Oxford Economic Papers, 2025, 1–20
https://doi.org/10.1093/oep/gpaf014
Article

OXFORD

The meritocratic illusion: inequality and the cognitive basis of redistribution

Arthur Blouin [1,*], Anandi Mani [2], Sharun W. Mukand [3], Daniel Sgroi [3]

[1]Department of Economics, University of Toronto, 150 St. George Ave, Toronto, ON, M5S 3G7, Canada
[2]Blavatnik School of Government, University of Oxford, 120 Walton St, Oxford, OX2 6GG, United Kingdom
[3]Department of Economics, The Social Sciences Building—The University of Warwick, University of Warwick, Coventry, CV4 7AL, United Kingdom

*Corresponding author. Department of Economics, University of Toronto, 150 St. George Ave, Toronto, ON, M5S 3G7, Canada. E-mail: a.blouin@utoronto.ca

Abstract

Can an inequality in rewards result in an erosion of broad-based support for meritocratic norms? We examine whether unequal rewards can affect social preferences for redistribution by driving a cognitive gap in the meritocratic beliefs of those who are successful and those who are not. Two separate experiments (conducted in the USA and the UK) show that the elite develop and maintain 'meritocratic bias' in the redistributive taxes they propose. This bias results in lower taxes on the rich and fewer transfers to the poor, including those who failed despite high effort. These social preferences at least partially reflect a self- serving meritocratic illusion that their own high income was deserved and reflected their ability. An incentivized Wason Card task confirms that individuals prefer to maintain their illusion of being meritocratic, by not expending cognitive effort to process any information that may undermine their self-image of being deserving.

Keywords: inequality; meritocracy; redistribution; populism; motivated reasoning; social preferences
JEL classifications: D63, D91, H23

1. Introduction

Democratic societies in the twenty-first century rest on the presumption of meritocratic norms and a pro-meritocracy worldview. Meritocracy is not only a cornerstone of the American Dream, but is a widely held belief that spans the income distribution (Cappelen et al. 2007, 2013, 2017) and is often championed by politicians of all ideological stripes (Sandel 2020). However, recently rising inequality has led to renewed criticism of meritocratic governance. Those opposed to a pro-meritocracy worldview have argued, for instance that 'the discourse of meritocracy … seems to serve primarily as a way for the winners in today's economy to justify any level of inequality whatsoever

while peremptorily blaming the losers for lacking talent, virtue, and diligence' Piketty (2020: 2).[1]

In this article, we seek to assess how success contributes to the emergence of pro-meritocratic beliefs. In doing so, we analyse not only how these beliefs develop, but also how such beliefs persist, despite evidence to the contrary. We do this through a series of experimental tests. First, we construct an environment in which pro-meritocratic beliefs can develop on potentially selfish grounds. We grant individuals the opportunity to succeed through their own efforts but also allow for the possibility that luck may be the driving force behind success. This enables us to witness the development of pro-meritocratic beliefs that may be grounded in effort or ability, but may also be based on nothing more than good luck. The latter may arise, for instance if individuals, for self-serving reasons, convince themselves that their success was merited, even if it was not. Second, we vary the information that individuals receive about whether their success was merited, or based on luck. This enables us to assess whether this information influences the pro-meritocratic beliefs of those who succeeded due to luck. Finally, we implement a task that allows us to observe whether people actively avoid information in order to maintain meritocratic beliefs for self-serving reasons.

To measure the extent to which potentially selfish pro-meritocratic beliefs can extend into the realm of social preferences, we examine decisions about redistribution. Importantly, these decisions do not affect the decision-maker—our tax setters do not face the taxes that they set. Furthermore, we control for uncertainty by requiring separate tax-rate decisions for effort and luck-based earnings.[2] Our aim is to go beyond selfish rationality and to measure something closer to an ideological worldview: something that applies beyond narrow payoff-relevancy for the individual.

We then examine the question of what individuals will choose to do, when faced with a decision that requires them to either process information inconsistent with their beliefs, or information that results in their earning more money. Accordingly, we introduce into economics a method that is standard within Psychology: a Wason task (Wason 1966). A Wason task is a pure test of conditional reasoning and throws light on whether an individual's process of learning is akin to hypothesis testing. In particular, it examines whether an individual is willing to search for evidence that could falsify a hypothesis. We adapt the Wason selection task to our context by asking subjects to indicate what information is needed to disprove a pro-meritocratic statement. Notice that this may mean that subjects who formed pro-meritocratic beliefs on poor logical grounds (if they succeeded through luck) will be faced with a payoff-relevant tension. This is because the lucky may wish to preserve beliefs that favours their own self-image that their success is deserved and reflected in their chosen meritocratic social preferences. However, doing so may require them to fail the Wason task which in turn damages their direct payoff.

We have three main findings. First, while there exist some papers linking meritocratic beliefs to success, and other papers linking success and social preferences (discussed below), ours is the first to bring these three elements together into a single experiment. We show that success—regardless of its source—can increase individuals' proclivity to adopt meritocratic social preferences. Our main result is that not only do social preferences for redistribution shift as a result of success, but they do so in a meritocratic way even if the success was not won through effort and ability. This main result contrasts with the existing literature, which has shown that lottery-task winners redistribute less of luck-based earnings

[1] Relatedly, Supplementary Appendix Figs. A1a and A1b from our survey of Americans show that both meritocratic beliefs and lower preference for redistribution are correlated with income, though do not provide a causal link.

[2] We follow the standard definition of social preferences as 'the human tendency to not only care about one's own material payoff, but also the reference group's payoff or/and the intention that leads to the payoff' (Carpenter 2010). Our reference group here is the group of subjects in each experimental session who are affected by the chosen tax rate.

(but not less effort-based earnings), while effort-task winners redistribute less of effort-based income (but not less luck-based earnings).

The reason for this difference has to do with imperfect information that participants have about how their success was achieved. In our experimental design, individuals do not always know how much of their success was driven by their effort and ability, and how much was luck. This doubt allows scope for self-delusion. When we introduce full information about the how success was achieved, the meritocratic illusion that lucky income was deserved disappears, leaving us with estimates that are much more consistent with past work. In particular, when we reveal to people in the luck task that they earned their money through the lottery, they no longer redistribute more effort-based income.

While this result identifies the important role of information in the preservation of meritocratic beliefs, we go further to document that (1) the mere possibility that some information may threaten their meritocratic beliefs is enough to make successful participants avoid it (which is possible because of the scope for self-delusion within our experiment) and, moreover (2) they avoid this information despite it being personally costly to do so. Indeed, our introduction and novel application of a logical reasoning (Wason) task allows us to provide rigorous evidence for the distinct mechanism driving information avoidance: the threat of cognitive dissonance (between meritocratic beliefs and alternative explanations) rather than selfish/rational concerns about one's own self-image or tax burden.[3] The persistence of meritocratic beliefs amongst the successful that we document sheds light on how rising inequality could drive a widening wedge in the beliefs of rich versus poor, resulting in increasing political polarization across the globe today.

We can highlight how this works at an individual level by following the thought process of a hypothetical subject who holds a pro-meritocratic viewpoint. Success in the initial experimental tasks generates a stronger belief that the world is meritocratic whether this success comes through luck or effort, for instance: 'I succeeded and in this meritocratic world that means I am deserving.' This leads to meritocratic social preferences, for instance: 'Those who succeed like me are also deserving and so should face lower taxes.' Moreover, this may even allow them to disregard logic even when they are incentivized to do the opposite: 'I am willing to lower my payoff to avoid considering a statement that highlights the fact that success does not imply deservingness.'

In fact, we do find that those who deluded themselves into believing that their success was merited (or conversely that their failure was due to luck) are less likely to solve a Wason task aimed at disproving the relevant self-serving belief. Again, this is true despite the fact that disproving (or not disproving) these statements has no bearing on whether the participant's success was or was not merited, and despite the fact that they earned more by completing the tasks successfully. Furthermore, this variation in solving Wason tasks is not due to unobserved ability in solving these tasks, since we control for participant's overall solving ability by asking them to also solve a task that aims to disprove a statement that disagrees (rather than agrees) with the self-serving belief that may have been generated by their experience in the experiment.

There is a large literature on meritocracy across the social sciences—including in political philosophy (Appiah 2018; Sen 2018; Markovits 2019), psychology (McCoy and Major 2007), economics (Arrow, Bowles, and Durlauf 2018), and of course, sociology.[4] Indeed, the term 'meritocracy' was invented by the British sociologist (and activist-politician) Young (1958) whose work included a discussion of negative aspects of meritocracy. Our article contributes to this extensive literature and shows the benefits of an interdisciplinary

[3] Following Festinger (1959) and Akerlof and Dickens (1982), we define Cognitive Dissonance as follows: '[a]t the most abstract level this means that persons are uncomfortable in maintaining two seemingly contradictory ideas'.

[4] Also related is Lerner's (1980) system-justifying *belief in a just world* that 'people generally get what they deserve.' In contrast, our results suggest a more complex picture, since a cognitive bias may undermine the system by driving polarization. Also see (Bénabou and Tirole 2006).

approach—the application to taxation and redistribution is not an area that psychologists and sociologists focus on, just as social preferences and cognition are not areas that are a central preoccupation of economists.

This article is also related to work on attitudes towards redistribution. Romer (1975) and Meltzer and Richard (1981) build on Arrow (1950) to argue that since individuals were self-interested, higher income will result in them preferring lower redistribution. Karadja, Mollerstrom, and Seim (2017) provide evidence consistent with this and show that individuals who discover they are richer than they thought, demand less distribution, and Cruces, Perez-Truglia, and Tetaz (2013) also discuss the effect of relative rank information on luck/effort beliefs. Meanwhile both Deffains, Espinosa, and Thöni (2016) and Brown-Iannuzzi *et al.* (2021) find that a similar effect can be generated in the lab.[5]

Other studies that analyse the impact of the socio-economic and political environment in shaping an individual's redistributive preferences, such as social identity (Klor and Shayo 2010); political institutions (Alesina and Fuchs-Schündeln 2007); genetics (Vu 2023); and macroeconomic shocks (Giuliano and Spilimbergo 2014). Doherty, Gerber, and Green (2006); Peterson (2016), and Hvidberg, Kreiner, and Stantcheva (2020) all show the impact of lucky income shocks on distributional attitudes. Cassar and Klein (2019) show that income generated through both a lottery and a tournament influence distributional attitudes and together with Deffains, Espinosa, and Thöni (2016) show that self-serving biases can impact redistribution decisions. However, neither paper experimentally varies information, and so neither shares our focus on how beliefs are maintained in the face of differences in information about the state of the world. Our design allows us to test whether cognitive dissonance is a mechanism underlying the redistributive preferences declared by participants, because of the combination of our information treatments and the Wason Task.[6]

Of course, disentangling broad-based social preferences for redistribution from selfish income related reasons is not easy and often has to rely on sophisticated surveys as well as experiments rather than simple observational data. This includes work by Fisman *et al.* (2020), Alesina, Stantcheva, and Teso (2018) (see Stantcheva (2021) for a recent survey). This work is also related to work that examines whether the source of income inequality matters for its tolerance (Konow 2000; Cappelen et al. 2007; Karadja, Mollerstrom, and Seim 2017). Moreover, Cappelen *et al.* (2013) find evidence that subjects do distinguish between the source of earnings and are willing to redistribute to compensate for earnings that seem undeserved. Our results suggest that attitudes towards meritocracy, especially by the higher earners in our experiment, seem to act as a limit to these fairness concerns.[7] More recently, Fehr and Vollmann (2020) also suggest that individual tax choices are affected since subjects misperceive success. In contrast, our design focuses on how motivated reasoning may affect *social* preferences for redistribution *and* associated meritocratic beliefs. There is also a related literature in development economics (and elsewhere) that examines the role of fairness when assessing earned vs unearned income in a variety of countries (Jakiela 2011; Rey-Biel, Sheremeta, and Uler 2011; Barr and Miller 2020). Almås, Cappelen, and Tungodden (2020) in a very large-scale study of subjects in the USA and Norway find evidence of strong cross-national variations in attitudes towards fairness in social preferences.

[5] Relatedly, Alesina and La Ferrara (2005) find that redistributive preferences depend on expected future income.

[6] We present a more detailed examination of the extent to which our findings differ from Cassar and Klein (2019) and Deffains et al. (2016) in the concluding comments. We argue that while our results are very different, this is largely because of a difference in focus and design, and so there is no inconsistency between our findings and those in the literature.

[7] In particular, Cappelen et al. (2017, 2021) show that both linking rewards to actions, as well as certainty about the role of luck—both make individuals more meritocratic. See also Cappelen and Tungodden (2019) for a good overview.

2. Research design

Any empirical investigation of whether an increase in income may cause a self-serving meritocratic shift in attitudes towards redistribution has to overcome several challenges and possible confounds. We discuss some of these in Section 4.3 (for a more comprehensive list see Supplementary Appendix Table A1) including the possibility of selection, an in-group bias amongst the rich, a confirmation bias as well as unobserved socio-cultural and other characteristics that can influence attitudes towards meritocracy and redistribution.

Below we sequentially describe the main experimental interventions that we use to overcome the above challenges to examine whether unequal rewards result in a self-serving meritocratic bias in distributional social preferences.

2.1 Experimental procedure

We ran forty-four experimental sessions with over 400 participants at the Economics Laboratory at the University of Warwick in the UK. We supplemented this with an online experiment/survey using an American general population sample that is quota representative. It is stratified through the platform Prolific on age, ethnicity and sex to be representative, and on income to get power in the tails of the distribution. Protocol and survey details for both experiments are outlined in Supplementary Appendix B1. In every session, after an initial questionnaire designed to supply us with controls (see Supplementary Appendix Table A2), all subjects participated in two tasks, either of which could earn them money: an *effort task* and a *lottery task*. All subjects knew that after completing both these tasks, a computer would randomly choose (with equal probability) either the effort or the lottery task to determine their earnings.

2.1.1 Effort and lottery tasks

The first potential income generating task was an *effort task* that required each subject to add as many five 2-digit numbers as possible in 5 minutes (e.g. $14+62+73+39+92=?$). If they correctly solved greater/lower than a threshold number of addition problems (where the unknown threshold was variable and stochastic) they were rewarded with a high/low wage (for simplicity we label them 'rich'/'poor'). The second task that all subjects faced was a random lottery—the *lottery task*. A subject could either become lucky and win the lottery and be paid a high reward (with probability 50 per cent) or turn out to be unlucky and lose the lottery and earn a low wage. While all participants saw their income increase, the high reward in the chosen task (lottery or addition) was £15, the low reward was £5. This lab-induced economic inequality between those who earn a high reward versus those who earned a low reward was designed to be large in relative terms, to capture the effects of income inequality.

2.1.2 Full and partial information

All subjects were also cross-randomized and assigned to either the full or the partial information treatment. In the *partial information* treatment, subjects were only informed about whether they had earned a high reward (i.e. become 'rich') or low reward. In the *full information treatment*, they were also informed about the source of their reward (lottery or effort task), their performance in the effort task (i.e. the number of correct additions), and the threshold that was used to determine whether they were paid a high or low reward.

A balance exercise showing the distribution of *P*-values for each control/treatment pair can be seen in Supplementary Appendix Fig. A2. Additionally, we checked the balance for the information treatment. Of the 120 observable characteristics we looked at, only three were significant at the 10 per cent level or below. These were third-year undergrads were less likely to get full information, computer science students more likely, those from south Asia more likely. For above/below the threshold, we also see three significant variables out of 120. They are for student loans (negative correlation), second-year undergrads (negative

correlation), and those from south-east Asia (positive correlation). Overall, we are confident that each of our main treatments is balanced, since in each case the number of statistically significant correlations with treatment status is consistent with random assignment.

2.1.3 Redistributive social preferences

After completing these tasks under the above treatments, all subjects were asked to sequentially choose two tax rates. The first *lottery tax* was chosen for those who earned a high reward due to the lottery. They selected a value between 0 and 100 to indicate how much they believed others who had success during the effort task should transfer to participants who did not. Similarly, an *effort tax* was chosen. There was *no* uncertainty as to the source of income being taxed—that is, whether it was the lottery or the effort task. This absence of uncertainty in taxation minimizes the prospect of confirmation or availability bias in driving any meritocratic shift in social preferences.

Note also that all subjects were informed that the tax rates implemented in the session would be randomly chosen from the set of tax rates proposed by subjects in the session. They were also informed that subjects would not face the tax rate they themselves proposed. In the event that the rate they proposed was chosen for all other subjects, the tax rate proposed by another subject would be applied to them. Thus, it was participants' *social* preferences over redistribution that were being elicited. Finally, as an attention check, we also asked subjects to propose the tax rate under the condition that they *would* face this rate, if it were chosen for all other participants (see Supplementary Appendix B). Supplementary Appendix Table A3 shows descriptive statistics of the proposed tax rates for each type of income.

2.1.4 The Wason Card selection task

After subjects had proposed their tax rates, they were asked to complete two Wason Card tasks. This was incentivized with an additional £2 reward for a correct response on one of the two tasks, randomly chosen (complete details are in Appendix B5).[8]

The Wason Card task is a test of logical (conditional) reasoning, widely used in cognitive psychology, but not so far, in economics. It requires subjects to evaluate a conditional statement of the form *'if P then Q,'* using four cards displayed on the screen (Wason 1966). These cards display the *P*, not-*P*, *Q*, not-*Q* segments of the statement individually. For each card that has the *P* portion of the statement on the front, the reverse has the *Q* portion of the statement, and vice versa. Each subject is given the task of picking and turning over the minimal set of cards that help demonstrate whether the original conditional statement is true or false.

From a purely logical perspective, only the combination of the two cards *P* and not-*Q* can violate this rule. Accordingly, the correct answer is to check the *P* card (to see if it has a not-*Q* on the other side) and the not-*Q* card (and check whether *P* is on the reverse side). Therefore, the correct answer for all Wason statements, is to flip the two cards *P* and *not-Q*.

Under full rationality, correctly solving the Wason logical reasoning task should be completely independent of any content or context. However, our goal here is to examine whether subjects exhibit motivated reasoning given the context of their (random/quasi-random) experimental outcome. To do so, we asked them to evaluate two Wason statements with differing content, as described below. The first was,

> *If a participant in the experiment becomes rich, (s)he must have been lucky and obtained a high lottery payout.*

[8] The experiment then ended with an opportunity to complete an anonymous on-screen comment box prior to payment.

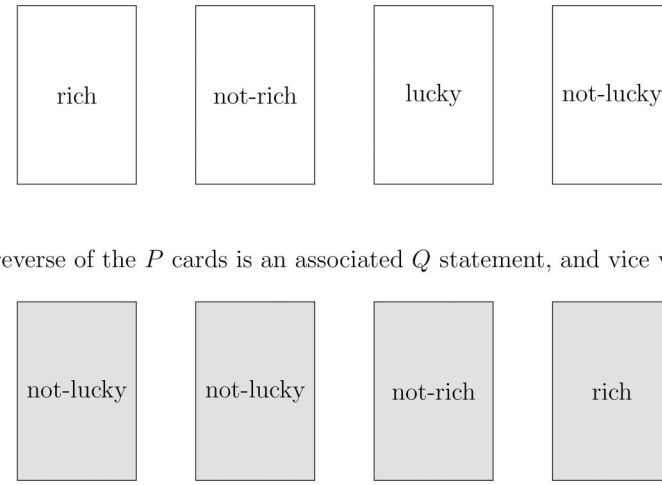

Figure 1. Illustration of Wason Cards.

The second, contrasting, statement that each subject saw was:

> *If a participant in the experiment becomes rich, then (s)he must have been hard working and hence scored above the threshold.*

The two statements express contrasting worldviews about the role of effort versus luck in success. As an example, consider statement 1, where P and Q take the forms of 'rich' and 'lucky', respectively, to match our meritocracy setting. The cards associated with statement 1 are displayed in Fig. 1. In order to evaluate, for instance, whether the first statement is true (if rich, then lucky) the correct answer would be to flip the first and fourth cards below. The second statement was associated with a different set of cards (since the 'if P then Q now takes the form 'if rich, then hard-working). We show these cards and describe the correct answer in Supplementary Appendix B5.

We hypothesize that subjects would be more motivated to scrutinize the statement that challenges their preferred beliefs more thoroughly. This is because this statement creates a cognitive dissonance between their preferred belief and performing the Wason task correctly, despite the financial incentive. For instance, a high-reward 'rich' subject would be more motivated to reason correctly in the second Wason Card task, while a low-reward 'poor' person would want to do so in the first task. As a result, motivated reasoning would result in a systematic gap in performance across the two Wason tasks, as a function of subjects' experimental outcome.

There is also an important role here that we predict for the information treatments. Participants who had full information about the source of their income—for instance, the lottery task—may find it harder to attribute success (theirs or others) to hard work. In contrast, those not informed about the source of their high reward might find it easier to persuade themselves that it was their effort that paid off.

Three additional points are noteworthy. First, the payoff in the Wason task is for the individual alone. Unlike with the tax rates s/he proposed, no one else benefits from how logically they reason (or not). This rules out *group identity* as the explanation for any gap in performance across the two tasks. Second, the within-person design allows us use the gap in performance across the two logical reasoning tasks as the outcome of interest, hence ruling out individual differences in *cognitive ability* as an explanation for observed outcomes. Finally, the statements being evaluated do not pertain to any individual subject in

particular, and in this sense they are purely hypothetical. To summarize, the combination of the Wason task and the partial information treatment allows us to test for cognitive dissonance with an individual's beliefs shaped by personal experience, as the mechanism for (financially) costly information avoidance.

2.1.5 The online experiment and the observer treatment

We also supplemented our laboratory experiment with an online experiment (protocol details are in Supplementary Appendix B). Here, a sample of American subjects could earn a $2 bonus either through winning a lottery or through successfully completing an effort-ability task (Tower of Hanoi puzzle).[9] As a function of the outcomes of these tasks, some subjects won the $2 bonus. Participants in the online experiment were *not* informed whether their payoff was due to winning the lottery, or their performance on the puzzle task.

The online experiment included a treatment arm of *observers* who did not participate in the tasks or earn any bonus. However, like the active participants, they saw all the rules and instructions of the experiment. All subjects (i.e. participants and observers) were then asked how they would redistribute the bonus income between the winners and the rest. Any subject who either exited partway through the experiment or timed out was excluded from our sample.

3. Empirical approach

We examine whether subjects have meritocratic social preferences and whether their attitudes towards redistribution vary across (1) differing sources of income—be it effort or pure luck and (2) differing information about the cause of their success. In effect, we examine whether there is a gap in distributional social preferences between those who succeed in our experimental tasks and those who fail, and the extent to which this gap is affected by information.

3.1 Measuring meritocratic social preferences

Our primary measure captures the notion that a 'meritocracy refers to the idea that whatever your social position at birth, society ought to offer enough opportunity and mobility for 'talent' to combine with "effort" in order to rise to the top.' (Littler 2013: 1). Accordingly, this measure emphasizes that individuals deserve and have a moral right to retain income generated through 'merit' - these rewards are 'just deserts' (Sandel 2020: 34). Our experiment measures meritocratic *social* preferences as the tax rate that an individual imposes on high earnings conditional on high effort, earned by others' (i.e. excluding themselves).

We also use a broader notion of meritocracy that encompasses not just attitudes towards income earned through an individual's effort, but also income earned through factors outside their control—that is, luck. This measure of meritocratic social preferences would be captured by the ratio of the tax on effort over the tax on income earned from the lottery, that is, $\frac{EffortTax}{LuckTax}$. While intuitive, this measure should be treated with some caution since Rawls (1999, sec. 17) suggests that, arguably, there is an element of luck in how distasteful an individual finds it to put in effort. If so, there is a morally arbitrary element in making a distinction between earnings from a lottery instead of effort.[10]

[9] The recruitment for the experiment was completed using the Prolific app, and we requested a nationally representative sample, according to their definition. This includes a stratification on age gender and ethnicity, in order to match US census information.

[10] There are some cases where individuals set both the effort and luck taxes to zero. We impute a value of zero in this case, since this reflects an attitude that there should never be redistribution, which seems to be a meritocratic perspective.

3.2 Lottery experiment

In this experiment, each subject had an equal chance to win a high reward through taking part in a random lottery. The specification is

$$Outcome_{is} = \beta_0 + \rho Lottery_i + \Gamma X_i + \varepsilon_{is} \tag{1}$$

$Lottery_i$ is a dichotomous variable indicating whether subject i earned a high income through winning the lottery. This means that ρ represents the causal effect of (lucky) high pay on taxation decisions.[11] X_i is a vector of controls that includes the gender, age, academic department, and political orientation of each respondent. ε_{is} is the error term for individual i in lab-session s, and standard errors are two-way clustered at the state and academic department level throughout. We consider as outcomes both measures of meritocratic social preferences for redistribution, and outcomes of the Wason task.[12]

3.3 Effort experiment

In the effort experiment, the subjects completed a number of addition tasks, and they were paid more if they completed more than a randomly determined threshold of correct answers. Accordingly, we can implement a regression discontinuity design (RDD) that allows us to compare the social preferences of subjects who *just* missed the threshold and earned a low reward with those (very similar) subjects who just made the threshold and earned a high reward.

This comparison allows us to estimate the causal effect of the high income reward on distributional social preferences, controlling for the effort level of the participant. We allow for a continuous relationship between effort and redistributive preferences, and test for a discontinuous jump at precisely the point where there was a discontinuity in pay. If such a discontinuity in preferences exists right at the high-pay threshold in effort, we interpret that as evidence that the difference in pay *caused* the difference in preferences. We implement the RDD using each of the local-linear, polynomial and kernel methods. We employ the local-linear specification as follows:

$$\begin{aligned} Outcome_{is} = \beta_0 + \rho \mathbf{1}[Effort_i > T_s] + \alpha_0(Effort_i - T_s) + \\ \alpha_1(Effort_i - T_s) \cdot \mathbf{1}[Effort_i > T_s] + \Gamma X_i + \varepsilon_{is} \\ | -h < (Effort_i - T_s) < h. \end{aligned} \tag{2}$$

Here ρ is the estimated size of the discontinuity at the effort threshold in the relationship between effort and taxation. We interpret this as the causal effect of higher earnings on our main outcomes (tax rates and Wason tasks), for an individual with effort at the threshold. We denote $Effort_i$ as the number of correct tasks completed by respondent i, while T_s is the threshold for high wage in session s. We allow the slope of the relationship between effort and tax choices to differ on either side of the threshold T_s. X_i is a vector of controls that is identical to those in the lottery experiment. h is the bandwidth that we use, and is estimated according to the MSE minimization routine described in Calonico, Cattaneo, and Titiunik (2014).

3.4 Full versus partial information treatments

One of the core contributions of our article is to investigate the role of information in the generation of self-serving beliefs.[13] Under full information, past work has identified that

[11] Whenever we discuss parameter estimates throughout the article and their precision, all references are to two-sided tests.

[12] Throughout the article, we two-way cluster standard errors on session and academic department. There are twenty-seven unique sessions in the data, and twenty-three unique academic departments.

[13] The tests described above are all using samples pooled across the information treatments described in this section.

lucky winners redistribute less from other lucky winners, but not effort-based ones. Likewise effort-based winners redistribute less from other effort-based ones, but not lucky ones (Cassar and Klein 2019). However, outside of the lab people often do not know the extent to which their success is due to luck and effort, and these results might be much different when people are able to delude themselves that they deserved their earnings when in reality they did not.

For instance, if the self-serving (i.e. motivated cognition) nature of meritocratic beliefs was a driver, then we should expect a difference between the partial and full information treatments. In this case, the rich who do not know for sure that they won the lottery (i.e. in partial information treatment), can find it much easier to delude themselves that they got rich due to their 'merit'. In contrast, the rich who are informed that their earnings are due to winning the lottery may find it harder to continue to delude themselves that their earnings are due to their effort/ability.

Our main interest in the information treatment is to better understand the role of information in sustaining meritocratic beliefs. However, one additional benefit of the information treatment is that it allows us to rule out some alternative mechanisms. Observe that across both treatments, all individuals know whether they earned a high or low reward. Therefore, it is plausible that this in-group affiliation amongst the relatively 'rich' and 'poor' may shape their attitudes towards the taxation of other rich. A low tax imposed on other rich individuals may be because of an in-group bias for other rich.

However, the informational treatments can also be used to examine whether our results are being driven by such a group-identity. In particular, if a group-identity driven in-group bias were driving social preferences, then we should expect the high reward 'rich' to behave in a similar way—whether or not they know the source of their high earnings. A summary of alternative mechanisms and how they can be ruled out appears in Supplementary Appendix Table A1.

4. Results: Does income cause meritocratic social preferences?

4.1 Results from the lottery experiment

We start with an analysis of the lottery experiment.[14] The results show that winning the lottery changed subjects' perceptions of what is a fair distribution—their social preferences became more meritocratic.

Figure 2a demonstrates that lottery winners taxed effort at a lower rate than non-lottery winners. In Fig. 2a, the gap is over 20 percentage points, while in 2 b it is about 10 points. This is consistent with Supplementary Appendix Table A4, which suggests that lottery winners taxed effort 21percentage points lower than lottery non-winners (column 1; $P < .01$) or about two-thirds of a standard deviation (column 3; $P < .01$).[15] Our alternative meritocratic measure suggests a similar shift. For instance, Supplementary Appendix Table A4, column 2 presents a 10 per cent decline in effort tax relative to taxation on lucky earnings ($P = .078$), which after normalization (column 4) translates into a difference of more than a fifth of a standard deviation ($P = .078$). This gives us a consistent message: earning more money (even in a lottery) makes a person more generous towards high-earning subjects by redistributing less of their income to the rest.

While more income may affect subjects' views of what is fair, we also examine other possibilities. For instance, is it possible that money is not the cause of the effect, and that even neutral observers would share a similar meritocratic bias? To assess this possibility, we

[14] The results presented in this section pool the partial and full information treatments, and later we explore heterogeneity by information.

[15] Supplementary Appendix Table A5 shows robustness to controlling for effort in the effort experiment. This is to mitigate concerns that that the proposed tax rates are influenced by the exerted effort irrespective of the outcome of the lottery.

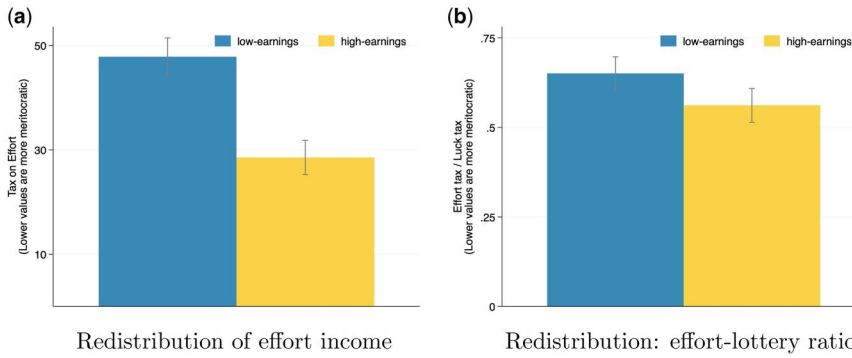

Figure 2. Lottery earnings makes social preferences more meritocratic.

Note: The figure graphs the impact of winning a lottery on our two main measures of preferences for redistribution. In panel (a), these preferences are measured as the tax rate that an individual imposes on high effort task income, earned by others (i.e. excluding themselves). In panel (b), these preferences are measured as the tax rate that an individual imposes on high effort task income *relative to lottery income*, earned by others (i.e. excluding themselves). Mean choices of individuals randomly assigned to high-earnings is in pink, and mean choices of individuals randomly assigned to low earnings is in blue. 90% confidence intervals are plotted for each. Each graph presents means plus residuals after conditioning out the standard set of controls, described in Sections 3.2 and 3.3.

also report our results from our online (US-based) experiment which included an observer-treatment. Our specification is identical to earlier, except that given the regional heterogeneity of our sample, we control for state fixed effects and cluster standard errors by state. While we were able to replicate our core luck (lottery winners vs lottery losers) result amongst a non-student population of Americans (Table 1, column 1, estimate = −14.44; $P = .014$), our focus is on the treatment arm of observers.

Column 2 examines a comparison between the winners and the observers, while in column 3 we compare those that participated in, but lost the lottery, to the observers. The result is nearly identical using the observers as the control group (column 2, estimate = −15.85; $P = .007$), since observers behave very similarly to lottery losers (column 3, estimate = 4.02; $P = .634$). Column 4 presents a joint-sample specification, and the results are again nearly identical (estimate=−12.12; $P = .040$). Taken together, these results suggest that winners of a random lottery develop social redistributive preferences that are quite different from both lottery losers and impartial observers.

These results are quite surprising, since they differ considerably from the literature. Past work has demonstrated that lottery winners redistribute less luck-based earnings, but not less effort-based earnings (Cassar and Klein 2019), while in our context even the ratio of effort-to-luck-based redistribution declines, a marked deviation from past work. We argue that this is because when there is scope for people to delude themselves that their income was earned, they do so. If this is true, then we should expect the results to be driven by winners in the luck experiment who do not know that they were assigned to the lottery. We could compare these winners to winners from the effort-task as long as they too develop more meritocratic preferences.

4.2 Results from the effort experiment

Before turning to the role of information, we consider the effort experiment for an additional test of whether income shocks can shift a subject's distributional preferences.[16]

We do this by implementing an RDD, where we exploit the discontinuity in rewards earned, around small differences in the number of questions solved. The results are

[16] As in the previous subsection, the results presented in this section pool the partial and full information treatments.

Table 1. Online experiment with observer treatment.

Dependent variable	EffortTax			
Sample	Lottery winners and losers (1)	Lottery winners and observers (2)	Lottery losers and observers (3)	All (4)
Lottery Winners	−14.44**	−15.85***		−12.12**
	(5.603)	(5.444)		(5.695)
Lottery Losers			4.016	4.132
			(8.366)	(7.489)
Age	√	√	√	√
Gender	√	√	√	√
Political Attitudes	√	√	√	√
State FE	√	√	√	√
Observations	219	177	152	274
R^2	0.264	0.295	0.298	0.234
Dependent Variable Mean	40.51	37.20	44.14	40.45

Standard errors are clustered at the state level.
***, **, and * represent statistical significance at the 1 per cent, 5 per cent, and 10 per cent levels, respectively. The table presents results from the online experiment. The table suggests that lottery winnings make people more meritocratic. These preferences are measured as the tax rate that an individual imposes on high effort task income, earned by others (i.e. excluding themselves). Lower values mean people are more meritocratic. Individuals were assigned to either be lottery winners, lottery losers, or observers. Column 1 compares high to low-earners, column 2 compares high-earners to observers, column 3 compares low-earners to observers, and column 4 compares high and low-wage to observers.

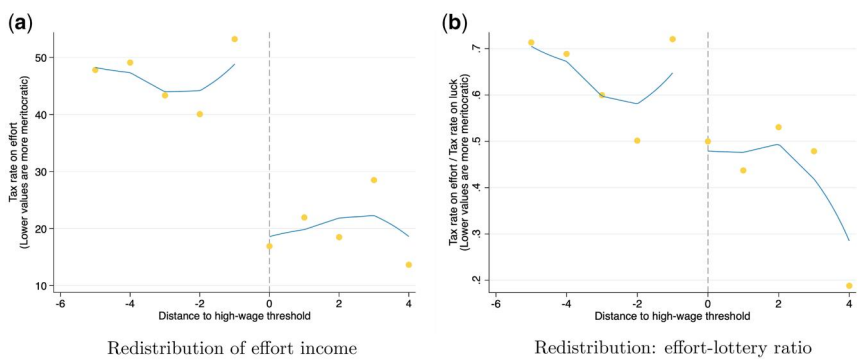

Figure 3. Merited earnings makes social preferences more meritocratic.

Note: The figure plots mean tax rates chosen in the effort experiment for our two main measures of preferences for meritocratic redistribution. In panel (a), these preferences are measured as the tax rate that an individual imposes on high effort task income, earned by others (i.e. excluding themselves). In panel (b), these preferences are measured as the tax rate that an individual imposes on high effort task income *relative to lottery income*, earned by others (i.e. excluding themselves). Each subfigure plots mean tax choices against the number of correct addition problems solved relative to the threshold determining high-earnings. The dashed line at 0 represents the threshold, so the right of the dashed line represents high-earnings individuals, and to the left is low-earnings individuals. Each graph presents means plus residuals after conditioning out the standard set of controls, described in Sections 3.2 and 3.3.

displayed graphically in Fig. 3a and 3b, respectively. In both cases, there is a large discontinuous shift in social preferences towards meritocracy at precisely the point along the effort distribution that people earn more money.

The effect on our primary measure of meritocratic social preferences (i.e. the effort tax) is particularly stark as we observe individuals reduce taxation on earned income by over 30

percentage points in our experiment (Supplementary Appendix Table A6, column 1, estimate = −30.18; $P < .01$). These estimates are robust to different bandwidth choices (columns 4–10 of table A6), and extend to more or less parameterization (columns 2–4 of Supplementary Appendix Table A6). The effect on the ratio of effort tax to luck tax is very similar (Supplementary Appendix Table A7).

4.3 Why do the rich become meritocratic?

There are several possible reasons why an increase in income would make an individual's social preferences more meritocratic. We discuss some of these mechanisms in light of our experiment.

First, is the possibility that talented and hardworking individuals are over-represented amongst the successful and rich. These people may be hardworking because they already held a meritocratic worldview.[17] While possible, our research design explicitly excludes this possibility since we examine the impact of *random* income shocks on an individual's distributional social preferences. Second, narrow self-interest alone could ensure that those with higher income favour less redistribution. However, our research design precludes this possibility as well, since a subject's choice of tax cannot affect their own income. Third, arguably higher income could alter an individual's group identity or loyalty to other successful and rich individuals. In other words, a group/tribal affinity towards the rich may *indirectly* affect a subject's distributional social preferences, such that they tax fellow members of their group at a lower rate. We further investigate this possibility in the Section 4.3.3.

Finally, an individual may be motivated to choose social preferences in a way that is consistent with their desire to maintain their self-image (Rabin 1994; Bénabou and Tirole 2016). Given the pervasive nature of meritocratic norms in contemporary society, successful individuals may persuade themselves that they deserved their rewards, that effort pays-off, and so they adopt meritocratic beliefs. Furthermore, we observe that subjects may find it easier to maintain this belief, and the associated meritocratic social preferences may be easier to maintain in the partial information treatment. In the next section, we further investigation this channel and describe our results.

4.3.1 Information provision and the meritocratic shift

As we discussed above, one explanation for our surprising baseline lottery results is that individuals may delude themselves that they deserved their success, even when they did not. Recollect that in the *partial* information treatment, subjects were not informed about the source of their income (i.e. lottery or effort). In the absence of information, winners may persuade themselves that they deserved the boost in their income. Moreover, this self-serving belief would be reflected in the adoption of meritocratic social preferences. For instance, once you find out that you got rich by winning the lottery, it becomes especially hard to persuade yourself that you merited your high income. Therefore, once a winner finds out that their income was due to a lottery win, we should *not* expect them to adopt meritocratic social preferences. This mechanism can be ruled out in past work, due to features of the experimental design.

In our case, we randomly varied information to highlight this mechanism. To investigate the extent that lottery winners delude themselves that they were deserving, we compare lottery- and effort-task winners who were exposed to full and partial information. To do so, we employ the following empirical specification:

[17] This, as separate from the possibility that the hardworking feel more entitled to their earnings (Demiral and Mollerstrom 2020)

$$Tax_{is} = \beta_0 + \rho Lottery_i \cdot FullInformation_s + \alpha_0 Lottery_i + \alpha_1 FullInformation_s + \Gamma X_i + \varepsilon_{is} | Earnings = high \quad (3)$$

Everything is as before, except here we condition the sample of those who earn a high reward. Our focus is on the interaction between full information and lottery earnings, since that is where we expect information to be most relevant. In the absence of information on how they got rich, we expect that lottery winners become meritocratic on the basis of a self-serving illusion that their earnings were merited. Informing these subject that their high income is because they are lucky lottery winners, may make their distributional preferences less meritocratic (i.e. higher tax rate on effort income). We would expect that information would have little effect for those that actually did deserve the money, and that the source of income would not matter whenever people did not know the source of income. Our hypothesis is therefore that $\rho > 0$, while $\alpha_0 = \alpha_1 = 0$.

Our results in Table 2 provide us with evidence that is consistent with this behavioural mechanism. The table includes only individuals who earned a bonus either through the effort or luck task. Interestingly, under partial information, those that earned money in the effort task were not more meritocratic than lottery winners (e.g. Table 2 row 3, column 2, estimate = −0.12; $P = .978$). Likewise, being exposed to full information in the effort task had no meaningful effect on meritocratic preferences (e.g. Table 2 row 2, column 2, estimate=−6.6; $P = .198$). Nor should it, since we expected these subjects to believe they deserved the money anyway. However in contrast, we find a robust and consistent decrease in the meritocratic preferences (i.e. an *increase* in tax on effort) of lottery winners once they learned the source of their income (e.g. Table 2 row 1, column 2, estimate = 12.18; $P = .020$).[18]

Providing information to lottery winners on how they became winners (i.e. lottery or effort), resulted in a shift in social preferences making them *less* meritocratic.[19] This suggests that the large meritocratic shift that we initially observed in Fig. 2a was being driven almost entirely by lottery winners who, under partial information, adopted meritocratic social preferences, and made decisions as if their earnings were merited. In other words, once we revert to full-information, as in past work, our results more closely reflect theirs.

This suggests that our findings imply a crucial role for information about the source of an individual's success. On the one hand, this could be quite important if outside of the lab many people do not have complete information about the roles of luck and effort in their success. On the other hand, in the long-run people may be exposed to enough information that the meritocratic illusion amongst the lucky is unsustainable. We explore this further using the Wason task.

4.3.2 Motivated cognition and the Wason Card task

Outside of a lab environment people are constantly receiving new information and updating their beliefs accordingly. How then, might the gap in meritocratic beliefs persist? The Wason Task looks directly at the tension between the truth and the self-serving beliefs highlighted by the information treatment. Given that information can undermine the illusion that success was merited when it was not, we are interested in testing whether people will avoid this type of information. In this case we want to see how the cognitive dissonance generated by the tension described above is resolved. When individuals are faced with a decision between considering information inconsistent with their beliefs and earning more money, what will they do?

[18] Supplementary Appendix Table A8 shows robustness to controlling for effort in the effort experiment. This is to mitigate concerns that that the proposed tax rates are influenced by the exerted effort irrespective of the outcome of the lottery. Likewise in Supplementary Appendix Table A9 we add earnings threshold fixed-effects, for similar reasons. None of these additional controls influence the interpretation of the results.

[19] Lottery results by full/partial information are in Supplementary Appendix Table A10 and Supplementary Appendix Fig. A3. The (lack of any) effect of information in the effort experiment can be seen in Supplementary Appendix Fig. A4. These results are all consistent with Table 2.

Table 2. Information about income source on meritocratic preferences.

Sample	High-earners in either lottery or effort experiment					
Dependent variable	Tax on effort income			Effort tax over luck tax		
	(1)	(2)	(3)	(4)	(5)	(6)
Full Information x Lottery Experiment	13.15***	12.18**	12.59*	0.339**	0.328**	0.330*
	(4.341)	(4.806)	(6.428)	(0.155)	(0.156)	(0.160)
Full Information	−7.554*	−6.616	−7.030	−0.175	−0.164	−0.165
	(4.370)	(4.967)	(5.386)	(0.119)	(0.119)	(0.121)
Lottery Experiment	2.959	−0.123	−3.092	−0.0510	−0.0868	−0.104
	(3.461)	(4.353)	(5.208)	(0.0971)	(0.106)	(0.0986)
Effort: Linear		√			√	
Effort: Cubic polynomial			√			√
Gender	√	√	√	√	√	√
Age	√	√	√	√	√	√
Political Attitudes	√	√	√	√	√	√
Academic Department	√	√	√	√	√	√
Observations	162	162	162	162	162	162
R^2	0.212	0.234	0.246	0.218	0.228	0.229
Dependent Variable Mean	24.33	24.33	24.33	0.524	0.524	0.524

Standard errors are two-way clustered at the session and academic department level.
***, **, and * represent statistical significance at the 1 per cent, 5 per cent, and 10 per cent levels, respectively.
The table shows that information about income source influences the meritocratic bias. In columns 1–3, meritocratic preferences are measured as the tax rate that an individual imposes on high effort task income, earned by others (i.e. excluding themselves). In columns 4–6, these preferences are measured as the tax rate that an individual imposes on high effort task income *relative to lottery income*, earned by others (i.e. excluding themselves). The table presents results on treatment effect heterogeneity by whether respondents randomly received partial or full information about their income. All respondents in the table were high-earners, and we compare tax rates by income source, and information about income source. Columns 2 and 4 include a linear effort control, while columns 3 and 6 add a cubic polynomial in effort. In each case, effort is measured by the number of correctly completed addition tasks.

We investigate this cognitive dissonance using the Wason Card Task. Observe that if people avoid information to protect their meritocratic beliefs, we should expect a subject's performance on the Wason Task to (1) differ across Wason Task 1 versus Task 2 in the partial information treatments (2) be relatively similar across Wason Task 1 and Wason Task 2 in the full information treatment. As we have already seen, there is no need to protect a meritocratic belief in the full information treatment, since in that case it has already been undermined. In other words, there is no sense for someone to avoid information that they have already processed.

Indeed, in the absence of information about the cause of success, it is much easier for lucky individuals to maintain a self-image of being meritorious and deserving ('rich' would delude themselves income due to effort and the 'poor' that they got unlucky). We might expect such a person to be much more likely to use cognitive reasoning to critically evaluate any statement that is unpalatable for their meritocratic self-image (Dawson, Gilovich, and Regan 2002). Accordingly, we hypothesize that Wason Task 2 is much more likely to be solved by the 'rich' (who are likely to find the hypothesis implicit in the statement more distasteful). Therefore, for the rich we should expect $W_2 - W_1 > 0$, where the proportion of subjects with high earnings who correctly solve the first Wason task is W_1 and for Task 2 is W_2.[20] For similar reasons, Wason Task 1 is more likely to be solved by the relatively 'poor' and we should expect $W_2 - W_1 < 0$. Notice that an appealing aspect of this empirical

[20] By correctly solve, we mean that they choose the two specific cards that need to be turned over to fully verify the statement under consideration.

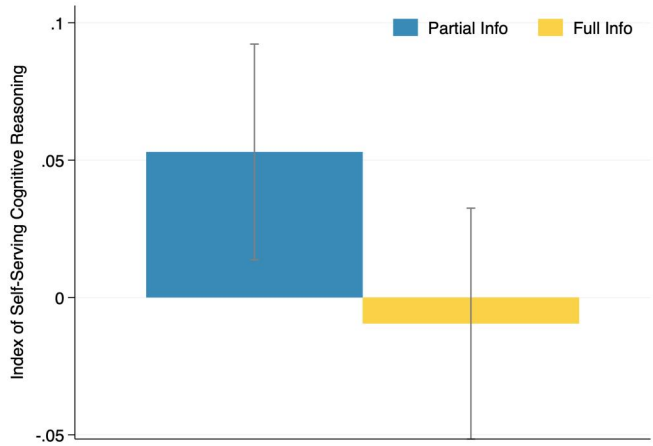

Figure 4. Quality of cognitive reasoning to protect self-image.

Note: The figure plots the tendency of subjects to solve the Wason tasks in a self-serving manner. The y-axis captures the asymmetry in cognitive reasoning between the two tasks, which may differ when a subject's self-image is at stake, that is, $I = |W_i - W_j|$, where $i, j \in \{1,2\}$. For the low earnings subjects, this means that they expend more effort to disprove the meritocratic statement (Wason Task 1) relative to Wason Task 2, that is, $W_2 - W_1 < 0$. For the high earnings subjects, this means the opposite $W_2 - W_1 > 0$.

design is that we examine *within-person* differences in performance across two (analytically identical) Wason Tasks.

Our results illustrated in Fig. 4 provide some evidence that people strategically avoid considering viewpoints that are inconsistent with their self-serving beliefs. On the y-axis is an index I that captures the asymmetry in cognitive reasoning between opposing statements, where $I = |W_i - W_j|$ where $i, j \in \{1,2\}$. Observe that it is under the partial information treatment that the asymmetry in cognitive reasoning is much higher. This is because it is precisely when there is an absence of information on their source of income that (1) the rich find it much easier to delude themselves that their high earnings were merited and the (2) relatively poor persuade themselves that were deserving but merely got unlucky by losing the lottery (see also Fig. 4 and Supplementary Appendix Table A11 columns 1 and 2).

The low-earners do much better at disproving the meritocratic statement (Wason Task 1), while the high-earners are better at disproving the unmeritocratic statement (Wason Task 2). This is reinforced in the more formal estimates, which show that the high-earners do much worse at disproving the meritocratic statement in both the lottery (Supplementary Appendix Table A11 column 1, estimate $= -0.248$; $P = .015$) and effort experiment (Supplementary Appendix Table A11 column 2, estimate $= -0.498$; $P = .015$) whenever information is incomplete. However, things change fairly dramatically under full information.

Under full information, it is much harder for subjects to delude themselves, and we would expect a much smaller difference in the expenditure of cognitive resources across the two Wason tasks. This is precisely what we see in Fig. 4 as well as Supplementary Appendix Table A11 columns 3 (estimate $= 0.0195$; $P = .723$) and 4 (estimate $= -0.087$; $P = .915$).[21]

Overall, the results from the Wason task are striking for a number of reasons. First, we see that even beliefs outside of the context of the experiment have shifted, and not just beliefs about the experiment itself.[22] Second, the results reinforce the importance of information in the persistence of divergent beliefs between those who are more and less

[21] Supplementary Appendix Fig. A5 shows the same result on the subset of lottery winners, and again, the pattern is the same.

[22] This reinforces work about information and polatization, for instance: Marino, Iacono, and Mollerstrom (2023).

successful. Indeed, the incorrect formation by the lucky of meritocratic beliefs generates a cognitive dissonance, between the true state of the world and the preferred one. And the Wason results suggest that to resolve this cognitive dissonance people are willing to incur financial losses in an effort to protect these beliefs, by endogenously avoiding viewpoints that might be inconsistent with them.

4.3.3 Group loyalty?

Here we examine the role of group loyalty effects in driving attitudes towards taxation. This is because such loyalty effects due to group identity—have the potential to confound our preferred explanation based on motivated reasoning. This is because the high-earning 'rich' (or poor) subjects may identify with other similar rich (respectively, poor) subjects, due to some kind of group/class based loyalty. This may give rise to an in-group bias, wherein 'rich' subjects impose relatively low taxes on effort income of other rich and the poor impose high taxes on the rich because of an out-group bias. In this case, the patterns of taxation may look very much like meritocratic social preferences—though in fact, these may merely be reflecting an in-group/out-group bias.

However, our results in Sections 4.3.1 and 4.3.2, together suggest that this particular mechanism is unlikely. Below, we briefly explain how.

1) Partial vs. Full Information treatment: We note that, in both these treatment, all subjects know whether they are 'rich' or 'poor'. The only difference is that those in the partial information treatment, do not know how they got rich/poor, i.e. through lottery winnings or performance in the effort task.

If the 'rich' were behaving as a group, we should expect *all the rich to propose very similar (low) taxes* on effort income—whether or not they know the source of their high earnings. In other words, we should not see any difference in the tax rates chosen by the rich under the partial versus full information treatment. However, as we can see from Table 2, the behaviour of the 'rich' under full information is significantly different from the rich in the partial information treatment.

2) The Wason Task: Our results from Section 4.3.2 also show a pattern that is remarkably consistent with (1)—the 'rich' do not behave as if in a group. In particular, here too we observe that the 'rich' behave very differently from each other—depending on whether they have information on the source of their income. It is the full information group that exhibits weaker cognitive reasoning in evaluating the claim that the rich are meritocratic.

Together (1) and (2) rule out the possibility of group identity driving our results. In Supplementary Appendix Table A1, we briefly discuss a subset of other biases that we rule out using our experimental design.

5. Concluding comments

Our results display a clear pattern across two independent experiments with two very different participant pools: higher inequality results in the successful, high-earning beneficiaries demonstrating a meritocratic bias. This bias provides a rationalization that enables individuals to justify their unequal rewards as being merited, even when those rewards are the result of pure chance. We can see strong evidence of this bias translating into a wedge between the high and low income subjects—in their meritocratic attitudes and the appropriate level of government redistribution. The results show us that higher earners cognitively process only information that boosts their meritocratic self-image and ignore evidence to the contrary.

Our work extends earlier work by Cassar and Klein (2019) and Deffains, Espinosa, and Thöni (2016).[23] In contrast to this work and the related literature described in the

[23] Cassar and Klein (2019) show that lottery winners engage in lower levels of redistribution from their lottery earnings, and those who earn from effort redistribute lower levels of effort-based income. In their article,

introduction, this article is specifically attempting to understand meritocratic beliefs by directly varying the information regarding the source of earnings. Our findings are consistent with the earlier literature and help rationalize the large difference between our main results and those presented in Cassar and Klein (2019) and elsewhere. We find that when people are given information about the source of their earnings, they are not able to maintain the incorrect self-serving belief about how deserving they are of their earnings. In particular, when we reveal to people that they earned their money through the lottery, they no longer redistribute more effort-based income.

At first glance this may seem like an optimistic result, as it identifies the important role that information can play in breaking down polarizing self-serving beliefs. It is a novel finding in the literature, and only possible with our introduction of randomized ambiguity in how income was earned. However, when people so readily update their belief in the face of information that reveals the true source of their income, this introduces a new question: do polarizing beliefs persist, and if so, how?

Using the Wason task, we provide the first evidence that people avoid allocating cognitive resources to issues that are inconsistent with their self-serving (and often incorrect) beliefs. The incentivized Wason task allows us to showcase the extent to which the beliefs formed in the earlier experiment is maintained even when it results in failing the task and thereby reducing the overall payoff in the experiment. This generates a direct tension between the desire to consider both a belief that enhances self-image, and an alternative belief that enables success in the Wason task and a resulting boost to payoffs. This is very much in the tradition of the cognitive dissonance literature which speaks to the difficulty of maintaining two competing sets of beliefs. The implication of our findings is that even though information updating does take place, it is very unlikely to occur in practice unless the information is made extremely salient. This implies that despite our evidence on belief updating, once beliefs are established they are far more likely to persist over time.

Supplementary material

Supplementary material is available at the Oxford Economic Papers Journal online. These are the data and replication files and the Supplementary Appendix.

Funding

This work was supported by the Centre for Competitive Advantage in the Global Economy; and the Social Sciences and Humanities Research Council.

Conflicts of interest statement. The authors declare that there are no conflicts of interest related to this research.

Acknowledgements

We are grateful for comments from George Akerlof, Robby Akerlof, Lorenzo Casaburi, Rachel Kranton, Kory Kroft, Dina Pomeranz. We also thank seminar and workshop

there is no variation in information, so those in the lottery experiment do not redistribute more in the effort experiment, and vice-versa. In our context, we leave the source of income ambiguous in our baseline setting and so generate different results. In particular, we find that lottery-winners even redistribute less effort-based earnings, giving rise to the possibility that people convince themselves that their earned income is deserving even when it comes entirely from luck. Similarly, while Deffains et al. (2016) do include imperfect information in their experimental setting, they are not able to pin down the role of information because this is not experimentally varied in their research design. Further, they observe redistribution as a whole, but do not observe redistribution specifically for earned and unearned income. Accordingly, since their study is not seeking to understand meritocratic beliefs it is not surprising that it is difficult in their design to distinguish between preferences for equality and meritocratic preferences.

participants at Georgetown University, University of Zurich, University of Warwick, University of Oxford, and Wilfred Laurier University. We are also thankful to the editor and two anonymous referees for their feedback, which improved the article. Ethical approval was granted by the University of Warwick Humanities and Social Sciences Research Ethics Committee, ref. HSSREC 113/20–21 & 72/16–17. The AEA registration and details of the pre-analysis plan for the online experiment are available online at: https://www.socialscienceregistry.org/trials/7355.

References

Akerlof, G.A., and Dickens, W.T. (1982) 'The Economic Consequences of Cognitive Dissonance', *The American Economic Review*, **72**, 307–19.

Alesina, A., and Fuchs-Schündeln, N. (2007) 'Goodbye Lenin (or Not?): The Effect of Communism on People's Preferences', *American Economic Review*, **97**, 1507–28.

Alesina, A., and La Ferrara, E. (2005) 'Preferences for Redistribution in the Land of Opportunities', *Journal of Public Economics*, **89**, 897–931. https://www.sciencedirect.com/science/article/pii/S0047272704000866

Alesina, A., Stantcheva, S., and Teso, E. (2018) 'Intergenerational Mobility and Preferences for Redistribution', *American Economic Review*, **108**, 521–54.

Almås, I., Cappelen, A.W., and Tungodden, B. (2020) 'Cutthroat Capitalism versus Cuddly Socialism: Are Americans more Meritocratic and Efficiency-Seeking than Scandinavians?' *Journal of Political Economy*, **128**, 1753–88.

Appiah, K.A. (2018) 'The Myth of Meritocracy: Who really Gets What They Deserve?' *The Guardian*,.

Arrow, K., Bowles, S., and Durlauf, S.N. (2018) *Meritocracy and Economic Inequality*, Princeton University Press.

Arrow, K.J. (1950) 'A Difficulty in the Concept of Social Welfare', *Journal of Political Economy*, **58**, 328–46.

Barr, A., and Miller, L. (2020) 'The Effect of Education, Income Inequality and Merit on Inequality Acceptance', *Journal of Economic Psychology*, **80**, 102276.

Bénabou, R., and Tirole, J. (2006) 'Belief in A Just World and Redistributive Politics', *Quarterly Journal of Economics*, **121**, 699–746.

Bénabou, R., and Tirole, J. (2016) 'Mindful Economics: The Production, Consumption, and Value of Beliefs', *Journal of Economic Perspectives*, **30**, 141–64.

Brown-Iannuzzi, J.L. *et al.* (2021) 'A Privileged Point of View: Effects of Subjective Socioeconomic Status on Naïve Realism and Political Division', *Personality & Social Psychology Bulletin*, **47**, 241–56.

Calonico, S., Cattaneo, M.D., and Titiunik, R. (2014) 'Robust Data-Driven Inference in the Regression-Discontinuity Design', *The Stata Journal: Promoting Communications on Statistics and Stata*, **14**, 909–46.

Cappelen, A.W. *et al.* (2021) 'Choice and Personal Responsibility: What is a Morally Relevant Choice?', *Review of Economics and Statistics*, 1–35. pp

Cappelen, A.W. *et al.* (2007) 'The Pluralism of Fairness Ideals: An Experimental Approach', *American Economic Review*, **97**, 818–27.

Cappelen, A.W. *et al.* (2013) 'Just Luck: An Experimental Study of Risk-Taking and Fairness', *American Economic Review*, **103**, 1398–413.

Cappelen, A.W. *et al.* (2017) 'The Merit Primacy Effect', *NHH Dept. of Economics Discussion Paper*, **1**.

Cappelen, A.W., and Tungodden, B. (2019) *The Economics of Fairness*. United Kingdom: Edward Elgar Publishing Limited.

Carpenter, J. (2010) 'Social Preferences' in *Behavioural and Experimental Economics*, pp. 247–252. United Kingdom: Springer.

Cassar, L., and Klein, A.H. (2019) 'A Matter of Perspective: How Failure Shapes Distributive Preferences', *Management Science*, **65**, 5050–64.

Cruces, G., Perez-Truglia, R., and Tetaz, M. (2013) 'Biased Perceptions of Income Distribution and Preferences for Redistribution: Evidence from a Survey Experiment', *Journal of Public Economics*, **98**, 100–12.

Dawson, E., Gilovich, T., and Regan, D.T. (2002) 'Motivated Reasoning and Performance on the Was on Selection Task', *Personality and Social Psychology Bulletin*, **28**, 1379–87.

Deffains, B., Espinosa, R., and Thöni, C. (2016) 'Political Self-Serving Bias and Redistribution', *Journal of Public Economics*, **134**, 67–74.
Demiral, E.E., and Mollerstrom, J. (2020) 'The Entitlement Effect in the Ultimatum Game–Does It Even Exist?', *Journal of Economic Behavior & Organization*, **175**, 341–52.
Doherty, D., Gerber, A.S., and Green, D.P. (2006) 'Personal Income and Attitudes toward Redistribution: A Study of Lottery Winners', *Political Psychology*, **27**, 441–58.
Fehr, D., and Vollmann, M. (2020) 'Misperceiving Economic Success: Experimental Evidence on Meritocratic Beliefs and Inequality Acceptance', Working Papers 0695, University of Heidelberg, Department of Economics. https://EconPapers.repec.org/RePEc: awi: wpaper : 0695
Festinger, L. (1959) *Cognitive Dissonance*. New York: Row, Peterson.
Fisman, R. *et al.* (2020) 'Do Americans Want to Tax Wealth? Evidence from Online Surveys', *Journal of Public Economics*, **188**, 104207. https://www.sciencedirect.com/science/article/pii/S0047272720300712
Giuliano, P., and Spilimbergo, A. (2014) 'Growing Up in a Recession', *Review of Economic Studies*, **81**, 787–817.
Hvidberg, K.B., Kreiner, C., and Stantcheva, S. (2020) 'Social Position and Fairness Views, Technical Report', National Bureau of Economic Research.
Jakiela, P. (2011) 'Social Preferences and Fairness Norms as Informal Institutions: Experimental Evidence', *American Economic Review*, **101**, 509–13.
Karadja, M., Mollerstrom, J., and Seim, D. (2017) 'Richer (and holier) than Thou? The Effect of Relative Income Improvements on Demand for Redistribution', *The Review of Economics and Statistics*, **99**, 201–12.
Klor, E.F., and Shayo, M. (2010) 'Social Identity and Preferences over Redistribution', *Journal of Public Economics*, **94**, 269–78.
Konow, J. (2000) 'Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions', *American Economic Review*, **90**, 1072–92.
Lerner, M. (1980) *The Belief in a Just World: A Fundamental Delusion*. Boston: Springer.
Littler, J. (2013) 'Meritocracy as Plutocracy: the Marketising of "Equality" under Neoliberalism', *New Formations*, **80**, 52–72.
Marino, M., Iacono, R., and Mollerstrom, J. (2023) '(Mis-) Perceptions, Information, and Political Polarization'.
Markovits, D. (2019) *The Meritocracy Trap*. UK: Penguin.
McCoy, S.K., and Major, B. (2007) 'Priming Meritocracy and the Psychological Justification of Inequality', *Journal of Experimental Social Psychology*, **43**, 341–51.
Meltzer, A.H., and Richard, S.F. (1981) 'A Rational Theory of the Size of Government', *Journal of Political Economy*, **89**, 914–27.
Peterson, E. (2016) 'The Rich are Different: The Effect of Wealth on Partisanship', *Political Behavior*, **38**, 33–54.
Piketty, T. (2020) *Capital and Ideology*. Cambridge, Massachusetts: Harvard University Press.
Rabin, M. (1994) 'Cognitive Dissonance and Social Change', *Journal of Economic Behavior & Organization*, **23**, 177–94.
Rawls, J. (1999) *A Theory of Justice*. Cambridge, Massachusetts: Harvard University Press.
Rey-Biel, P., Sheremeta, R.M., and Uler, N. (2011) '(Bad) Luck or (Lack of) Effort?: Comparing Social Sharing Norms between US and Europe'.
Romer, T. (1975) 'Individual Welfare, Majority Voting, and the Properties of a Linear Income Tax', *Journal of Public Economics*, **4**, 163–85.
Sandel, M.J. (2020) *The Tyranny of Merit: What's Become of the Common Good?* London: Allen Lane.
Sen, A. (2018) 'Merit and Justice', in S. B. Kenneth Arrow and S. N. Durlauf (eds) *Meritocracy and Economic Inequality*, pp. 5–16. Princeton: Princeton University Press.
Stantcheva, S. (2021) 'Why Do We Not Support More Redistribution? New Explanations from Economics Research', *Combating Inequality: Rethinking Governent's Role*.
Vu, T.V. (2023) 'Long-Term Relatedness and Income Distribution: Understanding the Deep Roots of Inequality', *Oxford Economic Papers*, **75**, 704–28.
Wason, P.C. (1966) Reasoning. In B. M. Foss (ed.) *New Horizons in Psychology*. Harmondsworth, UK: Penguin.
Young, M. (1958) 'The Rise of the Meritocracy', *Penguin. Qualitative Research in Education*, **1**, 133.

Oxford Economic Papers, 2025, 00, 1–20
https://doi.org/10.1093/oep/gpaf014
Article