically asked participants to think about individual incomes before taxes, not family incomes. The Eriksson–Simpson measure was similar, but instead of asking for percentages, it asked participants to provide the average income of people in each quintile. In both measures, participants were explicitly reminded that quintiles are rank-ordered.

For the income brackets measure, we asked participants to estimate the percentage of Americans with incomes in each of five income ranges. We took our ranges from the tax code for single filers and rounded to the nearest 10,000: less than $10,000, between $10,000-$40,000, $40,000-$80,000, $80,000-$160,000, and more than $160,000. Since few people earn more than $160,000, we used that as our top category. We converted these proportions into Gini coefficients by midpoint scoring each income bracket and setting the score of the top bracket to $1.5 \times$ its lower bound (or $240,000 for the $> \$160,000$ response). Because the value of the top income range could be set even higher, this is a conservative Gini estimate. Larger scores for the top bracket would yield even higher Gini estimates.

Participants completed the Stratification Belief diagrams measure by selecting the diagram which they believe comes closest to the actual level of stratification in the USA. In addition to analyzing the stratification belief diagrams as ranked categories, we also use their Gini equivalents, which Gimpelson and Triesman (2018) calculated as $A = 42$, $B = 35$, $C = 30$, $D = 20$, and $E = 21$.

Participants followed the same procedure for the inequality pie charts, selecting the one which most closely represents how they see the USA. We presented three pie charts (Fig. 2): one depicting relatively low inequality (Gini $= 35.5$), one depicting relatively high inequality (Gini $= 42.9$), and one in between (Gini $= 37.9$).

Fig. 2 Inequality Pie charts

Participants were presented with the countries that make up the G7 in random order and were asked to re-rank them from the highest inequality (1) to the lowest (7). In our analyses, we reverse this so that higher values represent higher inequality.

Participants tended to perceive inequality as high across all measures (Table 2). Mean Gini coefficients from the Norton–Ariely, Eriksson–Simpson, and income brackets measures were all above 40. Most participants completing the stratification belief diagrams rated the USA as either a Type A or B country—the two most unequal options. More than half chose the high inequality pie chart as representing the USA. And nearly two-thirds ranked the USA as having the highest inequality of the G7.

Justice Measures

To obtain income ratios, we asked participants to provide what they saw as the ideal income for each of seven occupations, and what they thought the actual average income was. We measured perceived actual and ideal incomes for two high-status occupations (chairman of a national corporation, doctor in general practice), two middle-status occupations (mid-level manager in a national corporation, high school teacher), two low-status occupations (unskilled factory worker, clerk [cashier]), and the participants' own occupation ('people in my occupation'). The high- and low-status occupations come from the GSS, ISSP, and related analyses (e.g., Kiatpongsan & Norton, 2014; McCall, 2013; Osberg & Smeeding, 2006).

To measure these justice perceptions less quantitatively, we also administered two parallel Likert scale items for the same occupations: we asked participants whether they felt that the typical income in that occupation was less than fair (-3), exactly fair (0), or more than fair (+3). Research on justice perceptions often uses a single item on fairness to summarize those perceptions (Auspurg et al., 2017; Clay-Warner et al., 2016). However, to capture the aspect of justice in which the under-rewarded deserve more while the over-rewarded deserve less (Hegtvedt & Isom, 2014), we also asked whether the typical income was less than what people in that occupation deserve (-3), exactly what they deserve (0), or more than they deserve (+3). We then

averaged the two items together. Reliabilities were somewhat low (mean $\alpha = 0.62$), but substantive conclusions of our analyses are the same whether items are analyzed together or separately. We therefore present analyses for the averaged items.

Participants perceived a large amount of injustice (Table 2). On average, they thought that CEOs earn 36 times as much as an unskilled factory worker, and that the inequality between them is two and a half times larger than it ought to be. On both ratio and Likert measures, they see CEOs as significantly over-paid, while most other occupations are under-paid.

Criteria Likert Scales

We also administered a questionnaire including our *Perceived Inequality* and *Inequality Attitudes* scales. We constructed these scales because most existing measures of inequality attitudes assess some evaluative or normative attitude (e.g., the GSS asks if income differences are *too* large), rather than simply having respondents indicate whether they perceive inequality as high or low. If other measures also assess the same constructs, they ought to correlate strongly with the Likert scales. A weak correlation, on the other hand, implies that the measure is assessing something that is only weakly related to that construct.

The Perceived Inequality Scale used four 7-point Likert items. Examples include "In your judgment, how large or small are the differences in income between the rich and the poor in the USA? (very small–very large)", "and "Compared to other developed countries, income inequality in the USA is (much lower–much higher)." The scale was reliable ($\alpha = 0.78$). The Inequality Attitudes Scale used seven 7-point items. Examples include "Income inequality in the USA is too high", "Most Americans would be better off if income inequality was lower", and "In the USA, wealthy people's incomes are higher than they should be." This scale was very reliable ($\alpha = 0.91$). Importantly, both scales include different conceptions of inequality and justice, e.g., the Perceived Inequality scale includes items comparing US inequality to the past and to other countries; the Inequality Attitudes scale includes items asking about individual interest as well as the justice of the incomes of different social classes. The correlations among the questionnaire items thus reflects similarity between these somewhat different views on inequality and justice. We will return to this point later in our analysis. Further detail on these scales is provided in Supplemental Material (Section 1). Table 2 shows that on the two Likert scales participants strongly saw inequality as high (Perceived Inequality mean = 5.55) and said that lower inequality would be better (Inequality Attitudes mean = 5.32).

Main Questions

Our analyses focus on several key questions. First, how well do estimates of inequality or justice attitudes correlate with other subjective measures? It seems unlikely that participants conceive of inequality or justice in precisely the terms that these measures do. Rather, their perceptions and attitudes are latent beliefs,

and the measures are all aimed at assessing those latent beliefs. Likert scales embody this reasoning, since they are composed of multiple indicators of the same latent construct. Thus, if the subjective measures in this study are good indicators of perceived inequality or justice attitudes, they should correlate with other subjective measures and with the Likert scales we constructed.

Second, we wanted to assess the size of method effects for inequality measures. That is, do the different measures of perceived inequality create differing estimates of inequality simply because they use different methods? For example, Eriksson and Simpson (2012) found that participants gave significantly higher estimates of inequality using their measure compared to the Norton–Ariely measure. In our study, we compare the Gini estimates from several measures to show how widely the mean estimates vary. Some degree of variation between measures seems quite likely, given that each measure is to some extent unique and likely comes with its own set of biases. However, it is an important question to ask because of the common practice in the literature on perceived inequality of comparing subjective perceptions against objective economic conditions (see Hauser & Norton, 2017 for a review). If different measures provide widely varying Gini estimates, then researchers' conclusions about participant 'misperceptions' of inequality may simply be an artifact of the method that they use to measure those perceptions. This, in turn, would make it difficult to justify comparing subjective estimates against objective statistics.

Third, we test how strongly the various measures relate to participants' political orientations. One of the reasons Norton and Ariely's (2011) seminal study gained wide attention was because it found that liberals and conservatives perceived—and preferred—similar levels of inequality. But a number of scholars have questioned these findings, given the major divides based on inequality and justice in American politics (Chambers et al., 2014; Eriksson & Simpson, 2012, 2013; Pedersen & Mutz, 2018). Research finds that conservatives believe higher levels of inequality to be fairer and prefer to have more inequality than liberals do (Cech, 2017; Kelley & Evans, 1993; McCall, 2013; Pedersen & Mutz, 2018), but some research also suggests that they perceive less inequality to begin with (Chambers et al., 2014; Eriksson & Simpson, 2012). Conservatives also have much lower preferences for economic redistribution, which itself tends to correlate with perceived inequality (Ashok et al., 2015; Gimpelson & Treisman, 2018; though see Schröder, 2017; Trump, 2017 on whether redistributive preferences are related to actual inequality). We thus expect that political orientation will correlate with justice measures, and with measures of perceived inequality.

Finally, we examine our results for evidence of low comprehension and heuristics known to bias estimates of inequality. Many of our measures ask participants to quantify perceptions or preferences, and previous research suggests that many participants do not understand these tasks. This leads them to rely on heuristics, e.g., anchoring income shares per quintile on 20% (Eriksson & Simpson, 2012, 2013) and anchoring 'just' incomes on whatever they perceive as the 'actual' income (Markovsky & Eriksson, 2012; Pedersen & Mutz, 2018). The cognitive intensity of our measures means that we should attend closely to whether participants are giving clear or logical responses.

Results

To simplify our analysis, we begin by detailing problems we found with particular measures that are large enough, in our view, to recommend not using those measures in future studies. We then proceed through several major problems that should be of interest to researchers looking to use any of the measures we analyze. The questions we focus on are not the only ones that can be addressed with our data. Given space limitations, we focus on what we consider particularly important issues for improving the methods and measures used to study inequality perceptions and attitudes. Full detail on our analyses is provided in the Supplementary Materials (Section 2). Further analyses (SM Section 3) compare results using bivariate linear regressions against regressions with demographic controls. The two approaches yield very similar results. Our findings are thus robust to demographic controls, strengthening our conclusions.

Problem 1: Many Participants Fail to Provide Rank-Ordered Responses When They are Required

The Norton–Ariely and Eriksson–Simpson measures require that participants understand that income shares (or mean incomes) must be rank-ordered. That is, participants must provide estimates such that the top quintile holds more income than the fourth quintile, which holds more than the third, and so on. However, many participants did not provide fully ranked responses: *nearly a third* ($N = 85$, 29.2%) of those completing the Norton–Ariely measure, and about one in sixteen ($N = 18$, 6.2%) of those completing the Eriksson–Simpson measure failed to give rank-ordered responses. This is not simply a result of some participants reversing the order of quintiles (providing the lowest incomes for the 'top' quintile, the highest incomes for the 'bottom', etc.): only 11 participants in the Norton–Ariely measure provided reversed rankings in this manner. Eight more provided equal-sized income shares, and each of the remaining quintile orderings had no more than seven observations each. These problems occurred even though we explicitly reminded participants that the income quintiles should be ranked.

Failing to provide ranked responses on at least one measure was associated with lower responses on the Perceived Inequality ($b = -0.42$, SE $= 0.12$, $p < 0.001$) and Inequality Attitudes ($b = -0.34$, SE $= 0.14$, $p < 0.05$) scales. This means the inequality estimates from the analytic sample might have an upward bias because people who perceive inequality as lower are more likely to be excluded for not rank-ordering responses. Non-random ranking errors thus undermine the reliability of the measure.

Importantly, failure of participants to give rank-ordered responses to the Norton–Ariely measure is not unique to our dataset. Both Norton and Ariely (2011) and Eriksson and Simpson (2012) found that participants failed to rank order, and resolved this issue by reordering non-rank-ordered responses so that the responses reflected a more logical quintile order (see Eriksson & Simpson, 2012:742 for a discussion).

We see two possible (and related) explanations for participants' failure to rank responses. First, participants may not have understood that they must provide rank-ordered responses, despite the fact that we provided them with an explanation of the concept of ranked quintiles prior to the start of the study, and reminded them of this requirement in the measure itself. Lay people are not broadly aware of statistical concepts like quintiles, and therefore are not accustomed to using them to organize their perceptions of society. A second, related possibility is that participants find these estimation tasks cognitively demanding. This would likely lead them to resort to heuristics (Eriksson & Simpson, 2012) or to simply give up on providing consistent (ranked) responses.

Cognitive demands would explain why non-ranking is five times more likely when using the Norton–Ariely measure than the Eriksson–Simpson measure: while most participants should be familiar with the notion of income, the concept of an 'income share' is likely more novel and likely to provoke non-ranked responses (see Eriksson & Simpson, 2012).

Given these problems, we caution researchers against using the Norton–Ariely method to study inequality perceptions. While the Eriksson–Simpson measure resulted in far fewer non-rank-ordered responses, there were still enough to merit caution. As noted in the Methods section, we excluded from further analysis all participants who failed to provide rank-ordered responses to the Norton–Ariely and Eriksson–Simpson measures. Doing so allowed us to assess the relationship of other our measures with the Norton–Ariely and Eriksson–Simpson measures without the results being skewed by poor quality data. The total remaining N in all remaining analyses is 680.

Overall, the Norton–Ariely measure should be ruled out for future use. Since the Eriksson–Simpson measure performs better in terms of rank-ordering, it might be used as a replacement, but we nevertheless caution researchers to take additional steps to ensure participant comprehension. Since rank-ordering is the minimal logical requirement for these measures, non-ranking does not bode well for participants' ability to provide logical quantitative estimates that reflect their inequality perceptions.

Problem 2: When Allowed, Participants Often Provide Extreme Income Estimates

A second problem emerged in our analysis of the Eriksson–Simpson measure: participants often gave estimates of typical top incomes that strained credulity: 9.6% of participants put the mean top quintile income at more than $10 million, 5.6% estimated more than $100 million, and the highest estimate was $50 *billion*. Similarly skewed patterns existed for other quintiles.

One possible explanation is that some participants added or dropped digits to convey their sense of extremity. This is supported by the fact that extremely high responses were more common for estimates of top incomes than for lower quintiles, and they were associated with higher responses on other measures of perceived inequality. Participants who provided at least one very high mean income (higher than $10 million) also gave higher responses on our Perceived Inequality ($b = 0.45$,

SE $= 0.23$, $p < 0.05$) and Inequality Attitudes ($b = 0.81$, SE $= 0.26$, $p < 0.01$) scales than those who completed the Eriksson–Simpson measure but did not provide any very high incomes. They also gave higher inequality estimates on the stratification belief diagrams ($b = 6.92$, SE $= 2.61$ $p < 0.05$) and perceived CEO/worker pay ratio ($b = 2.15$, SE $= 0.61$, $p < 0.01$) measures, though not the Norton–Ariely ($b = 1.47$, SE $= 5.59$, $p = 0.793$) or income brackets ($b = -2.81$, SE $= 5.38$, $p = 0.603$) measures.

On the one hand, a measure of perceived inequality ought to capture perceptions that inequality is extremely high. But extreme estimates of top incomes may cause other problems. Most obviously, they can skew the variable's distribution, complicating statistical analyses. But perhaps more importantly, the numerical response alone does not tell us whether it comes from participants' perceptions, indignation about high inequality, or innumeracy. Researchers using the Eriksson–Simpson should therefore use additional measures to determine whether they should take these responses "seriously but not literally." That is, researchers should pay attention to the variation in responses to the Eriksson–Simpson measure, and how those responses covary with other phenomena. But the issues outlined above show that the point estimates of inequality derived from the measure may not be as meaningful.

Problem 3: Responses Across Measures Are Often Inconsistent

Our design allows us to compare whether participants' responses were consistent (i.e., provided similar descriptive statistics or otherwise correlated highly) across measures, giving insight into how results from one measure may generalize to others. Of course, variation between measures is a given in most social research (Thye, 2000). But studies of perceived inequality and justice are somewhat atypical, in that they often compare participants' estimates against "reality," i.e., objective measurements obtained by social scientists (Chambers et al., 2014; Eriksson & Simpson, 2012; Gimpelson & Treisman, 2018; Hauser & Norton, 2017; Kuhn, 2019; Norton and Ariely 2011; Kiatpongsan & Norton, 2014). The emphasis on such precise comparisons means that method effects could affect substantive conclusions, e.g., that people overestimate vs underestimate inequality (Chambers et al., 2014). At the very least, method effects can affect the strength of estimation biases (Eriksson & Simpson, 2012). Thus, it is especially critical in this research area to understand how the measurement strategy affects participant responses. Supplementary Material Section 2 provides rank order correlations between measures.

We found that outcomes varied widely between measurement methods. For instance, mean Gini estimates ranged from a low of 32.7 (for the stratification belief diagrams) to a high of 54.2 (for the income brackets measure). Estimated Ginis had low correlations with each other (mean $\rho = 0.33$, five of six correlations significant at $p < 0.05$) and have slightly weaker correlations with our Perceived Inequality scale (mean $\rho = 0.26$, three of four correlations significant at $p < 0.05$). The Norton–Ariely measure tends to have the strongest correlations with other measures, but it is important to remember that these correlations exclude the third of participants who did not provide correctly ranked responses. This significantly undermines the finding that the remaining data correlate well with other measures. Overall, method

Table 3 Percentage of participants selecting closest option on pie charts and stratification belief diagrams	% Selecting closest option on…	
	Pie charts	Stratification belief diagrams
Norton–Ariely	48.4	39.0
Eriksson–Simpson	42.1	24.7
Income brackets	50.0	21.2

effects on Gini estimates appear to be substantial, and Ginis from one measure fail to correlate well with Ginis from other measures.

We also tested whether participants tended to select the inequality pie chart or stratification belief diagram closest to their responses on the Norton–Ariely, Eriksson–Simpson, and income brackets measures. To test this, we first grouped Gini estimates from these three measures according to whether they were closest to the low, medium, or high inequality pie chart, and to each of the five stratification belief diagrams. We then tabulated the proportion of participants who selected the pie chart or diagram that had the Gini coefficient closest to their own responses. As shown in Table 3, across the Norton–Ariely, Eriksson–Simpson, and income brackets measures, less than half of participants selected the inequality pie chart that was closest to the Gini generated from their responses and just 20–40% selected the closest stratification belief diagram. Overall, no measure gives clear associations with parallel graphical measures.

However, given that our quantitative measures tap into somewhat different aspects of inequality and attitudes about inequality, we should not expect them to correlate perfectly with each other. We therefore assessed consistency another way, by comparing the correlations among measures against the inter-item correlations from our criteria scales. The items in these Likert scales individually measure distinct aspects of perceived inequality and related attitudes, but as with any set of items in a scale, they correlate with each other because they also tap into a shared, latent conception of income inequality. If the quantitative measures also tap into this shared concept, they should also correlate to a similar extent.

The mean inter-item correlation in the Perceived Inequality scale was $\rho = 0.55$, while the mean absolute correlation between perceived measures was less than half that magnitude, $\rho = 0.24$ (nine of 21 correlations significant at $p < 0.05$; SM Table 2.2.1 for more detail). As a group, our perceived inequality measures are much less consistent than the items of a parallel Likert scale.

Participants also provided inconsistent responses to justice ratios and our Inequality Attitudes scale. On the inequality justice ratio, nearly a third (29.3%) of participants gave responses indicating that they thought inequality was too *high* but responded to the Inequality Attitudes scale saying inequality was too *low*, or vice versa. This replicates Pedersen and Mutz's (2018) finding that participants often give inconsistent responses to justice ratio measures and GSS measures of inequality attitudes.

Comparisons between justice ratios and Likert scales (SM Section 2) tended to fare better, but still lacked convergent validity; that is, participants' responses to different measures suggest that they had contradictory implications for their perceptions of occupational pay justice. Correlations between the two formats within occupations tended to be somewhat low (mean absolute $\rho = 0.24$, SD = 0.17, 21 of 56 correlations significant at $p < 0.05$) and participants often provided inconsistent responses: in total, 24.6% of responses to one measure were associated with inconsistent responses to the other, e.g., providing a CEO income ratio indicating over-payment while providing responses to the Likert items indicating either underpayment or exact income fairness. Participants were more likely to rate incomes as exactly fair on Likert measures than justice ratios (28.2% vs 20.6%).

Here again, we compared the consistency of justice measures against the mean inter-item correlation of our Inequality Attitudes scale. Correlations among justice measures tended to be low: ratio measures correlated at a mean absolute $\rho = 0.29$ (16 of 34 correlations significant at $p < 0.05$), while Likert measures correlated at a mean absolute $\rho = 0.27$ (14 of 21 correlations significant at $p < 0.05$; SM Table 2.2.2 for more detail). Both are much lower than the mean inter-item correlation from our Inequality Attitudes scale, mean $\rho = 0.60$. Thus, as with perceived inequality measures, justice measures as a group tend to be less consistent with each other than the items of a parallel Likert scale.

Inconsistency across measures appears to be high and ubiquitous, but becomes less surprising when we remember how cognitively demanding these tasks are. When faced with tough questions, especially numerical ones, people often rely on heuristics (Eriksson & Simpson, 2012, 2013; Markovsky & Eriksson, 2012; Pedersen & Mutz, 2018). Many participants may not know that responses on one measure can be converted into an equivalent response on another. Even if they are aware of this fact, they may not know how to perform that calculation even if they want to. Indeed, many of Eriksson and Simpson's (2012) participants admitted, when asked, that they were unaware there was any mathematical equivalence between the Norton–Ariely and Eriksson–Simpson procedures. Overall, the measures we studied were inconsistent with each other at a level beyond that which we might expect and accept from items in a reliable Likert scale. This higher level of inconsistency may be due to the cognitive demands that these measures impose.

Problem 4: Correlations with Criteria Likert Scales Vary, But Are Generally Low

We designed our Perceived Inequality and Inequality Attitudes scales to measure the same constructs as our key measures in a simpler, less cognitively demanding format. If our key measures tap into participants' beliefs, then they should correlate reasonably well with these criterion scales. However, most of the observed correlations between the scales and measures are small.

The average correlation between measures of perceived inequality and our Likert scale measure is $\rho = 0.29$ (SD = 0.05; all p's < 0.05), with the highest (the pie charts measure) being $\rho = 0.37$ (SM Section 2). The relatively low correlations across the board (lower than the inter-item correlations of the items composing the Perceived

Inequality scale, mean $\rho = 0.55$, SD $= 0.09$) means that perhaps the convergent validity of our tested measures is low across the board, and each measure in fact taps into somewhat different aspects of perceived inequality.

A similar pattern exists for justice measures. Correlations between justice measures and the Inequality Attitudes Scale are low, but quite variable, mean absolute $= \rho$ 0.32 (SD $= 0.16$; 12 of 16 correlations significant at $p < 0.05$). For occupation-specific measures, doctors and managers appear to be virtually irrelevant to attitudes about inequality (absolute ρ's ≤ 0.10, p's > 0.05). But even the strongest correlation, from the cashier income justice ratio, $\rho = -0.49$, is lower in absolute magnitude than the mean inter-item correlation within the Inequality Attitudes Scale (mean $\rho = 0.60$, SD $= 0.09$).

One potential reason why these measures have generally low correlations with criteria is because they may be more strongly influenced by random measurement error, and so are less reliable. Our results thus far have pointed to several different sources of such error, including low participant comprehension, the use of heuristics, and the cognitive complexity of quantitative estimation. It is also possible that the measures are each tapping into very different aspects of inequality that are not clearly related to each other, like income inequality within America vs America's international ranking in income inequality. However, we cannot distinguish between this possibility and the chance that low correlations are simply due to low reliability.

In short, the correlations between our measures and criterion scales are generally not large enough take any specific measure as a clear or strong indicator of participants' beliefs about the American economy. We therefore caution researchers against using any single measure to capture Americans' perceptions about the level of inequality or whether they think that level is objectionable.

Problem 5: Most Measures Overestimate Partisan Agreement

Researchers have recently debated whether there is partisan disagreement over perceptions of—and attitudes about—inequality. Some find that liberals and conservatives perceive similar levels of inequality or injustice (Kiatpongsan & Norton, 2014; Norton & Ariely, 2011). But others argue that this 'agreement' is illusory, a product of asking participants questions that they do not understand (Chambers et al., 2014; Eriksson & Simpson, 2012, 2013; Pedersen & Mutz, 2018). Given the salience of inequality and justice issues in American politics (e.g., Cech, 2017), we argue that prior to measurement, researchers should strongly expect perceived inequality and justice to correlate with political orientation. Apparent partisan agreement, then, is an outcome that should be investigated critically.

We found that for most measures (SM Section 2), conservatives tended to perceive somewhat less inequality than liberals and judge inequality as more just (mean absolute $\rho = 0.20$, SD $= 0.06$, six of seven correlation significant at $p < 0.05$). But on average, these correlations were much smaller than correlations between political orientation and our Perceived Inequality ($\rho = -0.39$) and Inequality Attitudes scales ($\rho = -0.52$). For our justice measures, the average correlation was slightly larger (mean absolute $\rho = 0.24$, SD $= 0.13$, 11 of 15 correlations significant at $p < 0.05$),

but still not as large as the Perceived Inequality or Inequality Attitudes scales. Justice attitudes for doctors and managers had the weakest association with political orientation (mean absolute $\rho = 0.05$, all $ps > 0.05$), while the other justice measures correlated between mean absolute $\rho = 0.20–0.40$. Overall, no measure correlated with political orientation as well as our Likert scales.

This result is suggestive, but could also simply reflect the fact that political orientation was measured using a similar 7-point Likert scale as our criteria. To assess this possibility, we conducted a second analysis using a measure of political affiliation adapted from the American National Election Studies (ANES). This measure provides political affiliation on a 7-point scale, from Strong Democrat to Strong Republican. But, for this measure, responses are gathered through two items: a first question asks participants whether they feel closest to Democrats, Independents, or Republicans, and a second question follows up and asks if they identify as a 'strong' Democrat or Republican, or if they identify as an independent, whether they lean toward one party or the other. This provides a measure of a construct strongly related to political affiliation, but without the same potential common method bias.

Here again, we find that our criteria scales correlated better with political affiliation than the perceived inequality and justice measures. While the Perceived Inequality scale correlated with political affiliation at $\rho = -0.28$, no perceived inequality measure correlated above absolute $\rho = 0.20$ (mean absolute $\rho = 0.12$, SD $= 0.06$, four of seven correlations significant at $p < 0.05$). The Inequality Attitudes scale correlated with political affiliation at $\rho = -0.39$. No occupational justice measure exceeded this level of correlation, and tended to correlate less than half as strongly (absolute mean $\rho = 0.16$, SD $= 0.10$, 10 of 15 correlations significant at $p < 0.05$). Thus the fact that our criteria scales correlate better with political views than our other measures does not appear to be driven by common method bias.

Why wouldn't measures of these perceptions and attitudes also be related to political orientation given their political salience? In addition to response heuristics (Eriksson & Simpson, 2012, 2013; Markovsky & Eriksson, 2012; Pedersen & Mutz, 2018), the comprehension issues we observe suggest another explanation—progressives and conservatives may *appear* to agree when comprehension is low on both sides of the political aisle. In short, when researchers discover that a measure appears to generate political 'consensus' on issues of inequality and justice, they should review their data and methods carefully to rule out the possibility that it is due to low comprehension or to the use of heuristics.

Problem 6: 'Ideal' Incomes Likely Anchor on Perceived Actual Incomes

Some scholars (Markovsky & Eriksson, 2012; Pedersen & Mutz, 2018) have argued that respondents tend to base, or anchor, 'ideal' incomes on their perceptions of actual incomes, whether they provide those actual incomes themselves or they are provided by the researcher. This casts doubt on any assumption that people carry ideals for what people ought to earn that are independent of what they believe they actually make.

Though we did not manipulate perceived actual incomes, we do find that the perceived CEO/worker income ratio correlates with the ideal CEO/worker ratio[4] at ρ = 0.59 ($p < 0.001$). That is, the more unequal participants thought incomes were, the more unequal they wanted them to be. A portion of this correlation likely stems from participants' actual beliefs: people often base their ideas of what 'ought to be' on what they think already 'is', especially when they benefit from the status quo (Homans, 1974; Schröder, 2017; Trump, 2017). But both Markovsky and Eriksson (2012) and Pedersen and Mutz (2018) show that it is also a matter of in-the-moment anchoring on potentially irrelevant information.

Summary

We found a range of problems across the different measures: low comprehension, inconsistency, low correlations compared to parallel Likert scales, and anchoring. In terms of the severity of criticisms, the Norton–Ariely and Eriksson–Simpson measures fared the worst. There, participants did not give ranked responses (particularly on the Norton–Ariely measure) or gave extreme responses (particularly on the Eriksson–Simpson measure). Even when we excluded problematic responses, these measures did not perform notably better than others in terms of consistency across measures, correlation with criteria, etc.

The stratification belief diagrams and international ranking measures fared the best. They had among the best correlations with our perceived inequality scales and offer a good range of response options. But even these measures come with important caveats. First, it is not clear that we could compare them against objective data on inequality since we also found significant method effects. Even a reliable measure of perceived inequality could plausibly show that most participants overestimate inequality *or* that most underestimate inequality, depending on how the method itself biases the distribution of estimates.

Second, the diagrams and rankings may appear to fare better in part because they are less able to betray low comprehension through extreme or illogical responses. That is, if many participants do not understand the Norton–Ariely or Eriksson–Simpson measures, we will know it from their responses. Indeed, some of the more positive outcomes for the Norton–Ariely measure, including greater correlations with other measures, may be due to the fact that we excluded so many invalid responses to that measure, which filters responses based on comprehension. However, we will not observe similar outcomes on other measures because they do not allow patently illogical responses in the same way. For instance, it is plausible that any given respondent could place the USA at any point in the international ranking and genuinely believe that this is where the USA stands—no rank is illogical or

---

[4]

$$\text{Ideal inequality ratio} = \ln \left(\frac{\text{Ideal CEO income}}{\text{Ideal factory worker income}} \right).$$

implausible. These alternative measures should therefore be packaged with comprehension checks to detect comprehension problems.

And third, the international ranking may have limited usefulness because it does not measure the same perception as the other measures. It compares US inequality against other countries, but it is not clear that Americans consider (or care about) other countries when judging US inequality as 'high' or 'low.'

We have organized our analysis around core problems that researchers studying inequality perceptions and attitudes should be aware of. Next, we move from these problems to advice for future research on perceptions of inequality and justice.

Discussion

Our results show that measuring lay perceptions of inequality is no easy task. And perhaps that should not come as a surprise. The magnitudes involved with inequality in a country like the USA can stymie professional social scientists, who spend considerable time measuring and interpreting inequality. It therefore seems unreasonable to expect lay people to be able to clearly quantify their perceptions of inequality. Nevertheless, understanding lay perceptions and attitudes is important for both social science and public policy. Our study took stock of various measures of perceived inequality and inequality attitudes, with the goal of improving research in this important and rapidly developing area. In so doing, we have documented a number of issues with current methods and measures. Our findings lead us to offer three prescriptions for researchers measuring perceptions of inequality and justice.

Use Measures That Participants Can Understand and Answer

Recent studies often ask respondents to put their thoughts or beliefs about inequality into (usually quantitative) terms that may be familiar to professional social scientists, but less so to laypeople. The appeal is clear: it allows the researcher to directly compare participants' judgments against objective measures of inequality to assess what extent respondents over- or underestimate inequality. Far from being confined to academic circles, the results of these studies have drawn attention from news media (e.g., Klein, 2013), making it especially critical that researchers exercise caution when deciding which methods to use and what conclusions to draw from them.

To ensure the validity and reliability of their findings, researchers need to ensure that participants understand the questions they are being asked and how to relate those questions to their beliefs. Lay people are unlikely to think about inequality using quantitative concepts like distributions, quantiles, ratios, and Gini coefficients that may be second nature to social scientists. Thus, while it may be straightforward for social scientists to quantify their beliefs about inequality, it is extremely unlikely that this ability is common to lay people.

Our findings lay bare comprehension problems for several of the main measures of perceived inequality. We can see this most clearly with the Norton–Ariely and Eriksson–Simpson measures, where many participants fail to give rank-ordered

responses (especially so for the Norton–Ariely measure). Rank-ordering is a defin-ing characteristic of those measures, and if that many participants fail to adhere to that requirement, it seems likely that many other participants did not fully under-stand the meaning of the measure or how they should relate it to their beliefs.

We also found that, for the Eriksson–Simpson measure, participants provided extremely high top income estimates, with some responses straining credulity. Though participants may mean these estimates literally, or at least as a meaningful expression of extremity, it is also possible that they did not fully understand the task or how to relate it to their attitudes. At the very least, these responses challenge any easy interpretation of results.

In a task as cognitively intense as quantifying subjective beliefs, it is inevitable that some participants will not fully understand their task. Given this, it is critical to measure comprehension rates. Researchers could do so by asking participants to correctly describe the task and what it means. Another option is to include (where possible) clearly implausible response options, like a pie chart or diagram showing complete equality of income.

Our study also joins previous work arguing against apparent agreement between liberals and conservatives on perceived inequality and inequality attitudes (Eriksson & Simpson, 2012, 2013; Markovsky & Eriksson, 2012; Pedersen & Mutz, 2018). American liberals and conservatives are heavily divided on these questions (Cech, 2017), and apparent agreement is more likely to mean that both liberals and con-servatives are using the same heuristics or are similarly confused about how to inter-pret the question, rather than sharing similar attitudes about inequality.

Researchers can take a mixed-methods approach to ensuring that measurement methods relate to participants' ways of understanding inequality. Researchers could draw on qualitative techniques to more fully understand how participants under-stand measurement methods and whether those methods tap into the ideas that are more central to their ideas about inequality (Small, 2011). Researchers can then use participant feedback to make their measures more comprehensible and valid to participants.

A recent example of this method comes from Cheng and Wen's (2019) study of social mobility perceptions. In addition to gathering data on a novel method for quantifying perceived mobility, they asked participants to describe how they think about mobility. This text feedback from participants showed that participants did, in fact, tend to think about mobility in ways similar to the researchers' mobility meas-ure, thus providing support for the measure's validity.

Use Multiple Measures to Make Conclusions More Reliable

We found that different ways of measuring perceived inequality and justice attitudes can yield dramatically different perceptions and attitudes. Researchers should there-fore not use just one measurement strategy to assess perceived inequality or justice attitudes. This leaves researchers unable to disentangle findings that are common across operationalizations from those that may be artifacts of a specific meas-ure. Indeed, additional analysis (SM Section 5.2) found that while all inequality

measures together explained more than a quarter of the variance in our perceived inequality scale, each measure individually explained only about 7.5%. Researchers can therefore make a stronger case for their conclusions if different kinds of measures give convergent findings.

This recommendation obviously isn't novel; researchers are already familiar with the need to use scales comprised of multiple indicators to make measurements more reliable (Thye, 2000). Measures of perceived inequality and justice attitudes are no exception. For instance, if researchers administered the Norton–Ariely and stratification belief diagram measures and found that higher perceived inequality was in both cases associated with some attitude like higher redistributive preferences, this would bolster the argument that higher perceived inequality is associated with redistributive preferences much more than if they found that outcome with just one of those measures. The use of multiple measures would also help to reduce the likelihood that conclusions regarding point estimates or estimation biases are driven by method effects.

For example, McCall (2013) analyzed numerous measures from the GSS, including Likert items measuring inequality attitudes, as well as perceived income and justice ratios. The variety of measures McCall analyzed brought depth and complexity to her account of Americans' inequality attitudes. And by testing her account against multiple attitude measures, she made a remarkably strong case for her argument about the circumstances under which Americans accept or reject inequality.

Researchers can also use analysis techniques to statistically combine data from different types of measures. As a proof-of-concept, we constructed CFA models with techniques for data with missing values to estimate the correlations between our criteria (the Perceived Inequality and Inequality Attitudes scales, political orientation, and political affiliation) and the latent factors underlying the more continuous measures of perceived inequality and occupational justice. The perceived inequality and justice factors correlated with criteria much better than the individual measures did on their own. Supplementary Information Section 4 provides full detail for this analysis. This analysis shows that combining different measures together using techniques like CFA or principal components analysis helps researchers measure these perceptions more reliably, recovering the useful information from each measure while making analyses more robust to the more unique aspects of each.

Take Steps to Avoid Reifying Measures

Many of the measures we reviewed can seem quantitatively precise, which invites comparisons to actual economic statistics and conclusions about whether participants understand these conditions accurately. We argue that this impulse is misguided as often as not. We reported ample evidence that responses vary depending on the way in which the researcher asks the question. And since lay people are unlikely to think of inequality primarily in quantitative terms, we argue that researchers should avoid reifying such measures. That is, researchers should avoid allowing their measures to be interpreted as uncomplicated, error- or bias-free, or 'true' representations of their participants attitudes.

One simple solution, if the data are approximately normal, is to standardize responses as *z*-scores. This would make scores more clearly relative to the distribution of responses rather than the original scale. For example, standardizing Gini coefficients would preserve the information they contain about participants' perceptions of inequality while discouraging one-to-one comparisons to Gini estimates in the actual economy. The exact statistics these measures provide may not be useful as literal estimates because of method effects or heuristic use, but the information they contain about the distribution of attitudes may still be recovered.

Researchers can also make the response distribution itself the subject of study, as done by Kelley and Evans (1993) and Osberg and Smeeding (2006) using tools like kernel density analysis. This strategy foregrounds distributional characteristics and allows the researcher to focus on questions (e.g., whether attitudes are multimodal) that are otherwise difficult to express through a focus on mean perceptions.

Another solution is to focus more on studying the determinants and effects of inequality perceptions and justice attitudes, as well as their longitudinal trends, than on the quantitative estimates themselves. For causal questions, (e.g., those examined by Clay-Warner et al. (2016), Kelley and Evans (1993) and Schneider and Castillo (2015)), the point estimates of those variables are less important than their statistical association with other important variables. And in longitudinal analyses (e.g., McCall, 2013), particular point estimates are less important than changes in responses over time. Longitudinal analyses also make method-specific biases less relevant, since those biases should be constant over time and thus should not affect temporal change. In causal and longitudinal analyses, there is more room for acknowledging some degree of error based on the measurement strategy.

The use of multiple measures also helps reduce the chance of reification. This is because, since each measure has its own properties and sources of error, the precise findings of each one will vary (Thye, 2000). Confronting these differences will help researchers and readers remember that no single measure is perfect. Techniques like factor analysis or principal components analysis (PCA) also help prevent reification, since they output standardized factor scores that do not use the original scales of the indicator variables.

Finally, researchers might eschew quantification and instead develop suitable Likert scales. Likert scales do not have inherently meaningful scales, making reification less likely. The disadvantage is that this would mean adopting obviously subjective measures with no relationship to empirical data on inequality. This approach would thus be less suitable for studies about the sources or consequences of economic misperceptions, which have seen a recent boom (Hauser & Norton, 2017).

The measures we have reviewed contain important insights, despite the flaws our analyses identify. For instance, all measures converge on the conclusion that most people believe that inequality is subjectively high, and higher than they prefer. And quantitative measures may offer advantages that Likert scales cannot; for example, quantitative measures allow researchers to put differences between groups into context by expressing them in terms of ratios and multiples. They also allow researchers to more clearly identify relationships between perceptions of inequality/justice and actual economic conditions, such as the finding that people who live in more

income-homogenous neighborhoods tend to anchor their perceptions of incomes more generally on the income level of their neighborhood (Dawtry et al., 2015).

More broadly, we believe that researchers can develop more useful and reliable measures of perceived inequality and justice if they remember that subjective perceptions and objective reality are quite different areas of study, and different types of measures may be better at measuring either construct.

Conclusion

We compared numerous different ways of measuring perceived inequality and justice and found a range of methodological problems. These include participant comprehension, inconsistent responses, low correlations with criterion measures, and vulnerability to anchoring. These issues cause problems for interpreting existing work and for designing future studies. But we are optimistic that if researchers keep in mind these problems and the aforementioned prescriptions, research on perceived inequality and justice attitudes will be improved.

References

Arsenio, W. F. (2018). The wealth of nations: International judgments regarding actual and ideal resource distributions. *Current Directions in Psychological Science, 27*(5), 357–362

Ashok, V., Kuziemko, I., & Washington, E. (2015). *Support for redistribution in and age of rising inequality: New stylized facts and some tentative explanations*. National Bureau of Economic Research Working Paper No. 21529

Auspurg, K., Hinz, T., & Sauer, C. (2017). Why should women get less? Evidence on the gender pay gap from multifactorial survey experiments. *American Sociological Review, 82*(1), 179–210

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. *Political Analysis, 20*(12), 351–368

Cech, E. A. (2017). Rugged meritocratists: the role of overt bias and the meritocratic ideology in trump supporters' opposition to social justice efforts. *Socius, 3*, 1–20

Chambers, J. R., Swan, L. K., & Heesacker, M. (2014). Better off than we know: Distorted perceptions of incomes and income inequality in America. *Psychological Science, 25*(2), 613–618

Cheng, S., & Wen, F. (2019). Americans overestimate the intergenerational persistence in income ranks. *Proceedings of the National Academy of Sciences, 116*(28), 13909–13914

Clay-Warner, J., Robinson, D. T., Smith-Lovin, L., Rogers, K. B., & James, K. R. (2016). Justice standard determines emotional responses to over-reward. *Social Psychology Quarterly, 79*(1), 44–67

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from mechanical Turk valid for research on political ideology? *Research and Politics, 2*(4), 1–9

Davidai, S., & Gilovich, T. (2018). How should we think about Americans' beliefs about economic mobility? *Judgment and Decision Making, 13*, 297–308

Dawtry, R. J., Sutton, R. M., & Sibley, C. G. (2015). Why wealthier people think people are wealthier, and why it matters: From social sampling to attitudes to redistribution. *Psychological Science, 26*(9), 1389–1400

Eriksson, K., & Simpson, B. (2012). What do Americans know about inequality? It depends on how you ask them. *Judgment and Decision Making, 7*(6), 741–745

Eriksson, K., & Simpson, B. (2013). The available evidence suggests the percent measure should not be used to study inequality: Reply to Norton and Ariely. *Judgment and Decision Making, 8*(3), 395–396

Evans, M. D. R., Evans, J., & Kolosi, T. (1992). Images of class: Public perceptions in Hungary and Australia. *American Sociological Review, 57*(4), 461–482

Gimpelson, V., & Treisman, D. (2018). Misperceiving inequality. *Economics and Politics, 30*(1), 27–54

Hao, L., Naiman, D. Q. (2010). *Assessing inequality*. Sage Publications.

Hauser, O. P., & Norton, M. I. (2017). (Mis)Perceptions of inequality. *Current Opinion in Psychology, 18*, 21–25

Hegtvedt, K. A., & Isom, D. (2014). Inequality: A matter of justice?" In J. D. MacLeod, E. J. Lawler, & M. Schwalbe (Eds.), *Handbook of the social psychology of inequality* (pp. 65–94). Springer.

Heiserman, N., & Simpson, B. (2017). Higher inequality increases the gap in the perceived merit of the rich and poor. *Social Psychology Quarterly, 80*(3), 243–253

Heiserman, N., Simpson, B., & Willer, R. (2020). Judgments of economic fairness are based more on perceived economic mobility than perceived inequality. *Socius, 6*, 1–12

Hertel, F. R., & Schöneck, N. M. (2019). Conflict perceptions across 27 OECD countries: The roles of socioeconomic inequality and collective stratification beliefs. *Acta Sociologica* 1–19. doi: https://doi.org/10.1177/0001699319847515

Homans, G. C. (1974). *Social behavior: Its elementary forms*. Harcourt Brace.

Hunt, M. O., & Bullock, H. E. (2016). Ideologies and beliefs about poverty. In D. Brady, & L. M. Burton (Eds.), *The Oxford handbook of the social science of poverty* (pp. 93–116). New York: Oxford University Press.

Jasso, G., & Wegener, B. (1997). Methods for empirical justice analysis: Part 1. Framework, models, and quantities. *Social Justice Research, 10*(4), 393–430

Kelley, J., & Evans, M. D. R. (1993). The legitimation of inequality: Occupational earnings in nine nations. *American Journal of Sociology, 99*(1), 75–125

Kiatpongsan, S., & Norton, M. I. (2014). How much (more) should CEOs make? A universal desire for more equal pay. *Perspectives on Psychological Science, 9*(6), 587–593

Klein, E. (2013). This viral video is right: We need to worry about wealth inequality. *The Washington Post*. Retrieved from (https://www.washingtonpost.com/news/wonk/wp/2013/03/06/this-viral-video-is-right-we-need-to-worry-about-wealth-inequality/)

Kluegal, J. R., & Smith, E. R. (1986). *Beliefs about inequality: Americans' views of what is and what ought to be*. Transaction

Kuhn, A. (2019). The individual (mis-)perception of wage inequality: Measurement, correlates and implications. *Empirical Economics*. doi: https://doi.org/10.1007/s00181-019-01722-4

Litman, L., & Robinson, J. (2021). *Conducting online research on Amazon Mechanical Turk and beyond*. Sage Publications

Litman, L., Robinson, J., & Abbercock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433–442

Markovsky, B., & Eriksson, K. (2012). Comparing direct and indirect measures of just rewards. *Sociological Methods and Research, 41*(1), 199–216

McCall, L., Burk, D., Laperrière, M., & Richeson, J. A. (2017). Exposure to rising inequality shapes Americans' opportunity beliefs and policy support. *Proceedings of the National Academy of Science, 114*(36), 9593–9598

McCall, L. (2013). *The undeserving rich: American beliefs about inequality, opportunity, and redistribution*. Cambridge University Press.

Melamed, D., Liu, Y., Park, H., & Zhong, J. (2018). Referent networks predict just rewards. *Sociological Focus, 51*(4), 304–317

Meltzer, A. H., & Scott, R. F. (1981). A rational theory of government. *Journal of Political Economy, 89*(5), 914–927

Niehues, J. (2014). *Subjective perceptions of inequality and redistributive preferences: An international comparison*. IW-Trends Discussion Papers No. 2. 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., & Smeeding, T. (2006). "Fair" inequality? attitudes toward pay differentials: The United States in comparative perspective. *American Sociological Review, 71*(3), 450–473

Pedersen, R., & Mutz, D. C. (2018). Attitudes toward economic inequality: the illusory agreement. *Political Science Research and Methods, 12*, 1–17

Reynolds, J., & Xian, He. (2014). Perceptions of meritocracy in the land of opportunity. *Research in Stratification and Mobility, 36*, 121–137

Sauer, C., & Valet, P. (2013). Less is sometimes more: Consequences of overpayment on job satisfaction and absenteeism. *Social Justice Research, 26*(2), 132–150

Schneider, S. M., & Castillo, J. C. (2015). Poverty attributions and the perceived justice of income inequality: A comparison of East and West Germany. *Social Psychology Quarterly, 78*(3), 263–282

Schneider, S. M., & Valet, P. (2017). Relative standards and distributive justice: How social comparison orientations moderate the link between relative earnings and justice perceptions. *Social Psychology Quarterly, 80*(3), 276–287

Schröder, M. (2017). Is income inequality related to tolerance for inequality? *Social Justice Research, 30*(1), 23–47

Small, M. L. (2011). How to conduct a mixed methods study: Recent trends in a rapidly growing literature. *Annual Review of Sociology, 37*, 57–86

Thye, S. R. (2000). Reliability in experimental sociology. *Social Forces, 78*(4), 1277–1309

Trump, K.-S. (2017). Income inequality influences perceptions of legitimate income differences. *British Journal of Political Science, 48*(4), 929–952