DEPARTMENT
OF POLITICAL AND
SOCIAL SCIENCES

The Race is Not to the Swift.

The Role of Social Cognition in Reproducing Class-based Inequality.

William Foley

Thesis submitted for assessment with a view to
obtaining the degree of Doctor of Political and Social Sciences
of the European University Institute

Florence, 25 January 2023

European University Institute

Department of Political and Social Sciences

The Race is Not to the Swift.

The Role of Social Cognition in Reproducing Class-based Inequality.

William Foley

Thesis submitted for assessment with a view to
obtaining the degree of Doctor of Political and Social Sciences
of the European University Institute

Examining Board

Professor Klarita Gërxhani, EUI (EUI Supervisor)
Professor Arnout van de Rijt, EUI
Professor Cecilia L. Ridgeway, Stanford University
Professor Carlo Barone, Sciences Po

Researcher declaration to accompany the submission of written work
Department of Political and Social Sciences - Doctoral Programme

I William Foley certify that I am the author of the work "The Race is not to the Swift. The Role of Social Cognition in Reproducing Class-based Inequality." I have presented for examination for the Ph.D. at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

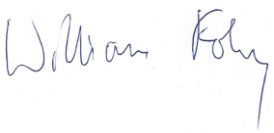
I certify that this work complies with the Code of Ethics in Academic Research issued by the European University Institute (IUE 332/2/10 (CA 297).

I declare that this work consists of 56,640 words.

Statement of inclusion of previous work:

I confirm that chapter 2 was jointly co-authored with Professor Klarita Gërxhani and Professor Arnout van de Rijt and I contributed 50% of the work.

Signature and date:



12/12/2022

Abstract

For sociologists, social class determines life outcomes in a myriad of different ways. But for the general public, class has never mattered as little. Class has become the "dark matter" of society, a potent yet invisible force shaping social reality. What accounts for the persistent, yet invisible power of class? In this thesis I argue that much of contemporary class inequality is driven by the widespread belief that "class = competence". This belief biases third party assessment of performance, directly affects performance and decision-making through internalising expectations, and buttresses the status quo by legitimising inequality as the result of meritocratic processes. Yet, potent as it is, the effects of this belief are not obvious because it appears to be nothing more than the common sense observation that those who achieve positions of power, wealth, and prestige are highly competent.

In this thesis, I explore how this belief is driven by social cognitive processes. In particular, I look at psychological heuristics designed to facilitate social action in the context of uncertainty. Each empirical chapter sets out to address three important mechanisms related to the "class = competence" belief. The first empirical chapter establishes that the "class = competence" belief leads to biased third party evaluations of performance. The second empirical chapter shows that working class students internalise the "class = competence" belief, decreasing their likelihood of attending university. The third chapter explores how the belief emerges by testing a mechanism based on cognitive dissonance reduction.

Contents

Acknowledgments

I would first of all like to thank my supervisor, Klarita Gërxhani. I am deeply grateful for her insight and motivation, and for how she indelibly marked how I understand the social world. She provided me with a sense of purpose and priorities, and facilitated my journey towards frontier lands of social analysis which otherwise would have remained unknown to me. I am also very grateful to Arnout van de Rijt for helping me to understand, among many other things, how to identify and present analytic solutions to problems of sociological importance.

I would also like to thank the (other) members of the thesis committee, Cecilia Ridgeway and Carlo Barone. Having read their influential and seminal contributions, it was an honour for them to comment on and critique my work.

Jan Skopek, who supervised my master thesis, provided the epistemological and practical springboard for my doctoral journey. Claudia Fanti provided inestimable and tireless administrative assistance throughout three-plus years of experimental administration and funding requests. The members of the interdisciplinary experimental working group provided rigorous critiques and invaluable advice during the design stage of each of the experiments presented in the thesis. The researchers and faculty of the European University Institute contributed to an inspiring intellectual atmosphere.

I'd like to thank friends and colleagues for the discussions, both formal and informal, which inspired the ideas behind this thesis: John Kennedy, Oisín Vince Coulter, Max Mansbart, Szymon Osmola, Grigoris Bacharis, John Trajer, Jan Dollbaum, Mar Cañizares, Francesco Colombo, Arianna Gatta, and many others. Special thanks to Eve Ryan for stimulating discussions and moral support.

Finally, I would like to thank my parents, Eilis and Anthony, without whom this would never have been possible. There is not enough space in this thesis to list my debts of gratitude. I dedicate this thesis to them.

Chapter 1. Introduction

I shall consider human actions and appetites just as if I were considering lines, planes, or bodies.

Spinoza.

According to popular wisdom, nowadays we truly live in the age of meritocracy. All careers are open to talent, and all inequalities are due to innate personal characteristics rather than social origins. Capitalism really has fulfilled its promise, having "pitilessly torn asunder the motley feudal ties that bound man to his 'natural superiors'" (Marx and Engels 2017:53). In this new age, we all get what we deserve.

But sociologists know, or think they know, better. Despite some brief popularity in the nineties (e.g. Pakulski and Waters [1996]), few scholars would now maintain that a person's social class plays no role in their life trajectory. Most researchers who study inequality (admittedly a self-selected cohort) would maintain quite the opposite. Put broadly, it is clear that class matters. The social class someone is born into plays a huge role in determining their life trajectory; in fact, it seems that social mobility has even declined somewhat in Western societies (Durlauf and Seshadri 2018; Hufe, Kanbur, and Peichl 2022; Jerrim and Macmillan 2015), meaning that the role of social origin has, if anything, been magnified.

Yet in an analysis of survey data from 24 Western countries, at least two-thirds of respondents in each country reported the belief that they live in meritocratic societies, and in half of them this proportion was at least three-quarters (Mijs 2021). Moreover, the belief that economic success derives from merit makes people more likely to accept inequality (Ahrens 2020; Alesina and Angeletos 2005), extinguishing class resentment. While class continues to matter, perhaps more than ever, it has never mattered so little in popular consciousness. Class has become the dark matter of society: a powerful yet invisible force which potently shapes the structure of social reality. How can this be explained?

The basic answer I propose in this thesis is that people widely believe the following equation to be true: "class = competence". In other words, they believe that high class people are generally more competent than lower class people, particularly when it comes to the skills, faculties, and abilities required to achieve economic success. Importantly, this belief is *not*

just a neutral appraisal of real population-level differences between social classes. Rather, it is a cognitive bias whereby objectively equal performances of higher and lower class persons will be judged unequally, such that the high class person will be seen as having performed better than the low class person.

This belief that "class = competence" is highly consequential for inequality since it works through several different mechanisms. In this thesis, I explore three of the most consequential mechanisms. First, the belief biases third-party assessments of individual ability, meaning that higher class people are more likely to receive rewards, remuneration, and access to valuable opportunities. Secondly, individuals internalise these beliefs, meaning that higher class people form more positive self-assessments of their ability and lower class people form more negative self-assessments. These self-assessments affect performance and influence decisions about what path to pursue in life. For example, if a lower class person may be less likely to go to college because they don't think they will succeed academically. Thirdly, the "class = competence" belief legitimates what might otherwise be seen as unfair distributions of rewards since it convinces people that the distribution of income and wealth is based on distinctions of individual merit rather than structural advantages and disadvantages. This mitigates against efforts to address the structural causes of class-based inequality.

These three mechanisms make the "class = competence" belief a powerful engine of inequality. Moreover, the belief has the power to render class inequalities invisible because the belief *seems* to be nothing more than the simple, common sense proposition that people who occupy prestigious and well-paying positions are able and hard-working. The combination of potency and invisibility makes the "class = competence" belief a prime candidate for explaining the "dark matter" of class inequality.

Put in these broad terms, this proposed explanation is not necessarily an original one. Bourdieu has pointed out how arbitrary markers of class distinction may be "misrecognised" as objective measures of merit (Bourdieu 1986). There is a large quantity of work exploring how class-based inequality can be reinforced, through various pathways, by assumptions concerning the relative ability of higher and lower class people. Given that the signals of class belonging, and the classed judgments they elicit, can be subtle, fluid and contextual, much of the work exploring these processes has naturally employed qualitative methodology (cf DiMaggio 2012; Friedman and Laurison 2020; Lamont 2000). That is not to say that there

is not also a large volume of illuminating work taking a quantitative approach, both experimental and non-experimental (cf Batruch, Autin, and Butera 2017; Jæger 2011; Rivera and Tilcsik 2016; Sullivan 2001). Both approaches have shed light both on the general empirical trends and associations, relevant concepts, and potential mechanisms and, in sum, have made immense contributions to our understanding of inequality between social classes.

However, there are several issues which the literature either has not tackled or has not overcome, each associated with the three mechanisms I identify above. Firstly, it is hard to assess whether third-party evaluations of an individual's ability are biased by the individual's social class because class is highly correlated with objective signals of competence such as educational attainment and occupation-relevant skills. So when we observe, for example, apparent discrimination against a working class person, what we may in fact be observing is that the 'discriminator' is simply using relevant indicators of performance correlated with class. An analogous example of this from a gender context is the fact that men typically pay higher premiums for car insurance (Aseervatham, Lex, and Spindler 2016). This is not likely to be due to a bias against *men* per se, but because male gender is a proxy for otherwise unobservable tendencies such as risk proclivity.

Secondly, although it might be theoretically plausible that people internalise "class = competence" beliefs about their class, it is very difficult to isolate the effect of these beliefs. The essential problem is that the internalised belief is "attached" to objective characteristics. For example, we observe that individuals from a working class background are less likely to attend college than their higher class peers, even controlling for academic performance. It might seem plausible that the difference in college attendance is due to working class students internalising the "class = competence" belief, which causes them to believe they are not competent enough to succeed in college. However, the belief that an individual internalises is naturally correlated with their objective class position – e,g, if they are working class, they internalise the belief about working class people. Hence it is hard to separate the effect of the belief from the effect of the many other factors associated with objective class position, such as financial resources, parental expectations, and relative opportunity costs.

Thirdly, it is unclear how the "class = competence" belief becomes widely held in the first place. For starters, holding this belief requires an individual to make a judgment about the distributions of income and competence in their society and the correlation between both

variables. However, people have only imperfect knowledge of the income distribution (Cansunar 2021), so how can they form such a judgment? Secondly, holding the "class = competence" belief requires working class people to, at least implicitly, derogate their own competence. It is not clear why they should acquiesce to such a belief, which would naturally have negative effects on their own self-esteem and sense of self-worth.

In this thesis, I aim to address each of these issues, taking an approach that is both theoretically oriented to the uncovering of mechanisms and empirically oriented to the identification of causal effects. I begin with the basic insight that people use heuristics to navigate a social world characterised by uncertainty. Such heuristics include the fundamental "class = competence" belief, but are not limited to that. I will also consider how heuristics (such as reference group comparisons) determine who the belief about class-based competence is applied to, and how heuristics for reducing cognitive dissonance could give rise to this belief. In doing so, I will aim to answer three broad research questions:

RQ1: Is there a widely-held belief that "class = competence", which leads to biased evaluation of performance?

RQ2: Do internalisations of the "class = competence" belief substantively contribute to inequality?

RQ3: How does the "class = competence" belief arise?

The scope of the ambition is pithily summarised in the epigraph to this chapter. The extent to which I fulfil this ambition is addressed in the conclusion.

1.1 Theoretical approach

1.1.1 The objectivity of uncertainty

Early sociological theorists often grappled with a paradox: to any given human, the rest of human society seems like an objective external structure. Yet that objective structure consists of the subjective actions and beliefs of the people who make it up. This conundrum was encapsulated by Marx, who said that humans "make their own history, but they do not make it just as they please; they do not make it under circumstances chosen by themselves, but under circumstances directly encountered, given and transmitted from the past" (Marx 1996:32). On this issue, it is common to divide Durkheim and Weber into opposing camps – Durkheim insisting on the objectivity of society, and Weber on the subjectivity of human

understanding and action. But the dichotomy is false, as Berger and Luckmann (1971) have pointed out. It is rather that to each individual, the subjective actions and beliefs of others will confront them as an objectively given fact.

I am interested in a subset of these beliefs which consist of heuristics to facilitate action in contexts of uncertainty. In order to form beliefs, people process information gleaned from their (social) context. However, this information is very frequently imperfect and the situation characterised by subjective and objective uncertainty. Müller-Kademann, adapting previous usages, outlines a threefold typology of uncertainty (2019):

Risk. People will usually not know with certainty whether a particular event will occur, or whether a possible state of the world will turn out to be true.

Ambiguity. Even when aware of all possible states, they will often be unable to form even an approximate probability distribution over all of them.

(Strict) Uncertainty[1]. Very often, people will simply be unaware of the existence, or possible existence, of a particular state of the world.

In short, we can (i) in the case of risk, be aware of all states and have a probability distribution over them; (ii) in the case of ambiguity, be aware of all possible states but lack a known probability for some or all of them; or (iii) in the case of strict uncertainty, be unaware of one or more possible state. I will generally use the term "uncertainty" in the looser sense, to cover all three case. However, it is worth reflecting on just how common even strict uncertainty is in everyday life. Introspection will tell us that one thing that we can always "expect" is the unexpected; people are constantly exposed to unanticipated events or revelations, especially in modern, fluid, and complex societies. The second type of uncertainty, ambiguity, is probably even more common in everyday life. Even when we are aware of all possible states, it may be very difficult for us to say which is likely to occur and which not. In short, there are few things in life we can be certain of, and so the information we have about the state of the world, even on a local level, is inherently imperfect.

[1] Muller-Kademann just refers to this type simply as "Uncertainty", but I add the adjective here to distinguish it from the more general concept of uncertainty which encompasses all items in this threefold typology.

Asserting the ubiquity of uncertainty is not necessarily a novel idea. The author of the biblical book of Ecclesiastes lamented the arbitrary connection between directed human action and its consequential effects: "I returned, and saw under the sun, that the race *is* not to the swift, nor the battle to the strong, neither yet bread to the wise, nor yet riches to men of understanding, nor yet favour to men of skill; but time and chance happeneth to them all." (Ecc. 9:11). In more recent times, John Maynard Keynes stressed that "we have, as a rule, only the vaguest idea of any but the most direct consequences of our acts" (Keynes 1937:213). And Frank Knight, "founding father" of Chicago economics, thought that uncertainty was an inevitable consequence of free will. Not being automata, human behaviour could never be perfectly predicted (Knight [1921] 1971). The "objective" social world, consisting as it does, of the beliefs and actions of other humans possessed of free will and driven by emergent creative processes, will always be characterised by uncertainty, even in the strict sense of the term.

1.1.2 Cognitive heuristics

Despite the constant presence of uncertainty in the social world, people must nonetheless form beliefs and take actions, which in turn requires that they must infer knowledge about the state of the world. Keynes supposed that people adopt a variety of conventions in order to form judgments under uncertainty and facilitate action (Keynes 1937). Hayek, who made the inescapability of uncertainty the cornerstone of his political philosophy and social theory, thought that humans evolved to learn, by trial and error, certain rules of conduct which would allow them to navigate environments (first physical, later social) about which they need not, and perhaps could not, have much knowledge of at all (Hayek 2013 [1973]). Tilly (1998) argued that human beings make use of social "scripts" to coordinate interaction under conditions of uncertainty, and that these scripts, when reproduced in an organisational context, reproduce and generate inequality between ascribed social categories.

The study of these conventions, rules of conduct, and scripts – and the biases they can sometimes give rise to – has been a long term pre-occupation of scholars of social cognition. In the middle of the 20th century, Fritz Heider developed his study of "common-sense" psychology, in which he argued that humans arrive to their own "naïve" theories about human psychology through formulating hypotheseses to explain the behaviour of others (Heider 1958). An example of such naïve theorising are attribution processes, where people attribute behaviour either to the "internal" characteristics of an individual, or to the "external" effects of the social environment (Kelley and Michela 1980). For example, someone's road

rage incident could be attributed either to their intemperate personality (an "internal" attribution) or stress caused by a long traffic jam on a hot day (an "external" attribution). While this can often lead to successful inference, it can also lead to systematic biases, such as the "fundamental attribution error" (Ross 2018), whereby people are likely to attribute behaviour which is the result of environment or circumstances to autonomous internal dispositions of an individual.

In opposition to the model of the "naïve scientist" that emerges from the theories of Heider and those that followed him, Fiske and Taylor proposed the model of the "cognitive miser" (Fiske and Taylor 1991). While the naïve scientist operates under certain constraints which may lead them astray, they generally act in a rational way, and their failures of cognition may, to some degree, be rendered analogous to the type of failures that befall "real" scientists – such as confounding with unobserved variables. The cognitive miser, on the other hand, wishes to preserve precious mental resources and energies. While naïve scientists might engage in a long and slow process of mental reasoning, cognitive misers will use cognitive heuristics – "good enough" rules of thumb to help them preserve mental energy. Tversky and Kahneman (1974) argued that such heuristics are particularly apt for forming judgments under uncertainty.

Here I use the term "cognitive heuristic" in a broad sense to refer to such rules or conventions which allow people to categorise information, form beliefs, and take action in the context of uncertainty. Although the terminology of "heuristic" may be more complementary with the cognitive miser picture of social cognition, it can also describe naïve scientist-style processes. Ultimately, both forms of cognition are at play when reasoning is conducted under uncertainty, and both can lead to biases and distortions. In the empirical chapters of this thesis, I explore how the use of such heuristics gives rise to cognitive biases and distortions that help reproduce social class-based inequality, primarily through the belief that "class = competence".

1.1.3 Status

The fundamental heuristic explored in this thesis is the concept of the "status belief", the paradigmatic concept of "status characteristics theory". According to status characteristics theory, or "status theory" for short, there are widely shared cultural beliefs which rank "social types" in a hierarchy of competence (Ridgeway 2019). Because of these status beliefs,

historically advantaged groups such as men and white people will be seen as inherently more competent than historically disadvantaged groups such as women and black people. Because in western societies, general competence is associated with respect and esteem, high status groups will also benefit from greater respect and esteem (Ridgeway 2019).

Status theory was developed by Cecilia Ridgeway, with significant contributions later being made by her collaborator Shelly Correll. The theory emerged from the small-group research of Bales et al (Bales et al. 1951), and the expectation states paradigm developed by Berger and colleagues (Berger, Cohen, and Zelditch Jr 1972) which sought to explain the unequal patterns of influence and interaction within small groups. Owing to these roots, one of the original scope conditions of status theory was that the status beliefs would operate in situations where a group was collectively working together to complete a task.

Correll, following work by Erickson (1998), extended the scope of the theory significantly by weakening the collective task orientation condition (Correll 2004). She argued that the same logic of relative comparisons that drove status processes in collective task settings would also be at work in individual task settings. In collective task settings, people form expectations of relative competence based on relative position within the status hierarchy. In individual task situations, anticipation of ranking creates pressure for individuals to estimate their performance in comparison to others who they believe will also be ranked. Here again uncertainty plays a crucial role. Since there is uncertainty over performance, and over the criteria for grading performance, individuals will use status beliefs to form expectations of their own competence viz-a-viz others'.

Ridgeway and Correll argue that the operation of status beliefs in individual settings is supported by other, closely-related theories such as stereotype threat (Steele 1998) and double standards theory (Foschi 2000), and they have included such theories under the broader umbrella of the status paradigm. Stereotype threat theory has provided evidence that awareness of stereotypes – i.e. widely held beliefs about relative competence – negatively affects the performance of lower status groups (Spencer, Logel, and Davies 2016). Double standards theory has shown that the performance of people from lower status groups is judged more harshly than that of people from higher status groups, even when performance is objectively identical (Foschi 2000).

Status beliefs are clearly heuristics, in the sense defined above. They are rules of thumb for categorising people, forming beliefs, and deciding on courses of action, and which serve to overcome uncertainty about competence and performance. Does status theory fit more into the cognitive miser or naïve scientist typology? It is difficult to say, and there are, perhaps, elements of both. According to status theorists, status beliefs emerge when differences in resource and power structures give Group A a systematic advantage in performance, influence and esteem over Group B (Ridgeway 2000). Personal experience will lead individuals to associate the salient characteristics of that social type with greater competence, giving rise to status beliefs. This mechanism is arguably an example of the "fundamental attribution error", which would place it in the "naïve scientist" account. People are rationally acting on a correlation between the variables they observe, though they mistakenly inflate the correlation because they omit the crucial variable of wealth or power resources.

On the other hand, there is a "cognitive miser" aspect to the theory as well. Status beliefs arise from "biased cognitive processes acting on ostensibly accurate performance information" (Correll and Benard 2006a:99). If a naïve scientist were to receive ostensibly accurate performance information showing that women perform as well as men, they would then revise their beliefs. But belief revision requires the expenditure of cognitive energy and – given the interconnection of beliefs – may entail the revision of other potentially salient beliefs, giving rise to cognitive dissonance. A cognitive miser would avoid this expenditure of mental energy.

1.1.4 Status and class

Status is a powerful tool for explaining how inequality can be reproduced through social cognitive processes. The pathways through which status can work are manifold. On the "demand side", gatekeepers and evaluators such as employers, teachers, and educational administrators will be biased by status beliefs in favour of high status individuals and against lower status ones. On the "supply side", status beliefs affect competence through boosting or knocking confidence and self-assessed ability, and encourage individuals to self-select into or out of prestigious jobs and educational tracks.

It would seem, therefore, that status theory would be an ideal sociological tool for explaining the "class = competence" belief. However, up to now, there has been virtually no work done on social class within the status theory paradigm. More recently, status theorists have

discussed the manner in which class can function as a status characteristic, albeit emphasising that some modifications or extensions are required to make status theory work in this context (Ridgeway 2019; Ridgeway and Fisk 2012). But there are few, if any, empirical papers which investigate the role of status in reproducing social class inequality. Ridgeway and Fisk state that "understanding the status dynamics of cross-class encounters is not a simple matter; it requires that we go well beyond standard accounts of interpersonal status process such as those offered by status characteristics theory." (Ridgeway and Fisk 2012:132).

There are at least three important difficulties in extending status theory to the domain of class. Firstly, the "salience" scope condition requires that social class be a readily perceptible nominal characteristic in order to be convertible into a status characteristic. But some might question how "obvious" an individual's social class might be. Secondly, there are strong conceptual disagreements among scholars over what class is, which are likely reflected in "lay" understandings of class. So how likely is it that consensual, widely-shared beliefs can form about such a seemingly unstable and ambiguous category?

Thirdly, social class is, to some degree mutable, and is in many ways popularly regarded as an achieved rather than ascribed characteristic (Ridgeway and Fisk 2012). Gender and ethnicity, while socially constructed and malleable, are in most cases effectively assigned at birth and strongly linked to obvious biological features. On the other hand, people can change social class. Moreover, in the age of meritocracy, social class is popularly associated with achievement and ability – people in higher class occupations are perceived as having achieved more, and got there on their merit. This may be seen as evidence for the strength of status beliefs. But we cannot assume what we have not yet proved. And it is difficult to empirically establish that status beliefs are attached to social class because it is closely associated with metrics of individual productivity and scholastic competence, making it difficult to "net out" any tendency to discriminate over-and-above appraisals of a seemingly neutral indicator of competence (Rivera and Tilscik 2016).

1.1.5 Class

It is not my intention to settle the debate – which has lasted as long as sociology has been a modern academic discipline – concerning the proper conceptualisation of social class. Rather I intend to provide a brief overview of my broad approach to class, and how that approach is operationalised in the thesis.

Broadly speaking, I conceive of social class as one's objective position within the system of relationships that characterises the prevalent mode of production. A position in a relationship entails a role, which entails both duties and rights. Relationships may also be hierarchical of course, and class relationships are usually conceived of in a hierarchical way, with income, authority, and prestige all increasing functions of position in the class hierarchy. Obviously, relationships have many dimensions, and so there are, in principle, many ways to divide a system of relationships into a class structure. Arguably the primary dimension of the class relationship is the Marxist distinction between those who own the means of production – the capitalists – and those who do not – the working class, in the Marxist sense of the term, who sell their labour to the owners of capital.

Certainly, class distinctions are not exhausted by this dichotomy, as the category of those who sell their labour technically includes all employees, from waiters to software engineers. Modern class analysis makes – and, to a large extent, focuses on – distinctions within the category of employees. The primary hierarchical distinction that is made is between blue collar and white collar workers – or labour and service contract employees, to adopt the terminology of the "EGP" framework (Erikson and Goldthorpe 1992), the most influential class schema in modern sociology (Smallenbroek, Hertel, and Barone 2021).

This labour / service distinction can also be derived from the individual's objective position within the system of relationships that characterise the mode of production. The distinction arises when we consider the nature of the work performed by the employee, and how that determines their bargaining power in relation to their employer. Labour contract workers tend to have low "asset specificity" and the costs of monitoring their performance is low (Goldthorpe 2000). In other words, it is easy to swap out their labour for that of another, and to keep tabs on how hard they are working. Hence, they can be hired on a short-term basis, paid per hour of labour, and easily replaced if they are not meeting their targets. Service contract employees have higher asset specificity and monitoring costs. They cannot be easily replaced and it is difficult to tell, on a day-to-day basis, whether they are performing at an adequate level without a costly inspection of their work. The distinction in terms of asset specificity and monitoring costs entails a different relationship between worker and boss – with the service contract employee enjoying a less hierarchical relationship, better remuneration, employment security, and prospects for promotion. Another, increasingly prominent – and newer approach – creates a class schema based not just on hierarchical

distinctions between based on authority, but horizontal distinctions based on "work logic" (Oesch 2006).

I tend to agree with Wright's argument that the Marxist, Weberian, and other mainstream traditions of class analysis both share valuable insights, and also have more in common than is usually conceived (Erik Olin Wright 2015). On the one hand, though the emphasis on ownership of the means of production is primarily associated with Marxian approaches, such a distinction is recognised, and even regarded as "primary", by the most the theoretical progenitor of the labour / service contract distinction (Goldthorpe 2000:1581). On the other hand, although Marxist approaches have tended to emphasise the unity of the employees, and hence to ignore the white / blue collar distinction, such distinctions can nonetheless, I would argue, be derived from Marxist analytical framework. Ultimately the Marxist story of class relations is a story of bargaining power. Under capitalism, the proletariat are in a sorry position because they lack capital and because their reduction to routine components of an industrial machine makes them highly substitutable: "Unfitted by nature to make anything independently, the manufacturing worker develops his productive activity only as an appendage of that workshop" (Marx 1992:482). However, in the passages dealing with division of labour in production, Marx distinguishes between workers who engage in only one routine activity in producing only a composite part of a commodity, and workers who freely engage in a number of different activities in the production of an entire commodity. The former type of labour is more regularised and policed than the latter, and also more exploited, in the sense of capturing a lesser share of the revenue. The distinction resembles that of the labour and service contract.

I also agree with Wright's point that, that to a large extent, the appropriate conceptualisation of social class depends partly on the level of analysis the researcher is working at (Erik Olin Wright 2015). Whether one focuses on ownership of the means of production, intra-employee distinctions, or differences in cultural knowledge and tastes, depends on the research question at hand. Explicitly Marxist approaches are best when the researcher is working at a systemic level, thinking in terms of longer time periods and broad changes to the social structure. The Marxist framework of struggle between big classes – peasants, lords, workers, bourgeoisie – can very adequately explain, for example, the path that different societies followed in their path to modernity (Anderson 2013; Brenner 1976; Moore 1993).

On the other hand, making finer distinctions within the broad category of employee is very useful when researching topics such as social mobility, support for political parties, the incidence of disease, and patterns of fertility or household formation. Indeed, focussing on the hierarchical distinction between labour and service classes may be most useful when trying to explain support for redistribution, whereas focussing on the horizontal distinction between "work logics" may be most useful when trying to explain support for same-sex marriage or minority rights. In these more fine-grained approaches, the capitalist is usually a neglected and residual category. For example, in the "EGP" schema, large employers are included in the "upper service class", alongside professionals. This is perfectly fine when the object of research is to document and understand population-wide regularities (Goldthorpe 2016), where the small size of the capitalist class makes it demographically negligible (Breen 2005). But, for that reason, such approaches will be found wanting if a researcher wishes, for example, to understand how economic elites directly and indirectly wield power in society (Lukes 2021).

Ultimately, the level and object of analysis is crucial for determining the class conception which the researcher adopts. In this thesis I am broadly concerned with inequality between the labour and service classes, which I usually refer to simply as lower class and higher class. Besides being in accord with the EGP classification, the most influential class schema in modern sociology (Smallenbroek et al. 2021), this bipartite distinction also seems to be reflected in popular, or vernacular, understandings of class structure, at least in the British case (Evans and Mellon 2016). This is relevant since the sample in Chapter 2 and Chapter 4 is drawn from a British population.

Perceptions are also of importance because they pertain to the level and object of analysis of this thesis. The thesis is concerned with status, which is a matter of perception, and, more generally, with the beliefs that people hold. Hence, it is crucial to take into account the fact that people do not always sort others into neat and consistent class categories. Though I generally conceive of class in a categorical and not a continuous way, there are certain attributes or correlates of class position – such as income, educational attainment, and occupational prestige – which can be conceived of as continuous, and which influence people's perception of their own, and of others, class position. At the margins, it is not always clear whether someone is "working" or "middle" class. Many occupations, for example, that tend to be placed in the "labour" category of the EGP schema often have remuneration

schemes and working conditions more characteristic of the "service contract" type (Breen 2005). In an environment where most people have labour contracts for example, having a bit more money, a bit more education, or a bit more autonomy and freedom in one´s work, can lead to a person being conceived as of higher class than they would be in an environment where most are on service contracts. In the second chapter, myself and my co-authors are aware of this distinction and control for it with the experimental design, in order to facilitate a binary low-high class analysis. In the third chapter, this ambiguity is exploited in order to uncover the effect of status net of objective factors.

1.2 Thesis overview

The empirical component of the thesis consists of three chapters.

1.2.1 Are there status beliefs about social class?

The second chapter is co-authored with Klarita Gërxhani and Arnout van de Rijt. It aims to address RQ1: Is there a widely-held belief that "class = competence", which leads to biased evaluations of performance? In particular, it aims to establish if there are status beliefs about class, which lead to discrimination in the assessment of individual performance.

Social class has been largely avoided as a focus of status research because unlike other status variables – ethnicity, gender, motherhood – class is strongly correlated with performance on status-relevant tasks. Productivity indicators such as education and occupational experience are conceptually intertwined with social class. As a result, a preference for high-class individuals can plausibly be attributed also to the rival theory of statistical discrimination, even after accounting for observable performance indicators. Statistical discrimination stipulates that discrimination arises when evaluators use differences in the group-level distributions of productivity to counteract uncertainty over individual ability. The strong connection between class and group-level productivity makes it hard to "control" for statistical discrimination in the case of social class by using conventional designs for studying status beliefs or discrimination.

Our innovation is to ask evaluators of higher and lower class workers to predict worker performance on a task using data on workers' past performance on that same task. We then study whether class discrimination tends toward zero as the number of data points on which past performance information is based increases. If instead discrimination stabilizes at some non-zero level, this provides evidence in favor of a status mechanism. We embedded this

design in an online experiment, studying also gender, a classic status variable, for comparison. A British sample of four hundred "evaluators" were hired to evaluate performance in two different tasks: a navigation task, where men should be expected to perform better; and a culture task, where higher class people should be expected to perform better. We find conclusive evidence in favor of status discrimination on social class, and little evidence for status discrimination on gender.

1.2.2 To whom do you refer? Studying the role of status in reproducing educational inequality between social classes.

Having, in Chapter 2, established the existence of a class-based status belief in third-party evaluations, I seek in the Chapter 3 to address RQ2: Do internalisations of the "class = competence" belief substantively contribute to inequality? Specifically, I seek to understand if the internalisation of status beliefs can help explain one of the most consequential forms of inequality between social classes: inequality of university attainment.

One problem with measuring the effect of status is that it is attached to particular ascribed characteristics such as gender, ethnicity, and class. It is therefore difficult to identify variation in status that is not confounded with the ascribed characteristic that it's attached to. I provide a solution by connecting status beliefs to the concept of a reference group. I argue that due to uncertainty about precise location in the class hierarchy, people use class position within their reference group as a heuristic for estimating their class position. Hence, an individual's reference group affects their perceived social class identity and, as a result, the status beliefs which are applied to them. Identifying exogeneous variation in reference group position should allow us to identify exogeneous variation in status beliefs.

One implication of this theory is that being placed in a classroom with a higher social class composition should *negatively* affect a working class student's educational attainment. Being placed in this setting, working class students will be seen by their peers as subjectively lower down the class ladder than they would be seen in a working class context. Hence they will also be seen as lower status and be disadvantaged by status beliefs stipulating that lower status individuals are less educationally competent. Internalising these beliefs will reduce their confidence to succeed in university, and make them less likely to attend.

This implication contradicts the consensus in the literature which maintains that having higher social class classmates benefits working class students through a positive peer effect

mechanism. But I contend that if status beliefs really do exert a powerful and socially consequential effect, they have the potential to overwhelm any positive peer effect.

I demonstrate this negative status effect in two different studies. The first implements a causal design using survey data from Sweden and Germany. The second study uses a vignette experiment in an Irish school. Taken together, the studies provide complementary evidence for a negative status effect generated by the reference group mechanism.

1.2.3 Worse things happen at sea. Testing a cognitive dissonance mechanism for belief in meritocracy.

The final empirical chapter aims to address RQ3: How does the "class = competence" belief arise? Empirical research has established that most people in most developed countries believe that their societies are meritocratic (Mijs 2021), and that people view economic inequality as legitimate when they think it is based on distinctions of merit (Ahrens 2020; Alesina and Angeletos 2005; Hufe et al. 2022). In other words, most people believe that higher class people have achieved their position through greater competence, and this encourages them to accept inequality.

Does belief that society is meritocratic emerge simply from neutral and accurate observation of society-wide distributions of income and merit? There are at least three reasons to doubt this. Firstly, as is well known, people have very imperfect knowledge of the income distribution in their country (Cansunar 2021). Secondly, belief in meritocracy is higher in countries which are more unequal (Mijs 2021), but more unequal countries are more socially immobile (Durlauf and Seshadri 2018; Jerrim and Macmillan 2015) so likely to be less meritocratic. Thirdly, endorsing the belief also requires that working class people "accept" they are incompetent and deserve less – why would people voluntarily endorse a belief, on the basis of imperfect evidence, that derogates themselves?

Following Elster's "sour grapes" theory, I argue that belief in meritocracy is a strategy for mitigating cognitive dissonance. People may experience cognitive dissonance when the belief that they live in a just society conflicts with an (apparently) unjust distribution of income. People can reduce dissonance by taking action to reduce inequality or improve their material position – an "action-based" strategy to reduce dissonance. But when action is impossible, they may convince themselves the distribution of income is fair because it is based on merit – a "belief-based" strategy. Once again, uncertainty plays a crucial role. People have only

incomplete information on the distribution of income and merit, giving them room to revise their normative and factual beliefs concerning income, effort and dessert.

In the experiment, 351 "Receivers" are matched with a "Proposer". The proposer offers them 40% or 10% of a money prize, corresponding to a low or high inequality scenario. Half of the receivers have the choice to reject the offer and half of them must accept it. Receivers are asked: (i) whether they think their proposer was gifted the money or earned it (by doing a maths task); and (ii) whether the offer was fair. I hypothesise that receiving a highly unequal offer may generate cognitive dissonance. Rejecting the offer is an action-based strategy to reduce dissonance. But if this option is not available, receivers may follow a belief-based strategy by convincing themselves that the proposer earned the money.

Findings show no difference between treatments when it comes to believing the proposer earned the money. However, Receivers who had no choice to reject the offer were more likely to think a highly unequal offer was fair. Qualitative analysis of open-ended questions suggests that these Receivers espoused a more pessimistic outlook on the nature of humanity, expressing the view that people should not be expected to treat others equally or with generosity. The results suggest that when lacking control over the distribution of income, people accommodate themselves to inequality through fatalistic resignation rather than by convincing themselves the distribution arises through merit.

1.3 Note on methodology

All three empirical chapters use experiments, and the experiments in chapters two and four are drawn from online samples. In this section I would like to briefly address common criticisms and concerns pertaining to the experimental method in general, and to online experiments in particular.

Perhaps the most common concern about experiments is that they lack "external validity". That is to say, the findings cannot be generalised from the sample to the actual attitudes and behaviours that people exhibit in social life. The primary issues with external validity arise from the fact that experimental samples are usually not random samples, and that the experimental context is rather artificial (List 2020).

To begin with the problem of "unrepresentative" samples, it is well known that participants in the vast majority of experimental studies are "WEIRD" – Western, Educated, Industrialised,

Rich and Democratic (Henrich, Heine, and Norenzayan 2010). Moreover, even within this population, samples can very often be skewed on key demographic variables – for example, undergraduate students are massively overrepresented amongst experimental samples (Arnett 2016).

There are a number of approaches one can take to address the problem of unrepresentative samples. Firstly, one should be very careful in making inferences beyond the demographic characteristics of the sample one has recruited. Secondly, one should try and diversify the sample as much as possible. I have tried to follow both approaches in this thesis. In the second chapter of this thesis, the experimental sample, though not randomly selected from the general population, is still quite diverse. Half of the sample is university educated, and half not. Half is female, and half is male. The sample median of 38 years old is not very far off the median British age of 40 years old[2], and the distribution was also fairly spread out, with a standard deviation of 13 years. In the third chapter, the sample consists of school students, which is appropriate since the theoretical argument concerned comparisons of school students to their student peers. Finally, in the fourth chapter, the sample was limited to participants without university degrees. This restriction was also theoretically motivated, since I was particularly interested in why working class people endorse the view that society is meritocratic. (Though I do not equate being working class with not having a university degree, most people in working class, or labour contract, style occupations do not possess one. This made it a convenient measure for targeting the sample). In other regards, the sample in this chapter is also reasonably diverse. It is gender-balanced, and has a median age of 44 years with a standard deviation of 13 years.

The diversity of the samples – within the limits of the target populations – should reduce the worry that results are driven by focussing on a narrow sliver of society. Nonetheless, since none of the samples were randomly selected from the population, it would be dubious to try and infer precise point estimates for population-level parameters. This is fine, if we conceive of the primary goals of experiments to uncover causal relationships, not to form precise point estimates. The aim is to test directional hypotheses: when x goes up, y goes down; and to be able to claim that the association between x and y is the causal effect of y. Since we don't aim

[2] https://www.statista.com/statistics/281288/median-age-of-the-population-of-the-uk/

to uncover precise point estimates, we do not have to worry so much if the experimental sample does not exactly map onto the target population.

The second big worry concerning the external validity of experiments is that the experimental situation is to some degree artificial, and so concerns might arise that participants behave differently in the experiment compared to how they behave in "real life". The experimenter should indeed take care that the participants are not unduly influenced by the fact that they know they are in an experiment. However, there are several reasons why I believe this worry should not cause too many problems in the case of the experiments presented in this thesis. Firstly, it is, in some sense, a "feature" not a "bug" that experiments are artificial. The point is to cut out noise and zero in on the theoretically relevant features of the causal relationship the experimentalist wishes to investigate. Secondly, how much the "artificiality" of the experiment matters depends on the human behaviour or attitude under investigation. Here the metaphor of a tree could be useful. The trunk of the tree can be taken to represent universal human traits; the branches can be taken to represent stable cultural divergences between human societies; and the leaves can be taken to represent easily disturbed and changeable human qualities which fluctuate substantially between and within individuals over time periods and contexts. The further "up" we go from the roots to the leaves, the more we would worry about that an experiment examining a given trait would suffer from major external validity concerns, since the trait is so sensitive to changes in context. Thirdly, if we are concerned that participants are not motivated to participate, or are perversely motivated to behave other than they would in the real-life situation we are trying to simulate, we may ensure that incentives are sufficiently large, and that they do not cause the participant to respond in a way that would be contrary to how they would in reality.

Given these considerations, how problematic is the "artificiality" of the experimental environment for the studies presented in this thesis? The second chapter considers whether participants discriminate when offered the choice to invest between high and low status individuals, with their payoff dependent on investing in the winning individual. Discrimination is a stable and durable feature of human societies, and of individual humans (Pearson, Dovidio, and Gaertner 2009; Vuletich and Payne 2019), hence we should not expect it to be very sensitive to contextual variations. Participants are monetarily incentivised, meaning that they should be motivated to pay attention and take the task seriously.

The experimental component of the third chapter asked participants to guess their classmates' views on what a hypothetical student would do after completing secondary school. This is a topic close to the daily lives of the student participants and concerns a consequential life decision to which at that stage – sixteen years old – they must have begun to give some serious thought (Winterton and Irwin 2012). Hence the experiment is asking them to "retrieve" pre-existing cognitive states rather than react to a novel stimulus, making their response more reliable. As with the second chapter, participants were monetarily incentivised to take the task seriously.

The third experiment aims to show how the experience of inequality affects attributions of effort or luck, and fairness judgments. These reflect attitudes relating to values and moral attitudes, dispositions which are known to be very stable by the time an individual reaches thirty (Kiley and Vaisey 2020; Vaisey and Lizardo 2016). Since about three quarters of the sample was over thirty, it is highly likely that the experiment was tapping into a stable trait. Furthermore, as with the first two experiments, participants were also incentivised.

Overall, I would argue that the three experiments elicit behaviours and attitudes based on stable pre-existing dispositions. Furthermore, they all use monetary incentives, which should ensure the participant's active compliance.

One final issue is that the sample was accessed online. Online samples can be considered less reliable since the researcher has less control over the participant's environment, and since internet participants are often conceived to have particularly "mercenary" motivations, aiming to speed through the experiment in order to maximise pay and minimise time spent (Snowberg and Yariv 2021). It is difficult for the researcher to do anything to gain control over an online participant's environment. Accepting less control is a trade-off that a researcher must make in return for improvements in other areas. For example, hosting an experiment online as opposed to in a lab allows the researcher to gain access to a wider and more diverse pool of participants, which can address the external validity question discussed above.

On the other hand, it is possible to improve the "quality" of participant responses. I sought to do this in two primary ways. Firstly, I tried to identify online environments which would yield an adequate quality of participant. For the experiments featured in the second and fourth chapters, I used the online platform Prolific to gain access to the subject pool. Unlike

Amazon's Mechanical Turk, which is also commonly used to hire participants, Prolific is specifically designed for the scientific community, and has been found to provide better quality participants (Eyal et al. 2021; Palan and Schitter 2018; Peer et al. 2017). For the third chapter, the experiment was conducted online through the Zoom video conferencing application. The rationale for the online format was that the research was being conducted during a period of the Covid-19 pandemic when children were attending school remotely. Since it was in a school context, with a teacher present, most students should have been motivated to perform the experiment with a certain degree of diligence because the presence of teachers in online environments increases student engagement (Bangert 2008; Roque-Hernández et al. 2021; Yu 2022).

The second way to improve participant quality is to offer substantial monetary incentives. In both experiments using the Prolific sample pool, the payoffs on offer were substantive and likely to be significantly above average. For example, in the main sample of the second chapter, the median earning was £5.22 and the median time taken to complete the experiment was 15 minutes, yielding an hourly equivalent wage of about £21. For the main sample of the fourth chapter the median time to completion was six and a half minutes, and the median earning was £3.45, yielding an hourly equivalent wage of about £31. These hourly wage equivalents are 2.3 and 3.4 times larger respectively than the £9 rate which Prolific *recommends* researchers pay, and 3.5 and 5.2 times larger than the minimum payment prolific allows of £6 per hour. For the school student sample in the third chapter, it is less easy to establish norms for good pay, but it is likely that the prospect of winning between 25 and 100 euro was motivating to the students, who normally do not earn money for completing activities during school hours.

1.4 Note on the title

The title of this thesis is taken from the Ecclesiastes verse already quoted above. The verse pithily encapsulates two core aspects of the thesis: the emphasis on the fundamental uncertainty of human endeavour, and the disassociation between true ability and actual (socioeconomic) success.

References

Ahrens, Leo. 2020. ‘Unfair Inequality and the Demand for Redistribution: Why Not All Inequality Is Equal’. *Socio-Economic Review* mwaa051. doi: 10.1093/ser/mwaa051.

Alesina, Alberto, and George-Marios Angeletos. 2005. ‘Fairness and Redistribution’. *American Economic Review* 95(4):960–80. doi: 10.1257/0002828054825655.

Anderson, Perry. 2013. *Lineages of the Absolutist State*. Verso Books.

Arnett, Jeffrey J. 2016. *The Neglected 95%: Why American Psychology Needs to Become Less American*. Washington, DC, US: American Psychological Association.

Aseervatham, Vijay, Christoph Lex, and Martin Spindler. 2016. ‘How Do Unisex Rating Regulations Affect Gender Differences in Insurance Premiums?’ *The Geneva Papers on Risk and Insurance - Issues and Practice* 41(1):128–60. doi: 10.1057/gpp.2015.22.

Bales, Robert F., Fred L. Strodtbeck, Theodore M. Mills, and Mary E. Roseborough. 1951. ‘Channels of Communication in Small Groups’. *American Sociological Review* 16(4):461–68. doi: 10.2307/2088276.

Bangert, Arthur. 2008. ‘The Influence of Social Presence and Teaching Presence on the Quality of Online Critical Inquiry’. *Journal of Computing in Higher Education* 20(1):34–61. doi: 10.1007/BF03033431.

Barone, Carlo. 2006. ‘Cultural Capital, Ambition and the Explanation of Inequalities in Learning Outcomes: A Comparative Analysis’. *Sociology* 40(6):1039–58. doi: 10.1177/0038038506069843.

Batruch, Anatolia, Frédérique Autin, and Fabrizio Butera. 2017. ‘Re-establishing the Social-class Order: Restorative Reactions against High-achieving, Low-SES Pupils’. *Journal of Social Issues* 73(1):42–60. doi: 10.1111/josi.12203.

Berger, Joseph, Bernard P. Cohen, and Morris Zelditch Jr. 1972. ‘Status Characteristics and Social Interaction’. *American Sociological Review* 241–55.

Berger, P., and T. Luckmann. 1971. *The Social Construction of Reality : A Treatise in the Sociology of Knowledge*. Penguin Books.

Bourdieu, Pierre. 1986. *Distinction : A Social Critique of the Judgement of Taste*. Routledge.

Breen, Richard. 2005. ‘Foundations of a Neo-Weberian Class Analysis’. Pp. 31–50 in *Approaches to Class Analysis*, edited by E. O. Wright. Cambridge: Cambridge University Press.

Brenner, Robert. 1976. ‘Agrarian Class Structure and Economic Development in Pre-Industrial Europe’. *Past & Present* 70(1):30–75.

Cansunar, Asli. 2021. ‘Who Is High Income, Anyway? Social Comparison, Subjective Group Identification, and Preferences over Progressive Taxation’. *The Journal of Politics* 83(4):1292–1306. doi: 10.1086/711627.

Correll, Shelley J. 2004. ‘Constraints into Preferences: Gender, Status, and Emerging Career Aspirations’. *American Sociological Review* 69(1):93–113. doi: 10.1177/000312240406900106.

Correll, Shelley J., and Stephen Benard. 2006. ‘Biased Estimators? Comparing Status and Statistical Theories of Gender Discrimination’. Pp. 89–116 in *Advances in Group Processes*. Vol. 23, *Advances in Group Processes*, edited by S. R. Thye and E. J. Lawler. Emerald Group Publishing Limited.

DiMaggio, Paul. 2012. ‘Sociological Perspectives on the Face-to-Face Enactment of Class Distinction’. Pp. 15–38 in *Facing social class: How societal rank influences interaction*. New York, NY, US: Russell Sage Foundation.

Durlauf, Steven N., and Ananth Seshadri. 2018. ‘Understanding the Great Gatsby Curve’. *NBER Macroeconomics Annual* 32(1):333–93. doi: 10.1086/696058.

Erickson, Kristan Glasgow. 1998. *The Impact of Cultural Status Beliefs on Individual Task Performance in Evaluative Settings: A New Direction in Expectation States Research*. Stanford University.

Erikson, Robert, and John H. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Clarendon Press.

Evans, G., and J. Mellon. 2016. ‘Social Class: Identity, Awareness and Political Attitudes: Why Are We Still Working Class?’ *British Social Attitudes* 33.

Eyal, Peer, Rothschild David, Gordon Andrew, Evernden Zak, and Damer Ekaterina. 2021. 'Data Quality of Platforms and Panels for Online Behavioral Research'. *Behavior Research Methods*. doi: 10.3758/s13428-021-01694-3.

Fiske, Susan T., and Shelley E. Taylor. 1991. *Social Cognition*. Mcgraw-Hill Book Company.

Foschi, Martha. 2000. 'Double Standards for Competence: Theory and Research'. *Annual Review of Sociology* 26(1):21–42. doi: 10.1146/annurev.soc.26.1.21.

Friedman, Sam, and Daniel Laurison. 2020. 'The Class Ceiling: Why It Pays to Be Privileged'.

Goldthorpe, John H. 2000. 'Rent, Class Conflict, and Class Structure: A Commentary on Sorenson'. *American Journal of Sociology* 105(6):1572–82. doi: 10.1086/210465.

Goldthorpe, John H. 2016. *Sociology as a Population Science*. Cambridge University Press.

Hayek, Friedrich August. 2013. *Law, Legislation, and Liberty*. London: Routledge.

Heider, Fritz. 1958. *The Psychology of Interpersonal Relations*. Hoboken, NJ, US: John Wiley & Sons Inc.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. 'The Weirdest People in the World?' *Behavioral and Brain Sciences* 33(2–3):61–83.

Hufe, Paul, Ravi Kanbur, and Andreas Peichl. 2022. 'Measuring Unfair Inequality: Reconciling Equality of Opportunity and Freedom from Poverty'. *The Review of Economic Studies* rdab101. doi: 10.1093/restud/rdab101.

Jæger, Mads Meier. 2011. 'Does Cultural Capital Really Affect Academic Achievement? New Evidence from Combined Sibling and Panel Data'. *Sociology of Education* 84(4):281–98. doi: 10.1177/0038040711417010.

Jerrim, J., and L. Macmillan. 2015. 'Income Inequality, Intergenerational Mobility, and the Great Gatsby Curve: Is Education the Key?' *Social Forces* 94(2):505–33.

Kelley, Harold H., and John L. Michela. 1980. 'Attribution Theory and Research'. *Annual Review of Psychology* 31:457–501. doi: 10.1146/annurev.ps.31.020180.002325.

Keynes, John Maynard. 1937. ‘The General Theory of Employment’. *The Quarterly Journal of Economics* 51(2):209–23.

Kiley, Kevin, and Stephen Vaisey. 2020. ‘Measuring Stability and Change in Personal Culture Using Panel Data’. *American Sociological Review* 85(3):477–506. doi: 10.1177/0003122420921538.

Knight, Frank H. (Frank Hyneman). 1971. *Risk, Uncertainty and Profit*. University of Chicago Press.

Lamont, Michèle. 2000. *The Dignity of Working Men : Morality and the Boundaries of Race, Class, and Immigration*. Russell Sage Foundation ; Harvard University Press.

List, John A. 2020. ‘Non Est Disputandum de Generalizability? A Glimpse into The External Validity Trial’.

Lukes, Steven. 2021. *Power : A Radical View*. Third edition. Red Globe Press.

Marx, Karl. 1992. *Capital: Volume 1: A Critique of Political Economy*. Reprint edizione. edited by E. Mandel. London ; New York, N.Y: Penguin Classics.

Marx, Karl. 1996. ‘The Eighteenth Brumaire of Louis Bonaparte’. Pp. 31–127 in *Marx: Later Political Writings, Cambridge Texts in the History of Political Thought*, edited by T. Carver. Cambridge: Cambridge University Press.

Marx, Karl, and Friedrich Engels. 2017. *The Communist Manifesto*. Pluto Press.

Mijs, Jonathan J. B. 2021. ‘The Paradox of Inequality: Income Inequality and Belief in Meritocracy Go Hand in Hand’. *Socio-Economic Review* 19(1):7–35. doi: 10.1093/ser/mwy051.

Moore, Barrington. 1993. *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*. Vol. 268. Beacon Press.

Müller-Kademann, Christian. 2019. *Uncertainty and Economics: A Paradigmatic Perspective*. Routledge.

Oesch, Daniel. 2006. ‘Coming to Grips with a Changing Class Structure: An Analysis of Employment Stratification in Britain, Germany, Sweden and Switzerland’. *International Sociology* 21(2):263–88.

Pakulski, Jan, and Malcolm Waters. 1996. *The Death of Class*. SAGE.

Palan, Stefan, and Christian Schitter. 2018. 'Prolific.Ac—A Subject Pool for Online Experiments'. *Journal of Behavioral and Experimental Finance* 17:22–27.

Pearson, Adam R., John F. Dovidio, and Samuel L. Gaertner. 2009. 'The Nature of Contemporary Prejudice: Insights from Aversive Racism'. *Social and Personality Psychology Compass* 3(3):314–38. doi: 10.1111/j.1751-9004.2009.00183.x.

Peer, Eyal, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. 'Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research'. *Journal of Experimental Social Psychology* 70:153–63.

Ridgeway, Cecilia L. 2000. 'The Formation of Status Beliefs Improving Status Construction Theory'. Pp. 77–102 in *Advances in Group Processes*. Vol. 17, *Advances in Group Processes*. Emerald Group Publishing Limited.

Ridgeway, Cecilia L. 2019. *Status*. New York: Russell Sage.

Ridgeway, Cecilia L., and Susan R. Fisk. 2012. 'Class Rules, Status Dynamics, and "Gateway" Interactions'. Pp. 131–51 in *Facing social class: How societal rank influences interaction*. New York, NY, US: Russell Sage Foundation.

Rivera, Lauren A., and András Tilcsik. 2016. 'Class Advantage, Commitment Penalty: The Gendered Effect of Social Class Signals in an Elite Labor Market'. *American Sociological Review* 81(6):1097–1131. doi: 10.1177/0003122416668154.

Roque-Hernández, Ramón Ventura, José Luis Díaz-Roldán, Adán López-Mendoza, and Rolando Salazar-Hernández. 2021. 'Instructor Presence, Interactive Tools, Student Engagement, and Satisfaction in Online Education during the COVID-19 Mexican Lockdown'. *Interactive Learning Environments* 0(0):1–14. doi: 10.1080/10494820.2021.1912112.

Ross, Lee. 2018. 'From the Fundamental Attribution Error to the Truly Fundamental Attribution Error and Beyond: My Research Journey'. *Perspectives on Psychological Science* 13(6):750–69. doi: 10.1177/1745691618769855.

Smallenbroek, Oscar, Florian R. Hertel, and Carlo Barone. 2021. ‘Measuring Class Hierarchies in Postindustrial Societies: A Criterion and Construct Validation of EGP and ESEC across 31 Countries’.

Snowberg, Erik, and Leeat Yariv. 2021. ‘Testing the Waters: Behavior across Participant Pools’. *American Economic Review* 111(2):687–719. doi: 10.1257/aer.20181065.

Spencer, Steven J., Christine Logel, and Paul G. Davies. 2016. ‘Stereotype Threat’. *Annual Review of Psychology* 67(1):415–37. doi: 10.1146/annurev-psych-073115-103235.

Steele, Claude M. 1998. ‘Stereotyping and Its Threat Are Real.’

Sullivan, Alice. 2001. ‘Cultural Capital and Educational Attainment’. *Sociology* 35(4):893–912. doi: 10.1017/S0038038501008938.

Tilly, Charles. 1998. *Durable Inequality*. Berkeley: University of California Press.

Tversky, Amos, and Daniel Kahneman. 1974. ‘Judgment under Uncertainty: Heuristics and Biases’. 185:10.

Vaisey, Stephen, and Omar Lizardo. 2016. ‘Cultural Fragmentation or Acquired Dispositions? A New Approach to Accounting for Patterns of Cultural Change’. *Socius* 2:2378023116669726. doi: 10.1177/2378023116669726.

Vuletich, Heidi A., and B. Keith Payne. 2019. ‘Stability and Change in Implicit Bias’. *Psychological Science* 30(6):854–62. doi: 10.1177/0956797619844270.

Wright, Erik Olin. 2015. *Understanding Class*. New York: Verso Books.

Yu, Zhonggen. 2022. ‘The Effect of Teacher Presence in Videos on Intrinsic Cognitive Loads and Academic Achievements’. *Innovations in Education and Teaching International* 59(5):574–85. doi: 10.1080/14703297.2021.1889394.

Chapter 2. Are there status beliefs about class?

With Klarita Gërxhani and Arnout van de Rijt

Abstract

This paper tests the existence of status beliefs in the case of social class. Social class has been largely avoided as a focus of status research because unlike other status variables – ethnicity, gender, motherhood – class is strongly correlated with performance on status-relevant tasks. Productivity indicators such as education and occupational experience are conceptually intertwined with social class. As a result, a preference for high-class individuals can plausibly be attributed also to statistical discrimination, even after accounting for observable performance indicators. Our innovation is to ask evaluators of higher and lower class workers to predict worker performance on a task using data on workers' past performance on that same task. We then study whether class discrimination tends toward zero as the number of data points on which past performance information is based increases. If instead discrimination stabilises at some non-zero level, this provides evidence in favour of a status mechanism. We embedded this design in an online experiment, studying also gender, a classic status variable, for comparison. A British sample of four hundred "evaluators" were hired to evaluate performance in two different tasks: a navigation task, where men should be expected to perform better; and a culture task, where higher class people should be expected to perform better. We find conclusive evidence in favour of status discrimination on social class, and little evidence for status discrimination on gender.

1 Introduction

This paper aims to address the first research question of the thesis, RQ1: Is there a widely-held belief that "class = competence" that biases evaluation of performance? In theoretical terms, this means asking if there are status beliefs about class.

Status theory is the dominant sociological framework for explaining discrimination (Pager and Shepherd 2008; Rivera 2020). The core contention of status theory is that discrimination operates through a cognitive bias, where individual members of certain historically advantaged groups are believed to be generally more competent than their historically disadvantaged counterparts. Over the past few decades, status theory has generated a broad range of insights and findings that have advanced our understanding of discrimination, particularly in the case of gender, but also for ethnicity and other characteristics (see Ridgeway (2019) for an overview). But in contrast with the abundant evidence for gender and ethnicity, there is virtually no empirical work that seeks to investigate whether there is status discrimination based on social class.

As status theorists have acknowledged, class presents a more difficult case for status theory than other characteristics (Ridgeway and Fisk 2012) .The fundamental problem of testing the existence of status discrimination in the case of social class is that an individual's social class position is very highly correlated with productivity. This means that class discrimination could be entirely explained by the rival theory of statistical discrimination. Statistical discrimination stipulates that discrimination arises because there is uncertainty about individual performance, and evaluators then use aggregate information on the productivity of the social group that the individual belongs to (Correll and Benard 2006b). Statistical discrimination does not postulate that people discriminate due to a cognitive bias – merely that they neutrally evaluate productivity relevant information. It is therefore plausible that controlling for enough productivity indicators may reduce statistical discrimination to negligible levels in the case of gender and ethnicity, meaning residual discrimination can be attributed to status. However, class position is conceptually interlocked with common indicators of productivity. Core determinants or correlates of class such as occupation and education carry highly productivity-relevant information making it difficult, if not logically impossible, to design or statistically adjust away the statistical discrimination.

The solution to disentangling status and statistical discrimination is to get rid of uncertainty concerning individual productivity by providing sufficient information (Correll and Benard 2006b). However, given the close connection between indicators of class position and productivity indicators, it is hard to say how much information is enough to eradicate uncertainty. The solution we propose here is to treat information as a continuous variable on a domain from zero towards infinity. Both theories of status and statistical discrimination make distinct predictions about the behaviour of discrimination as information increases to infinity and uncertainty decreases to zero. Statistical discrimination theory has discrimination approach zero; status theory predicts that discrimination will stabilise at a non-zero level.

We design a novel experiment to leverage these diverging predictions, focusing on the categories of social class *and* of gender; the latter is included as a benchmark as it is the most studied characteristic within both theories. In our design, there is strict consistency between the task the performers are evaluated on and the task they are evaluated for, and we use a single, quantitative information metric rather than a cluster of qualitatively different pieces of information as would be in resumé-based studies. This allows us to monotonically reduce uncertainty towards zero. The results show strong evidence for status-based discrimination in the case of social class, but little evidence for gender.

2 Discrimination theories and social class

According to status theory, there are widely shared "status beliefs" which attribute greater competence to higher status groups such as men and white people, and lower competence to lower status groups such as women and people of colour (Ridgeway 2019). Status beliefs entail that performance is not viewed neutrally, but through the lens of ascribed social characteristics.[3] Importantly, status beliefs are "inscribed in the brain" (Kalkhoff et al. 2020:26). They are the result of "distortions and bias" rather than neutral ratiocination, meaning that people have an ingrained tendency to see higher-status people as more competent and more worthy of esteem. While status theory allows that some degree of bias may be eroded by repeated demonstrations of competence, it proposes that the effects of

[3] Owing to its social psychological origins in small group theory (Berger, Cohen, and Zelditch Jr 1972), one of the original scope conditions of status theory was that status beliefs operate in environments where a group is working together towards a common goal. However, subsequent theoretical and empirical work extended status theory into individual task settings (cf Correll and Ridgeway [2006]).

discrimination “never really go away because the actor disadvantaged by the characteristic has to be otherwise better and more competent than others to be seen as equally worthy of status and influence in the context” (Ridgeway 2019:108–9).

Status theorists have theoretically discussed how social class could function as a status characteristic, while emphasizing that some modifications or extensions are required to make status theory work in this case (Ridgeway 2019; Ridgeway and Fisk 2012). The main reason for the latter is that class is seen as changeable, and hence an achieved characteristic rather than one assigned at birth like gender or ethnicity are. Gender and ethnicity, while socially constructed and malleable, are typically seen as fixed attributes which one is born with. Class is generally perceived as a contingent achievement: the translation of individual merit into a particular social position, with a higher class position supposedly indicating greater work ethic and superior competence (Markovits 2019; Ridgeway 2014).

The tendency to see class as a position won by competence and effort is reinforced by the close entanglement, both conceptual and statistical, between class and productivity indicators. In the case of social class, the association with putatively objective markers of productivity is much less arbitrary than it is for gender and ethnicity. For example, education – both in terms of quality of institution attended, grades and level attained – is highly correlated with social class. Yet, education is also a typical indicator of productivity, hence making it difficult to “net out” any tendency to discriminate over-and-above appraisals of a seemingly neutral indicator of competence (Rivera and Tilscik 2016). Even the use of cultural capital and other seemingly arbitrary markers of class distinction may reflect that such markers actually do convey information on otherwise unobservable indicators of productivity (Santos, Reis Neto, and Verwaal 2018).

It is thus likely that when investigating class as a status characteristic, status beliefs are confounded with statistical discrimination. According to the theory of statistical discrimination (Arrow 1971; Phelps 1972), discrimination can arise when there is uncertainty concerning individual performance. Evaluators make up for this uncertainty by utilizing information on the performance distribution of the social group an individual belongs to. For example, an employer will choose the candidate from the group with higher average productivity or, if they are risk-averse, with lower variance. This means that under conditions of uncertainty about true aptitude for the task at hand, an evaluator could infer from

productivity-relevant indicators closely correlated with class position, that a higher class individual would perform better at the task. In the absence of uncertainty, by contrast, if the evaluator were given convincing information that working class Jo is just as good as upper class Jim, they would have no statistical reason to prefer the latter to the former.

3 Disentangling the mechanisms behind class discrimination

What would count as "convincing information"? As Correll and Benard (2006b) have pointed out, the key to distinguishing between statistical and status mechanisms is providing sufficient information to eradicate uncertainty. The statistical discrimination mechanism is operable only when there is uncertainty over individual performance. On the other hand, even providing abundant information on individual performance is not be sufficient to erase status beliefs: "status theories assume that cultural beliefs distort cognition, even when information is perfect" (Correll, Benard, and Paik 2007:1302).

The problem of disentangling these mechanisms in the case of social class, is that class is so tightly interlinked with productivity indicators that we can never be sure we have provided enough information to eradicate uncertainty. Much of the foundational work in status theory consisted of studies following a fairly uniform "standardized setting" to establish the existence of status beliefs for various (Berger 2014; Correll and Ridgeway 2006; Wagner and Berger 1997). In order to avoid associations with known abilities, the standardized setting usually involved the participants being told they would be evaluated on a "newly discovered" skill (Berger 2014), in fact an arbitrary task without an objectively correct answer. The use of an arbitrary task is an entirely logical design choice, given that the experimenters wished to eliminate influences other than that of the partner on the research participant's choice. However it does allow for some degree of uncertainty, precisely because it is supposedly new and therefore the participant will have no information concerning their ability at the task. Because of the particularly close entanglement of class with conventional indicators of productivity, the presence of uncertainty in the design may be sufficient to activate statistical discrimination. For example, participants could reason that well-educated people are adept at responding to new challenges, such as the task involving the "newly discovered" skill. The standardized setting therefore would have limited efficacy as tool for differentiating status from statistical discrimination on class.

Other common designs for studying discrimination are also ineffective for disentangling the status and statistical mechanisms for class. Probably the most common empirical design in discrimination studies over the last few years has become the audit or correspondence study (Baert 2018; Neumark 2018). The basic format of this design involves sending fake resumés to actual employers, randomly varying some characteristic of the resumé to signal the social group the applicant belongs to (e.g. female, white, etc). A common strategy is to vary the applicant's name. Since all other relevant characteristics are kept the same – education, experience, etc – it is logical to attribute any variation in callback rates to status discrimination.

It is plausible, given the more arbitrary connection between, for example, gender and ethnicity on the one hand and conventional productivity indicators on the other, that "enough" information can be included on a resumé to reduce statistical discrimination to negligible levels (Neumark 2018). However, given the close *statistical* connection between class and conventional indicators of productivity, greater precision of control is in order. But the *conceptual* interwovenness of class and productivity renders control more difficult. The challenge is how to vary "only" arbitrary signals of social class while thoroughly controlling for objective differences in productivity.

One strategy is to look at discrimination based on *class background*. This could occur, for example, when individuals with the same educational background but different social origins apply for a job in an elite profession. Arguably having a working class background – as signalled by cultural capital or other such putatively arbitrary indicators – could be a basis for discrimination. Some correspondence studies have uncovered discrimination following this strategy (Jackson 2009; Rivera and Tilcsik 2016; Thomas 2018), yielding important evidence of discrimination based on class background. But can correspondence studies really rule out statistical discrimination, even when restricted to discrimination based on class origin rather than current class position? Almost certainly not: it is highly likely that there will still be some significant sources of uncertainty which will lead to statistical discrimination.

Firstly, unless class advantage is fully mediated by observable markers of educational and occupational achievement, individual from a higher class backgrounds will be possessed of other, unobservable, advantages over equally credentialled individuals from a lower class background. Hence, those from a higher class background will perform better on average.

Complete or near-complete mediation is plausible when the task at hand is tightly coupled to formal qualifications. Yet in labour market and analogous settings, the tasks that the candidate is being evaluated *on* are always different from the task they are evaluated *for*. For example, a resumé will give information on educational attainment, past occupational experience, and so on. However, the tasks which the applicant had to perform in school and university, and also in other jobs, will inevitably differ to some degree from the tasks they will have to perform in the new job. Even within the same industry, the exact nature of the work will differ from firm to firm. In a different task, a candidate may perform better or worse, hence information on past performance in dissimilar tasks will only be imperfectly indicative. This problem of "task inconsistency" generalizes to other settings such as political competition.

Secondly, uncertainty is likely to be a concave function of stated applicant characteristics – at a certain point, providing *more* bullet points will not further decrease uncertainty and perhaps even increase it. There are indeed many experiments which control for statistical discrimination by varying the degree of information given. This is typically done in audit or correspondence studies, meaning that the treatments consist of varying the number of items on a candidate's resumé (cf Auspurg, Hinz, and Sauer (2017) and Thijssen, Coenders, and Lancee (2021) for examples). For example, in one treatment there could be only five items, in another eight, or twelve. However, providing a dozen different bits of information may not decrease uncertainty much, since an evaluator is likely not paying attention to all of them. Research using eye-tracking software has shown that people do not pay (equal) attention to all the information given on a resumé (Lahey and Oxley 2018). The more information that is given, the more likely the evaluator is to be bored or overwhelmed, and hence to adopt some idiosyncratic heuristic to weight or ignore information.

Thirdly, even the most honest, informative and detailed resumés, scrutinized by the most scrupulous and patient eye, cannot provide information on every aspect of an individual's abilities. It is a generally agreed upon principle among scholars that it is difficult if not impossible to observe all relevant facets of ability (Belzil and Hansen 2002; Blackburn and Neumark 1992; Taber 2001). In the absence of sufficient individual information, the predictive power of group membership is substantive. Many studies looking at gender and ethnic gaps in performance show that, even when controlling for various performance-relevant variables, it is common that between 50% and 100% of the unadjusted gap in

performance remains (Azmat and Ferrer 2017; Battaglia et al. 2018; Gallen 2018; Mayer and Rathmann 2018; Reardon, Robinson-Cimpian, and Weathers 2014). Given that social class is much more tightly correlated with productivity indicators, it is highly likely to remain a reliable proxy of ability. In practice, the only way we could be sure of eliminating uncertainty is by providing perfect information on individual productivity. However, that is not possible.

These problems are all magnified if we want to study discrimination based on *current* social class and not just on social origin. Because in this case it is likely not possible to meaningfully control for differences in educational background, occupational experience and so on, since these differences are conceptually bound to social class itself. For example, correspondence and audit studies keep productivity-relevant information such as education and work experience constant and only vary arbitrary signals of group membership such as name. However, education and work experience are logically part of the distinction between higher and lower class groups, so cannot be kept constant. In the next section, we present a novel design to disentangle status and statistical mechanisms of discrimination despite the inevitability of uncertainty to uncover whether there is status discrimination based on *current* social class position.

4. Research design

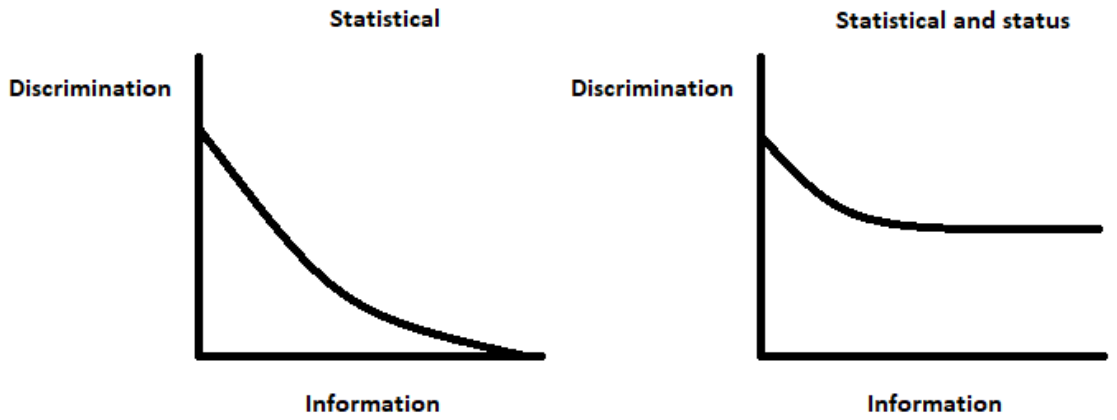
4.1. Overview of design

How can we disentangle status from statistical mechanisms when social class remains such a good proxy for productivity? Ideally one would control away most productivity-relevant variables correlated with status, so that status group membership has negligible informativity. But we have argued that while this may be a plausible strategy for gender and ethnicity, it is not for class. Fortunately, status and statistical theory do not just imply different predictions about discrimination when productivity information is excellent and uncertainty is thus minimal; they also imply different predictions about how discrimination evolves as we traverse the information continuum along which uncertainty decreases. Statistical discrimination would predict that, as we approach perfect information, discrimination should converge to zero. On the other hand, status theory would predict convergence to a non-zero value: there is a hard core of discrimination which is resilient even to sufficient counterevidence; at a certain point along the continuum, discrimination should stabilise at a non-trivial level.

Our proposed solution is thus to conceive information as a continuum, from zero to perfect information. While perfect information is not attainable, we can approach it along the continuum, getting arbitrarily closer as we provide more and more information. Hence, we disentangle the two mechanisms not by achieving perfect information, which is probably impossible, but by studying how the pattern of discrimination evolves.

Figure 2.1 illustrates in a simplified, stylized format the form of the relationship between information (measured on the x-axis) and discrimination (measured on the y-axis) that ought to prevail under the different circumstances. The graph on the lefthand side of the figure depicts a situation where there is only a statistical mechanism at play, where increasing information eventually reduces discrimination to zero. (Note that though the line is convex, it could also easily be concave, or straight. The point is it reaches zero.) The graph on the right-hand side depicts a situation where there is both a statistical and a status mechanism. In the right-hand side graph, at a certain point the curve flattens into a horizontal line parallel with, and above the x-axis, indicating that there is some residual level of discrimination which, due to the status mechanism, will never be reduced to zero no matter how much information is provided.

Figure 2.1. Stylized representation of the two scenarios



In order to study this pattern, we design and implement a novel experiment. The crucial and innovative features of this design are (i) a strict consistency between the task the performers are evaluated on and the task they are evaluated for; and (ii) a single quantitative measure of the level of information which is entirely distinct from the measure of performance. Feature

(i) deals with the problem of "task inconsistency" discussed above, where, as is necessarily the case in correspondence studies, subjects evaluate candidates based on their past performance in tasks which will not be the same as the task they are being evaluated for. Feature (ii) deals with the problem that providing more discrete bits of information, as is done in some correspondence studies, may eventually overwhelm and bore the reader and they may stop paying attention. With the design implemented here, we can be sure that uncertainty monotonically decreases as information increases. Our experiment was pre-registered[4] and carried out over two stages, which involved hiring two distinct samples.

Stage 1: We hired 240 "task performers" to complete two tasks: a male-favouring navigation task, and a higher class-favouring culture task.

Stage 2: We hired 401 "task evaluators", who evaluated the performance of the task performers in each task. They evaluated the performance of the task performers by reviewing profiles of the performers which contained information on the performers' social class, gender, and performance scores. Specifically, they evaluated two side-by-side profiles— we call these side-by-side profiles "profile pairs". It is from the evaluations of these pairs that we derive our dependent variable and generate the data for the main analysis (each row in the dataset corresponds to a single evaluation of a profile pair by a task evaluator). We varied the quantity of information available on the competence of the performers, thereby manipulating the degree of uncertainty.

4.2. Stage 1 — The task performers

On the 15th of December 2020, 240 British nationals were hired online through the Prolific[5] platform to carry out the first step of the study[6]. These participants were the "task performers". The task performers were paid a participation fee of £5.55, and could win a

[4] https://osf.io/46yrm Note that the pre-registration was only uploaded to the OSF platform after a period of time, due to a delay involved with importing the registration from the now-defunct EGAP registry. However, the registration confirms the original time when the pre-registration was uploaded to the EGAP registry, on the 14th of December 2020.

[5] Prolific is an online platform for hiring participants, based in Oxford. Unlike Mechanical Turk, it is specifically designed for the scientific community, and has been found to provide better quality participants (Eyal et al. 2021; Palan and Schitter 2018; Peer et al. 2017). URL: https://www.prolific.co/

[6] British nationals are an apt choice for a study on discrimination in the case of social class because it is the empirical setting for many influential studies on class from the classic (e.g. Marx (Marx 1965 [1867]) to the contemporary, e.g. Bukodi and Goldthorpe (2018).

reward of up to £4. The median earning was £6.55 and the median time taken to complete the experiment was 25 minutes. The sample was split into four equally sized sub-samples, depending on their social class and gender: high-class men, low-class men, high-class women, low-class women.

The task performers did two tasks: the navigation task and the culture task. The order of the tasks was randomized and balanced within each subgroup, so that half of the task performers did the navigation task first, and the other half did the culture task first.

Each task was designed to favour one group within the gender and social class categories. The navigation task was designed to favour men, and the culture task was designed to favour higher class people. More specifically, the tasks were designed so that there could reasonably be group-information based performance expectations as well as task-based status expectations.

The rationale for using the navigation task was existing empirical literature showing that men are expected to possess greater navigation skills than women, and that there are stereotypes that men are better in this regard (Hausmann et al. 2009; Rosenthal et al. 2012; Shurchkov 2012). Hence, we could plausibly expect both statistical discrimination and status-based discrimination to occur for this task, if it occurs at all.

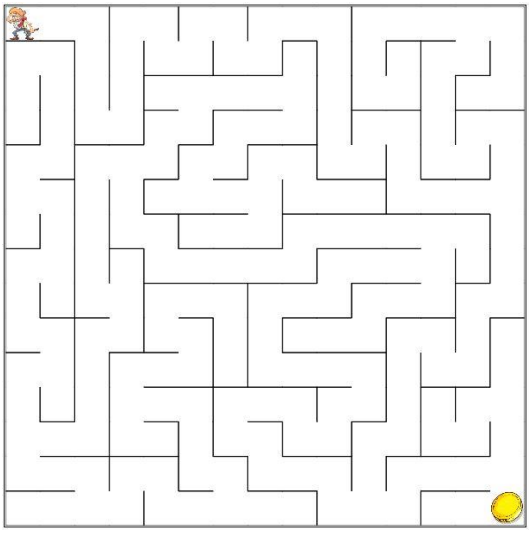
The culture task was designed to evoke the perception that individuals who possess greater cultural capital would perform better. We expected that higher class people would be perceived as having greater cultural capital; specifically, a familiarity with highbrow arts and culture. The argument that the familiarity of higher class people with highbrow culture leads to them being (mis)perceived as more competent is a familiar one (Bourdieu 1984). While the notion of what constitutes high(brow) cultural capital is rather slippery, recent work has tended to confirm that highbrow cultural capital is associated with higher class, and that higher class people benefit from this assumption (Friedman and Laurison 2020; Rivera and Tilcsik 2016; Thomas 2018).

Prior to the experiment, we ran a preliminary study to see if there would be performance expectations favouring men in the navigation task and higher class people in the culture task. Participants played a round of either task and afterwards were asked which group (men or women / higher or lower class) they thought would do better. They were given a bonus of £1.20 if they guessed correctly. 73% of participants thought men would outperform women

in the navigation task, and 79% of participants though higher class people would do better in the culture task. (More details on the preliminary study are given in Appendix A2.1.)

The specific type of navigation task we implemented was a two-dimensional maze task where the participant had to navigate a gold miner through a maze to a gold coin. Figure 2.2 shows a picture of a maze, with the miner and the gold coin in opposing corners. Participants were incentivized to guide the miner through the maze as fast as possible: the faster they completed the maze, the more money they received. Participants completed six rounds of this task, plus a practice round to familiarize themselves, with their score in each round based on how quickly they completed the task. Instructions for the task are given in Appendix A2.2.1.

Figure 2.2. The maze task.



In the culture task, participants in each round had to read short biographies of three notable cultural figures from British public life. They then had to answer five questions on these figures. The figures were selected from guests on the BBC radio show "Desert Island Discs" — a long-running radio show that interviews individuals who have made notable contributions to the world of arts and culture. We chose figures whose work would generally be classed as contributing to "highbrow" culture, such as writers of literary fiction, classical musicians, and Shakespearean actors. The text for their biographies was based on the individual's Wikipedia entries. As with the navigation task, the task performers completed six rounds of this task, plus a practice round to familiarize themselves. Their score was based on how fast they completed the task, with a time penalty applied for getting answers wrong. As

with the maze task, completing a round faster earned them more money. Instructions for the task, and a sample round, are given in Appendix A2.2.3 and A2.2.4.

To construct the profiles which we would then show to the task evaluators, we gathered information from the task performers on three variables: performance score (in each round of each task), gender, and social class. Their gender data was taken from demographic data which Prolific automatically supplies for all participants in a study. We had to gather the social class data ourselves, by placing questions within a survey given at the end of the task performer component. We wanted to show the task evaluators two indicators of the task performers' social class: their occupation, and their cultural capital.

We opted for two indicators since social class is a multidimensional concept (Bourdieu 1984; Markus and Fiske 2012; Savage 2015), consisting both of more "objective" measures such as occupation and income, and "subjective" measures such as culture and lifestyle. Since individuals can therefore occupy differential positions – having, for example, a high-class occupation, but retaining, from their upbringing perhaps, a working class cultural orientation (Friedman 2016) – we wished to ensure that the task performers presented to the evaluators unambiguously and consistently belonged to a higher or lower social class position. Hence, the task evaluators were always shown unambiguously high class or low class task performers – the players with the higher class occupation always had high cultural capital and the players with the lower class occupations always had lower cultural capital.

Modern class schemas, used pervasively in the social sciences, such as the Erikson-Goldthorpe-Portocarero schema (Erikson, Goldthorpe, and Portocarero 1979), and the more modern Oesch schema (Oesch 2006), are based on occupational position. Hence we opted to use occupation as one of our indicators. We elicited occupation with a survey question at the end of the task performer component. To cover the cultural and lifestyle dimension of social class, we asked the performers whether they would rather attend a football match or visit an art gallery. Preferring to watch a football match, a sport traditionally associated with the working class in Britain (Goulstone 2000), would be an indicator of relatively low cultural capita (Jackson 2009); whereas attending an art gallery is a classically highbrow, and higher class, cultural activity (Chan and Goldthorpe 2007; Silva 2008).

4.3. Stage 2 — The Task evaluators

We next recruited 401 British nationals to complete the evaluation assignment. These participants carried out this component of the task on the 28th of January, 2021. This sample was balanced so that half the participants would have a university education and half not, and that both educational groups would be gender-balanced.

The task evaluators were paid a participation fee of £3.50, and could win a reward of up to £4. The median earning was £5.22 and the median time taken to complete the experiment was 15 minutes. If we divide the participants' individual payments by the time each took to complete the experiment, then the median earnings per minute was £0.33, equivalent to an hourly "wage" of £19.8. This is a substantial payment, comparable to conventional labouratory experimental payments, and significantly above conventional payments for the Prolific platform.

Besides the desirability of providing a decent payoff for participants, we also assumed that providing a relatively large reward would improve the quality of responses. It is common in online experiments to incorporate some sort of attention or manipulation check to see if the participant is taking the task seriously, with the intention to exclude those who fail the check. However, this approach has been criticized. For starters, it can prime the participant to be more rationalistic than they would otherwise be (Hauser and Schwarz 2015); can introduce demographic bias if it involves removal of participants who fail checks (Vannette 2017); and lead to the introduction of "Hawthorne" or experimenter demand effects (Clifford and Jerit 2015). Hence, we opted for significant incentives to ensure quality responses, without introducing demographic bias, experimenter demand effects, or unintended pro-rationalistic treatment effects.

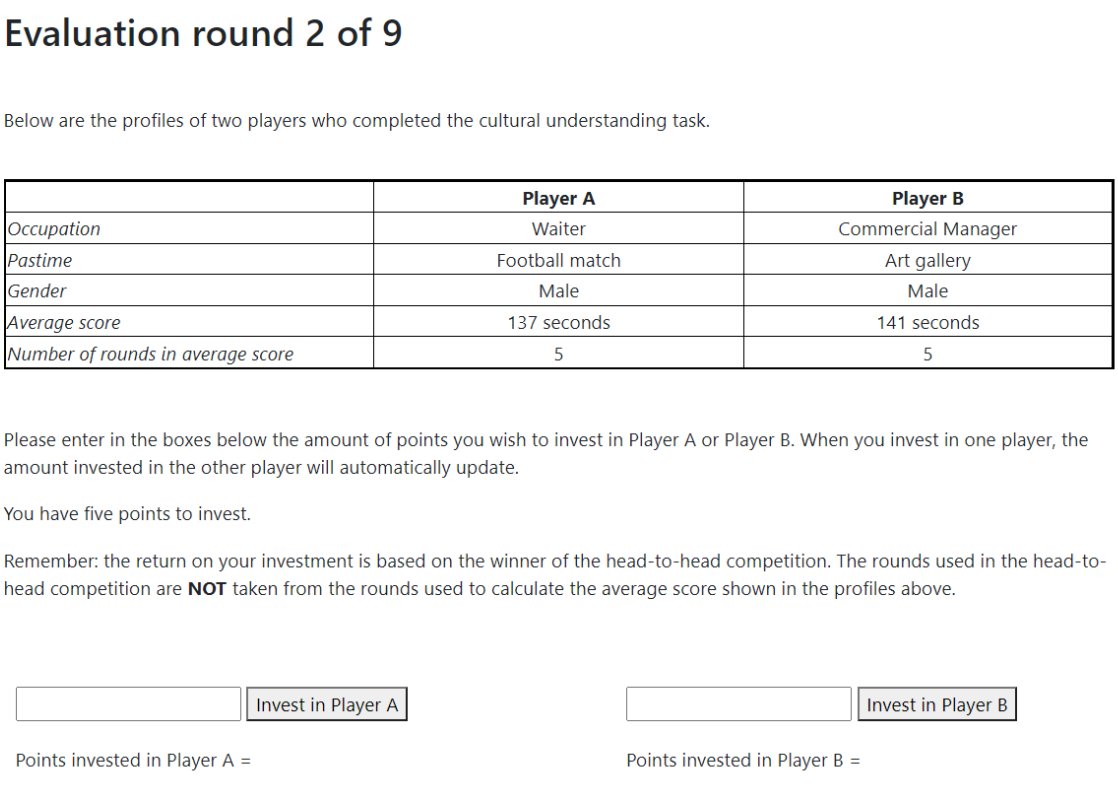
In the evaluation assignment, task evaluators reviewed the profiles of task performers for both the navigation task and the culture task. They reviewed the profiles in pairs, seeing two task performers' profiles side-by-side. These task performers were referred to as Player A and Player B. The allocation of performers to the Player A or Player B slot was randomly varied each time the profile pair was sampled, so as to avoid order effects.

The evaluators reviewed seven pairs of profiles for each task, so fourteen overall. Task evaluators also did two "dummy" evaluations per task, meaning that they actually reviewed eighteen profile pairs in total. The purpose of the dummies was to reduce social demand bias

(Rissing and Castilla 2014). For the navigation task, all profile pairs except the dummy pair were gender heterogeneous – i.e. one of the profiles belonged to a male performer and one to a female performer; while for the culture task all pairs were class heterogeneous (one higher class and one lower class). A consistent pattern of class or gender heterogeneous pairs may have disproportionately emphasized class or gender as variable of interest. Hence, the dummy pairs always showed gender- and class-homogeneous pairs. The dummy pairs are not part of the empirical analysis.

The task evaluators began their evaluation assignment with either the navigation task or the culture task (the order was randomized). First they read instructions on how to evaluate performance in the task — these instructions are reproduced in Appendix A2.3. They then did a practice round of the task, to experience it from the point of view of the performers. This practice round was exactly the same as that done by the task performers. Then they did a practice evaluation of a profile pair to familiarize themselves with the nature of their assignment.

Figure 2.3. Profile pair and evaluation decision.



Evaluation round 2 of 9

Below are the profiles of two players who completed the cultural understanding task.

	Player A	**Player B**
Occupation	Waiter	Commercial Manager
Pastime	Football match	Art gallery
Gender	Male	Male
Average score	137 seconds	141 seconds
Number of rounds in average score	5	5

Please enter in the boxes below the amount of points you wish to invest in Player A or Player B. When you invest in one player, the amount invested in the other player will automatically update.

You have five points to invest.

Remember: the return on your investment is based on the winner of the head-to-head competition. The rounds used in the head-to-head competition are **NOT** taken from the rounds used to calculate the average score shown in the profiles above.

Invest in Player A

Points invested in Player A =

Invest in Player B

Points invested in Player B =

Following the practice evaluation, the task evaluators evaluated nine profile pairs, reviewing one pair per evaluation. As mentioned above, we created the profiles based on the data taken from the task performers. The profiles contained information on the task performers' gender, social class, a performance metric, and an information metric. See Figure 2.3 for a screenshot of a profile pair shown to task evaluators. Gender was signalled simply by stating whether the performer was "male" or "female". Social class was signalled through the task performer's occupation and cultural capital. Again, social class was never "ambiguous" -- either the player was high class, and had a high class occupation and high cultural capital, or they were lower class, and had lower cultural capital.

The performance metric reported in the profiles consists of the average score of the task performer over a number of rounds of the task. The information metric reports the number of rounds used to create the average score. The number of rounds used to create the average score is a very important piece of information, because it corresponds to the quantity of information that the task evaluator receives about the competence of the task performer. The more rounds that are used to create the performance score for a given profile pair, the more information the task evaluator has about the underlying competence of the task performer.

Hence, the quantity of information is inversely related to the degree of uncertainty. A score from one round may be a mere product of chance: a performer could have done well or poorly by accident. A score averaged from five rounds, however, is likely to express a numerical quantity closer to the task performer's true competence – though not, of course, with perfect certainty. This is how we ensure that information is varied continuously in our experiment. Care was taken so that the instructions would highlight the importance to the evaluator of understanding the "number of rounds in average score" dimension. Information on the dimension is made to stand out from the rest of the text by use of bullet points and the evaluator is told that the pertinent information is "important" (see the instructions in Appendix A2.3). Subject behaviour varied significantly with information conditions (see section 4.2), confirming that evaluators took notice.

For the empirical analysis, we refer to the number of rounds averaged into the score as the "information condition". There are four information conditions, each corresponding to a different number of rounds. These conditions are:

Zero information condition: zero rounds

Low information condition: one round

Medium information condition: three rounds

High information condition: five rounds

For the evaluation of any given profile pair, the information condition of each performer in the pair will be the same. So if the profile of Player A shows a score averaged over three rounds (medium information condition) then the profile of their opponent in the pair, Player B, will also show a score averaged over three rounds.

4.4. Matching of profile pairs

The profile pairs were formed by matching task performers on certain criteria. The first criterion is that the performance score be similar. If the performance score were widely varying, the decision for the task evaluator would be trivial, and would be a waste of data and resources from our point of view. Profile pairs were matched so that their scores were within a tenth of a standard deviation. While some variation was necessary to make the task meaningful and minimize suspicion, we wished to avoid large deviations in score, which we suspected would dominate the evaluator's decision-making.

The second criterion is that the number of rounds used to estimate the performance score be the same — i.e. the information condition should be the same within the pair. It should be said that the rounds used to estimate the score are drawn randomly from the true playing history of the task performers. It is not *necessarily* the case that in, for example, the medium information condition, where the number of rounds = 3, that the first three rounds done by the task performer are used to estimate their performance score. *Any* three rounds can be chosen (without replacement). This is done so that the task evaluators do not make assumptions about learning effects. They are informed of this random selection.

The third criterion is that, for the navigation task, all profile pairs are class homogenous and gender heterogeneous; and, for the culture task, all profile pairs are class heterogeneous and gender homogeneous. Putting it more plainly, for all the profile pairs which were seen by the task evaluators when evaluating the navigation task, each task performer within the pair had the same social class as the other — higher class and higher class or lower class and lower class. But they had opposite genders. So, the task evaluator will have evaluated, say, a pair consisting of a male architect and a female lawyer, or a pair consisting of a male cashier and

a female bus driver. But they will *not* have evaluated a male cashier and a female lawyer, or a male architect and a female bus driver. Likewise, for the culture task, the task performers in each profile had the same gender, but opposing social class. This is because if we had crossed class *and* gender, we would then have introduced interaction effects which, while interesting, would have introduced complications outside the scope of our analysis, and which would have diminished the statistical power.

The fourth criterion concerns the creation of class homogeneous and class heterogeneous pairs. The formation of heterogeneous and homogeneous pairs is very straightforward in the case of gender, since all participants were either male or female. On the other hand, there was quite a diversity of occupations represented in the task performer sample and it is not necessarily the case, for example, that a given performer's occupation will be seen as equivalent to another performer's when matched in a "class homogeneous" pair. An analogous problem is that the gap between the lower and higher-class occupation may not be seen as equivalent for all potential "class heterogeneous" pairs. For example: task evaluators may see there as being a greater difference between a doctor and a cleaner than an engineer and a skilled machine tool operator.

We avoided this problem by making sure that pairs were also matched on a numerical scale, so that supposedly homogeneous couplings were *really* homogeneous and supposedly *heterogeneous* couplings were equivalently heterogeneous in all instances. For that, we chose to map the occupations onto Ganzeboom et al.'s (1992) International Socio-Economic Index of Occupational Status (ISEI). Every occupation reported on the sample was mapped to its rank in the index. For the evaluation of the navigation task, where profile pairs were class homogeneous, each occupation was within 3 points of each other on the ISEI scale. A similar approach was adopted for the culture task, where profile pairs were class heterogeneous, and in which the gap between the high and low class performer in the pair was restricted to between 35 and 45 points within pairs.

Once we had generated a list of all possible matches, we iteratively sampled from this list until we achieved a balanced universe of profile pairs. The universe was balanced so that, within each information condition, in 40% of pairs the performer from the advantaged group had a better performance score, in 40% of pairs where the performer from the disadvantaged group had a better score, and 20% of profile pairs had equal scores.

Table 2.1. Order of evaluation rounds shown by condition

Round	Condition
1	Zero information
2	Information (low, medium, or high)
3	Information (low, medium, or high)
4	Dummy
5	Information (low, medium, or high)
6	Information (low, medium, or high)
7	Dummy
8	Information (low, medium, or high)
9	Information (low, medium, or high)

As mentioned above, the evaluators saw seven profile pairs, plus two dummy profile pairs. They saw one pair from the zero information condition, which was always the first evaluation round they were shown[7].

For the subsequent eight evaluation rounds, they were shown either a dummy profile or a profile pair randomly sampled without replacement from the non-zero information conditions. There was no pre-determined order in which the evaluator saw a given information condition, nor did they see a fixed proportion from each condition. They always saw a dummy pair for their fourth and seventh evaluation. The order is given in Table 2.1.

4.5. Dependent variable

The dependent variable is the number of points invested by the task evaluator in the higher class (male) player in the culture (navigation) task. For every evaluation, the task evaluator was given a "budget" of 5 points. They had to distribute this budget of points, in any way that

[7] Evaluators were only shown one profile pair from the zero information condition for both theoretical and practical reasons. Theoretically, we wanted to capture the participants' prior expectations about the difference between genders / classes, before they were shown profile pairs with actual performance scores that would cause them to potentially "update" their expectations. Hence, if we showed more than one zero information condition, then we would have to show them in a row, which might strike the participant as suspicious or strange. And, since the score differences between the performers would be small in the other information conditions, we expected to observe a larger effect of gender / class in the control condition, hence less data would be required to reliably detect a statistically significant difference. Further, because task evaluators received no information on scores in the zero information condition, we assumed that evaluations by the same evaluator would be very similar, leading to little gain in statistical power.

they wished, to each of the task performers in the pair. They were informed that they could invest between 0 and 5 points into Player A, and that the remainder was automatically invested in Player B[8].

The task evaluator was incentivized to invest more points in whichever task performer they deemed better. This is because they would win back the points they invested in the "winning player"; the points were then converted into a monetary reward. In order to determine the winning player, a round was randomly drawn from the playing histories of Player A and Player B. The winning player was whoever performed better in the randomly drawn round. This randomly-drawn round *did not* come from any of the rounds which were averaged into the performance score shown to the task evaluators. This mechanism of determining the winning performer in the pair was explained to the evaluator.

We assume that the number of points invested in the performer reflects the evaluator's underlying expectation regarding the relative competence of that performer versus their paired opponent. If the task evaluator invests more points in A than B, that is because they expect that A is more likely to win them back their points.

4.6. Hypotheses

Both statistical discrimination and status discrimination would predict discrimination in the zero information condition. We therefore specify the following pre-registered[9] hypotheses:

Hypothesis 1: In the evaluation of the culture task, under the zero information condition, there will be discrimination against lower class people.

Hypothesis 2: In the evaluation of the navigation task, under the zero information condition, there will be discrimination against women.

[8] Because the total number of points to be invested is an odd number it was impossible to invest an equal number of points in both candidates. This boosted the design's capacity for revealing even slight biases, aiding statistical power.

[9] The pre-registration is available here: https://osf.io/46yrm. As already mentioned, this was registered via the EGAP (Evidence in Governance and Politics) platform, hosted on OSF. This platform has since migrated fully to OSF. Hence the pre-registration plan is registered on OSF at the date of migration – September 3, 2021. However, the form also notes the timestamp of the original submission to EGAP: 14th of December 2020.

In line with Hypotheses 1 and 2, we expect that the task evaluators, when in the zero-information condition, will invest more points in higher class versus lower class (in evaluating the culture task) and in men versus women (in evaluating the navigation task). This is a necessary condition for the experimental design to meet, since if there is no discrimination in this baseline condition, then there is not discrimination to be explained by either mechanism.

However, these hypotheses do not allow us to discern whether there is a status discrimination mechanism. In order to establish that, we must observe whether increasing information can reduce discrimination to zero. If there is some non-zero level of discrimination which cannot be further reduced, no matter how much information is provided, this is a good basis for claiming support for the status mechanism. If not, discrimination is entirely statistical. Hence[10]:

Hypothesis 3: In the culture task, discrimination will tend to zero as information increases.

Hypothesis 4: In the navigation task, discrimination will tend to zero as information increases.

5. Results

5.1 Main analysis

The following multilevel models were estimated, with random intercepts fitted to each evaluator (represented by the *j* subscript), to account for participant-level variability in average responses. The full dataset and replication package are available online[11].

Model 1 – culture task

$$PointsCenteredHighClass_{ij} = \beta_{0j} + \beta_1 LowInfo_{ij} + \beta_2 MediumInfo_{ij} + \beta_3 HighInfo_{ij} + \beta_4 ScoreDifference + u_j + \varepsilon_{ij}$$

[10] In our pre-registration, the hypotheses were erroneously stated in a manner that is not strictly implied by status theory: Hypothesis 3: In the culture task, under the high information condition, and controlling for performance, there will be no discrimination against lower class people; Hypothesis 4: In the navigation task, under the high information condition, and controlling for performance, there will be no discrimination against women. Regardless of whether we use the original formulation or the one used here, the test results and substantive conclusions are the same.

[11] Full dataset and replication package available here: https://osf.io/dexjy/.

Model 2 – navigation task

$$PointsCenteredMale_{ij} = \beta_{0j} + \beta_1 LowInfo_{ij} + \beta_2 MediumInfo_{ij} + \beta_3 HighInfo_{ij} + \beta_4 ScoreDifference_{ij} + u_j + \varepsilon_{ij}$$

The dependent variable is the number of points invested in either the higher class player in the profile pair (culture task) or the male player (navigation task). The dependent variable is centred around its mean of 2.5 points. A significant deviation from the centred mean is interpreted as a significant difference in points invested between higher and lower class, or male and female. The independent variables of interest are the categorical information variables, where the reference variable is zero information. Treating information as a categorical variable models information in the most agnostic way, not assuming any formalism for the shape of the relationship. The estimates are adjusted for the difference in average performance score between the higher class (male) player and the lower class (female) player in the profile pair.

The estimated coefficients from Models 1 and 2 are presented in Table 2.2, and the average marginal effects of the information conditions are graphed in Figures 2.4 and 2.5. Models 1 and 2 were designed to test hypotheses 1 and 2. As can be seen in the table and figures, there is, in the zero information condition, discrimination against women in the navigation task and against lower class people in the culture task. The evidence supports hypotheses 1 and 2 and thus establishes that there is some baseline bias in expected performance between genders and classes.

Hypotheses 3 and 4 stated that discrimination would tend towards zero as information increases, consistent with a pure statistical mechanism, leaving no room for status discrimination. We can already see from Figure 2.5 that discrimination reaches zero in the navigation task – hence hypothesis 4 is confirmed. There does not appear to be status-based discrimination against gender. Studying Figure 2.4, we see that going from no information to low information reduces bias. However it also appears that there is no further effect of going from low to medium, and from medium to high. Hence bias appears to stabilise which seems to be evidence *against* hypothesis 3, and in favour of the existence of status discrimination in the case of social class. Rather than eyeballing a graph however, we test this result for the culture task more formally by, using both a Wald test and a goodness of fit test against a simpler model The aim of the tests is to see whether there is any difference in effect between

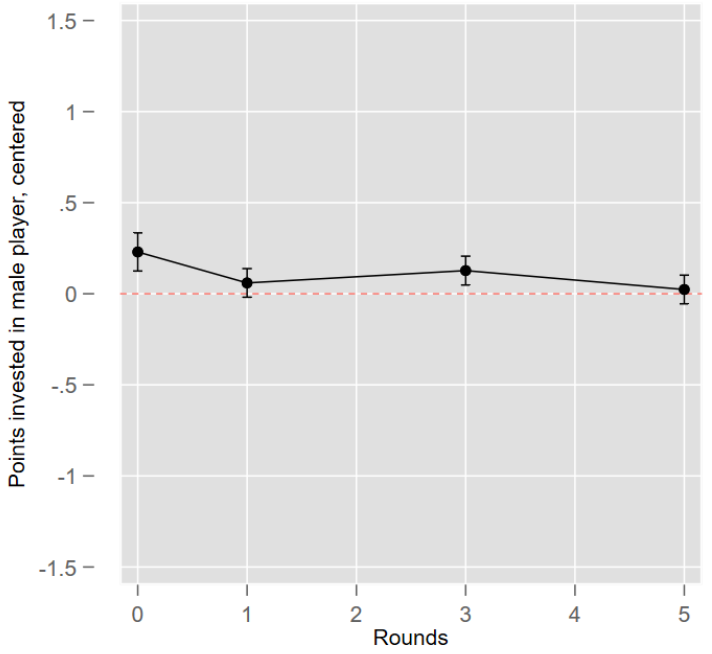
the low, medium, and high information conditions. If not, then providing more than "low" information will not reduce discrimination, and residual discrimination can be attributed to status discrimination, if there is any residual discrimination.

Table 2.2. Multilevel estimates for Models 1 and 2

	(M1) Culture task	(M2) Navigation task
Low Info	-0.70***	-0.17**
	(0.06)	(0.06)
Medium Info	-0.69***	-0.10
	(0.06)	(0.06)
High Info	-0.63***	-0.21***
	(0.06)	(0.06)
Score Difference	-0.22***	-0.48***
	(0.01)	(0.01)
Constant	1.09***	0.23***
	(0.05)	(0.05)
Participant-level error (std dev)	0.51	0.34
	(0.03)	(0.03)
Observation-level error (std dev)	0.98	1.01
	(0.01)	(0.01)
Observations	2,807	2,807
Number of participants	401	401

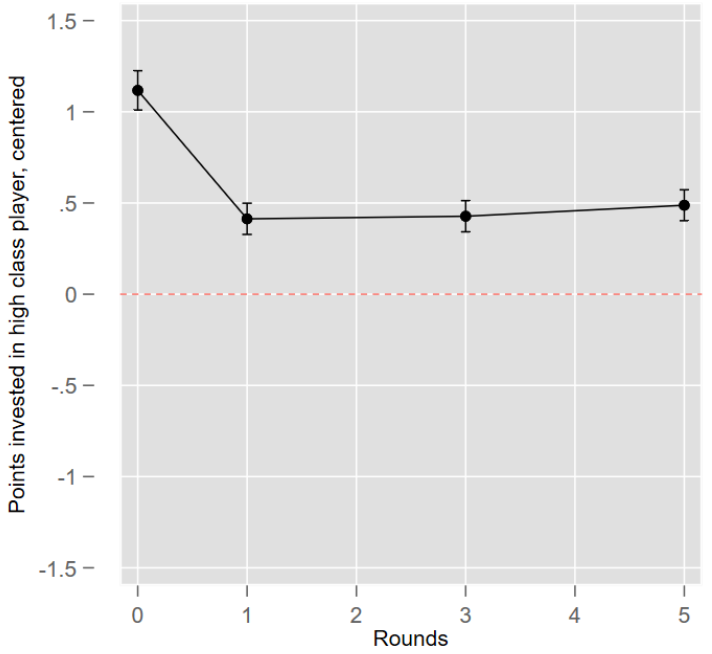
Standard errors in parentheses
Two-sided p-values: *** p<0.001, ** p<0.01, * p<0.05

Figure 2.5. Average marginal effects for the navigation task



Zero info condition = 0 rounds. Low info condition = 1 round.
Med info condition = 3 rounds. High info condition = 5 rounds.

Figure 2.4. Average marginal effects for the culture task



Zero info condition = 0 rounds. Low info condition = 1 round.
Med info condition = 3 rounds. High info condition = 5 rounds.

For the Wald test, we test for both model 1 and model 2 the null hypothesis that $\beta_1 = \beta_2 = \beta_3$. In other words, we test if there is a statistically significant difference between the

coefficients for the low, medium, and high information conditions. For the model fit comparison, we use a likelihood ratio test to compare model 1 to a simpler model where the categorical information variables LowInfo, MediumInfo, and HighInfo are replaced by a single categorical variable AnyInfo, as in model 3 below. AnyInfo is a binary variable that indicates whether any information on performance is given at all. It takes a value of 0 if the observation belongs to the zero information condition and 1 otherwise. The LR test tests whether the model that specifies individual parameters for each information condition fits the data better than a model that merely provides a binary indicator of whether information was given at all. If the first model doesn't perform any better, then this is indicative that there was no difference between the low, medium, and high information conditions – in other words, that once "low" information was given, providing more information did not further reduce discrimination.

Model 3 – culture task

$$PointsCenteredHighClass_{ij} = \beta_{0j} + \beta_1 AnyInfo_{ij} + u_j + \varepsilon_{ij}$$

The p-values for the Wald tests and the likelihood ratio test for the culture task (models 1 versus 3) is presented in Table 2.3. As can be seen the p-values are large (and identical to the third decimal place) and the null of no difference between coefficients (Wald test), and no difference between models (LR test) is not rejected. Providing more information does not reduce discrimination and residual discrimination can be attributed to status discrimination. Discrimination does not tend to zero in the culture task, hence hypothesis 3 is rejected. Examination of the task performer data shows that the task evaluators bias in the culture task cost them money. As we show in Appendix A2.5, while there is a gap in expected performance between lower and higher class performers in the low information condition, this gap shrinks to zero in the high information condition. Hence, the evaluators had no *statistical* reason to discriminate in this condition.

Table 2.3. Wald test and likelihood ratio tests for effect of information in the culture and navigation tasks.

Wald test p-value	0.294
Likelihood Ratio test p-value	0.294

5.2 Heterogeneities by even and uneven scores

In the design section we mentioned that score differences within profile pairs were deliberately restricted to very tight margins – within a tenth of a standard deviation. It was expected that small differentials would not meaningfully affect the task evaluators' decisions. However, inspection of the data showed that this assumption was not borne out. Task evaluators were fairly sensitive to variations in score difference between task performers within pairs. As such, we present here a brief analysis of discrimination for two subsamples: i) "even-scores" – all profile pairs where the task performers had the same score; ii) "uneven-scores" – all profile pairs where the task performers had differing scores. The even-scores make up 20% of the profile pairs when the zero-information condition is excluded.

The split analysis is theoretically motivated. Double standards theory (Foschi 2000), which has been integrated into the broader status theory paradigm (Correll and Ridgeway 2006), predicts that lower status individuals who perform well are judged more harshly than higher status individuals who perform well. People are more willing to accept that a high status person has excelled compared to a low status person. Hence, if double standards are at play, the task evaluators should invest more in the higher status task performer when the higher status task performer has a better score compared to their lower-status counterpart than they invest in the lower status task performer when the lower status task performer has the better score.

For the even-score subsample, we replicate the analysis of models 1 and 2, but restrict the sample to only profile pairs where the task performers had the same score. We also exclude the zero-information condition since these profiles by definition did not have performance scores. The regression output is omitted here and placed in Appendix A2.4.1. The marginal effects are presented here in Figure 2.6 and 2.7 for the culture and navigation tasks respectively. As we see in Figure 2.6, while there is a slight downward curve it is clearly levelling off and for 5 rounds of information remains substantively above the red dashed line marking the "no-discrimination" level. Even when restricting to only strictly even scores, there is still substantive status discrimination for the culture task. Turning to Figure 2.7, we see a straight line lying exactly on the dashed red line. When restricting to even scores, there is no discrimination at all in the navigation task from 1 round of information onwards.

Figure 2.6. Average marginal effects for the culture task, even-score subsample

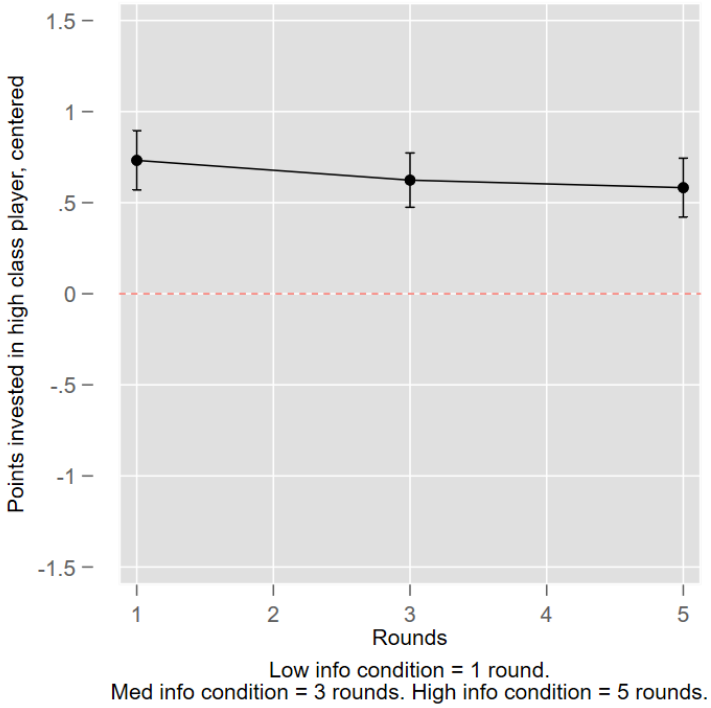
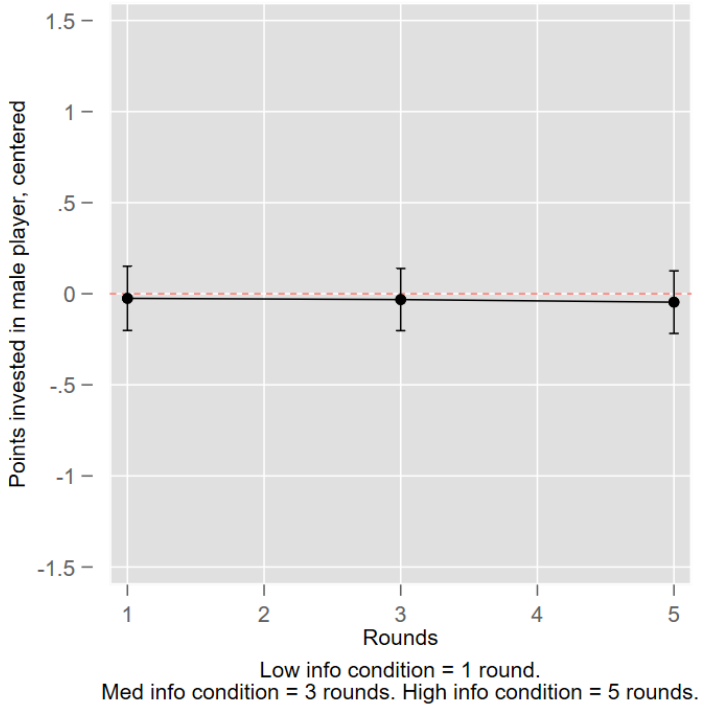


Figure 2.7. Average marginal effects for the navigation task, even-score subsample



We now turn to the "uneven-score" subsample. Here, the analysis is a little different by necessity. To conduct this sub-analysis, we create a new dependent variable,

PointsCentredBetterScorer, which is the number of points invested in the task performer with the better score in the profile pair, centred around the midpoint of the points scale (2.5). For example, if the task evaluator invests 4 points in Player A, and Player A is the better scorer in the profile pair, then the value of PointsCentredBetterScorer is 4 – 2.5 = 1.5. We construct two more variables, one for each task. BetterScoringClass takes a value of 0 if the better-scoring player is lower class, and a value of 1 otherwise. BetterScoringGender takes a value of 0 if the better-scoring player is female, and a value of 1 otherwise. If there is discrimination for either task then BetterScoringClass or BetterScoringGender will be positive and statistically significant. In our models, we interact BetterScoringClass/Gender with the information condition to test if there is variation in discrimination over information conditions. We estimate the following models:

Model 4 – culture task

$$PointsCentredBetterScorer_{ij} = \beta_{0j} + \beta_1 BetterScoringClass_{ij} + \beta_2 MediumInfo_{ij} + \beta_3 HighInfo_{ij} + \beta_4 MediumInfo_{ij} * BetterScoringClass_{ij} + \beta_5 HighInfo_{ij} * BetterScoringClass_{ij} + \beta_6 ScoreDifference_{ij} + \varepsilon_{ij}$$

Model 5 – navigation task

$$PointsCentredBetterScorer_{ij} = \beta_{0j} + \beta_1 BetterScoringGender_{ij} + \beta_2 MediumInfo_{ij} + \beta_3 HighInfo_{ij} + \beta_4 MediumInfo_{ij} * BetterScoringGender_{ij} + \beta_5 HighInfo_{ij} * BetterScoringGender_{ij} + \beta_6 ScoreDifference_{ij} + \varepsilon_{ij}$$

For these models, ScoreDifference equals the average score of the better scorer minus the average score of the worse scorer. The models are multilevel, with random intercepts fitted to participants. The regression output is given in Appendix A2.4.2. The average marginal effects are graphed in Figures 2.8 and 2.9.

Figure 2.8. Average marginal effects for the culture task, uneven-score subsample

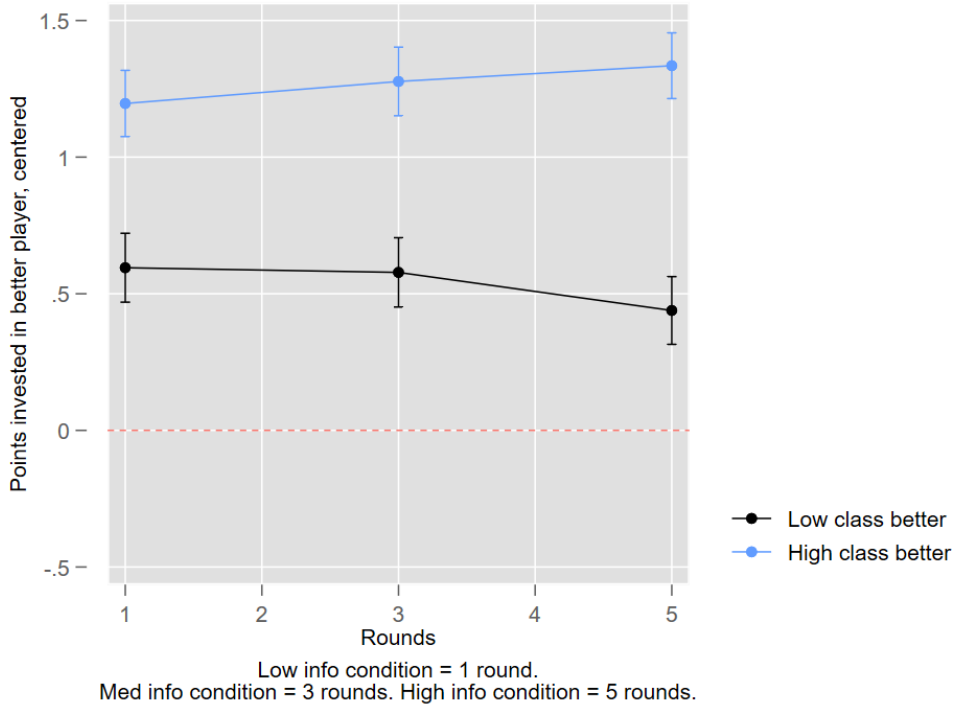
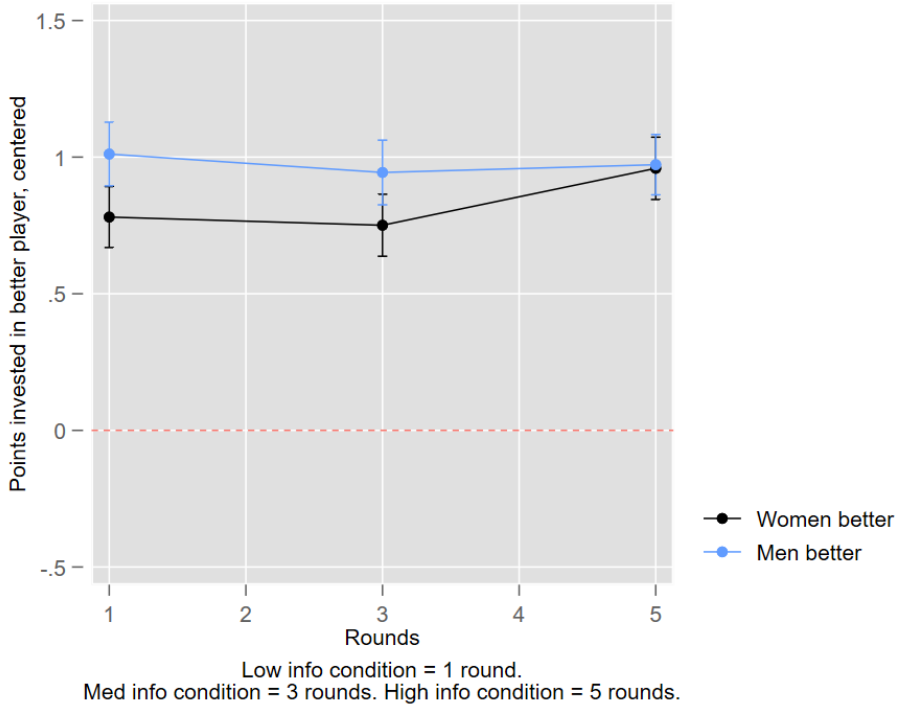


Figure 2.9. Average marginal effects for the navigation task, uneven-score subsample



The figures show diverging patterns that reinforce the earlier conclusions drawn with respect to hypotheses 3 and 4. In Figure 2.8, discrimination by class does not tend toward zero, again

rejecting hypothesis 3. Discrimination seems, if anything, to increase as more information is given. The evidence supports status discrimination on class. Interestingly, Figure 2.9 reveals that when women have a better record than men in the low information condition they get invested in less than what men get invested in when they have a better record than their female counterpart in the same condition. Nonetheless, this gender difference does tend to zero with increased information. The difference is already a bit smaller in the medium information condition and is gone entirely in the high information condition. The evidence suggests that because task evaluators do not have a detectible status bias in the case of gender for this specific task, statistical discrimination decreases as information increases and task evaluators increasingly favour whichever player has the better score.

6. Discussion and conclusion

The existence of a status mechanism in the case of social class has received strong support from the findings presented here. To put it in terms of RQ1: there are widely-held beliefs that "class = competence", which lead to biased assessment of performance. This result is notable at least in part because social class has not been a prominent characteristic of analysis for status theory, or wider work in the sociology of discrimination. The evidence presented here strongly suggests that many decisions to favour individuals on the basis of supposedly objective measures of competence actually serve to obscure a deep bias against lower class people and in favour of higher class people.

In order to disentangle the status and statistical mechanisms in the case of social class we implemented an original design which could also serve to untie this empirical knot for other ascribed characteristics. Because social class is a novel object of empirical research within the status theory paradigm, we also opted to investigate gender, the most-studied characteristic in the paradigm. Our study does not provide strong evidence for the existence of a status mechanism in the case of gender. But caution should be exercised here in concluding that a gender-based status mechanism does not exist at all. While gender discrimination in the past has been based on false beliefs about inherent ability, recent research has emphasized that status-based gender discrimination runs through beliefs concerning commitment or effort instead (Correll and Benard 2006b). That is to say, contemporary status beliefs about gender may not stipulate a distinction in inherent ability or

competence between genders, but may stipulate an inherent difference in effort and commitment to tasks.

For example, more recent work in the status field, as well as the broader discrimination literature has stressed the role of a "motherhood penalty" (Correll et al. 2007; England 2010; Gough and Noonan 2013) which derives from the belief or stereotype that women invest relatively more of their effort and energy in family life when compared to men, and hence relatively less in their work; it is this supposed lower commitment, rather than lower "innate" ability that drags down their expected productivity in the eyes of employers. Our experimental design was more apt for capturing ability rather than commitment – the participants were not asked to commit to working sixty hour weeks, but to solve puzzles – hence it would not have captured the status beliefs about commitment which may be the predominant basis of status discrimination in many contemporary societies.

More speculatively, it may help to contextualize our findings in terms of changes in inequality and attitudes regarding class and gender over the last few decades. Status beliefs ultimately derive from group-level differences in resources (Ridgeway 2019). Recent decades have shown diverging trends in resource gaps between men and women on the one hand, and higher and lower class on the other. While there are still important gaps, particularly at the top, women's participation in the labour market has increased substantially in recent decades (Fernandez, Castilla, and Moore 2000; Thévenon 2013). Women now outperform men academically in most western countries (Clancy and O'Sullivan 2020; Vincent-Lancrin 2008), contributing to a historic reduction in the gender wage gape (Goldin 2014; Goldin, Katz, and Kuziemko 2006). As the inequalities between genders have lessened, attitudes have begun to catch up, with gender egalitarianism becoming increasingly the norm (Knight and Brinton 2017; Scarborough, Sin, and Risman 2019). Belief in the equality of competence between genders has increased very substantially over the last few decades, at least in the USA (Eagly et al. 2020a), which suggests that status beliefs about gender may have weakened considerably.

The corresponding trends for social class display an almost mirror image. Whereas material inequalities between men and women have lessened in the past few decades, they have increased between classes, as indexed by rising income inequality (Goda and García 2016; McCall and Percheski 2010; Stockhammer 2013; Thewissen et al. 2018). Rising inequality

has been accompanied by, and may contribute to, falling social (Durlauf and Seshadri 2018; Hufe et al. 2022), partly because disparities in financial resources between classes drive disparities in educational outcomes (Jerrim and Macmillan 2015; Rauh 2017).

In short, individual ability may contribute less to economic success now than it has in previous decades. But despite the objective conditions, there has been a generalized increase in the belief that societies are broadly meritocratic, whereby success is determined by individual talent and drive, and this belief is strongest in the most unequal societies, just where we would expect the strongest structural barriers to mobility (Mijs 2021). Expressing the view that society is meritocratic is equivalent to expressing the view that higher class people are broadly more competent than lower class people, semantically equivalent to a status belief about class. It is possible that the belief that society is a meritocracy is "really" a status belief – this would explain why the belief is most prevalent in the countries which are most unequal and, as research suggests (Durlauf and Seshadri 2018; Hufe et al. 2022), most socially immobile. While this reasoning is necessarily speculative. Chapter 4 develops this line of thought in a more substantive way.

This paper has contributed to the literature on discrimination and status beliefs by demonstrating that there is cognitive bias against lower class people. While status theorists had discussed class-based status beliefs on a theoretical level (Ridgeway and Fisk 2012), establishing them empirically was acknowledged to be much less straightforward than for other characteristics such as gender and ethnicity. The method we present here allowed us to overcome the threat of confounding with statistical discrimination, and can potentially be used in other cases where an ascribed characteristic is highly correlated with indicators of productivity. Although the paper establishes the existence of a cognitive bias in the case of social class, we do not know exactly which realworld processes of inequality the bias may contribute. Given the strength of the effect, it is possible that status beliefs about class contribute substantially to inequality between social classes in diverse fields such as the education system, the labour market, and political representation. Future research can investigate the existence of status beliefs in these, and other, contexts.

Another limitation of the paper is that the experiment signalled social class in a fairly unambiguous way, through both listing an individual's occupation and a culture preference signalling their high or low cultural capital. However, such information may not be

immediately received or perceived in most contexts. If one or both pieces of information are not given, will other signals of an individual's class belonging – which may be ambiguous or not clearly perceptible – be sufficient to activate status beliefs? Future research could investigate whether status beliefs about class are activated when information on individual class position is less explicit.

Finally, the culture task for which discrimination was found against working class people was designed to evoke stereotypes about class because its subject matter pertained to (highbrow) cultural capital. It would be interesting to see if the status beliefs about class remain as strong when it comes to other tasks involving intellectual abilities, but which do not so obviously evoke class stereotypes.

Appendix A2.1 Preliminary study

In this preliminary study, 80 participants did the navigation task; and another 81 participants did the culture task. For each task, the 80 participants were equally divided into the gender/class subgroups. The participants played two rounds of the navigation task and culture task, respectively. The first was a "practice round", and the second was a "reward round". Performance in the latter was incentivised. In the navigation task, participants could win up to £0.30 depending on how quick they were. In the culture task, participants could likewise win up to £0.30, depending on how quickly *and* how accurately they answered the questions. They were also paid £1.25 and £1.50 for the navigation task and culture task, respectively.

Having completed the task, the participants in the navigation task were asked if they thought male or female participants would perform better in the navigation task. Participants in the culture task were asked if they expected higher or lower class people to perform better in the culture task. Note that we did not give the option for participants to select "neither". They had to choose either men or women, or higher class or lower class. The participants were incentivised to respond truthfully: if they correctly predicted which group would do better, they would get a reward of £1.20.

Although we intended that this study be a between-subjects design, where no-one who had done the navigation task would also do the culture task, due to an error by the researcher who was administering the experiment, almost half of the 160 participants did both. However, the

results are very similar whether or not we exclude data either from participants who were doing their second task.

Table A2.1 shows the results depending on different criteria for including participants: column 1 includes data from all participants, including those who did both tasks; column 2 includes data from all participants who were doing one of the tasks for the first time. The columns show the number (and percentage) of participants who favoured men and higher class, respectively. The table also includes the sample size (N) of those who meet the inclusion criterion for that column, as well as p-values for a one proportion test with the null hypothesis that the true proportion of participants who favour men / higher-class is 0.5. As can be seen, the results are substantively the same whether the full or restricted samples are used. In the article, we report the data from column 2: from participants who were doing their respective task for the first time, meaning that the reported data is between-subjects.

Table A2.1. Proportion favouring men / higher class.

	All participants	N	P-value	Only first-time participants	N	P-value
Number favouring men (percent)	56 (71%)	79	0.00	33 (73%)	45	0.00
Number favouring higher class (percent)	65 (80%)	81	0.00	61 (79%)	77	0.00

Appendix A2.2 Instructions for the task performers

A2.2.1 Instructions for the navigation task

There are six rounds in this task, plus a practice round. Each round contains a maze puzzle which you must complete.

The maze puzzle is a randomly generated maze, such as the one shown below. Your task is to direct the gold miner  to pass through the maze and reach the destination marked by the gold coin  . You can use the arrow keys on your keyboard to move the gold miner in all four directions.

The goal is to complete the maze as quickly as possible.

Reward

Depending on your performance, you can earn a reward of up to £1..

After you have completed all six rounds and finished this activity, one of those rounds will be randomly chosen as your "reward round".

Your performance in this randomly-chosen reward round will determine your total reward.

Performance is determined by how fast you can complete the round. The round is completed by getting to the end of the maze.

As soon as the page with the maze puzzle loads, a timer will begin. This timer determines your reward.

If you complete the reward round within 20 seconds, you will earn the full £1.

If you take longer than 20 seconds, you will earn less. Every extra second will reduce your earnings by 4 pence. If you take longer than 45 seconds to complete the round, your reward will be reduced to zero.

Please note that your reward will not go below zero, even if you take longer than 45 seconds.

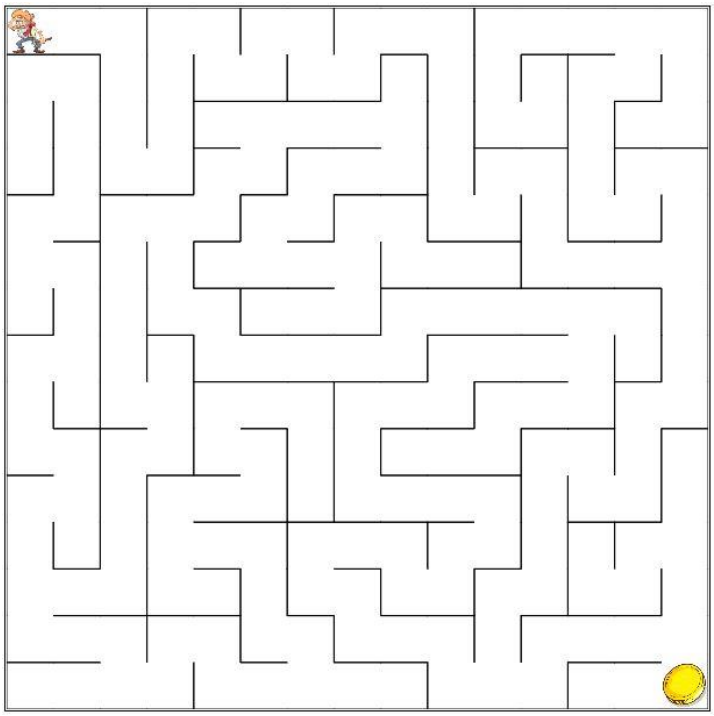
Practice round

Before doing the six rounds, you will do a practice round of the maze puzzle. The practice round cannot be chosen as your reward round.

A2.2.2 Sample round from navigation task

Figure A2.1 displays a screenshot from the maze puzzle. Participants had to guide a player character – the "miner" character shown in the top left – through the maze to the gold coin in the bottom right.

Figure A2.1. Maze puzzle



A2.2.3 Instructions for the culture task

There are six rounds in this task, plus a practice round. In each round, you will be asked to read short biographies of three notable persons. Each person has made an important contribution to the world of arts and culture.

You will then have to answer five questions about these persons. You must answer all questions.

When you have answered the questions, press the "Submit answers" button at the bottom of the page to submit your answers and continue to the next page.

Reward

Depending on your performance, you can earn a reward of up to £3.

After you have completed all six rounds and finished this activity, one of those rounds will be randomly chosen as your “reward round”.

Your score in this randomly-chosen reward round will determine your total reward.

The score is calculated as follows:

As soon as the round loads, a timer begins.

The timer stops when you press the “Submit answers” button.

If you answer all the questions correctly within 60 seconds, then you will win the full £3. If you take longer than this, you will start to lose money. After 60 seconds, your reward will be reduced by 4 pence for every extra second that you take to answer the questions.

You will be penalised for wrong answers. Every wrong answer in the reward round will add 70 seconds to your score.

Lower scores represent quicker and/or more accurate answers; higher scores represent slower and/or less accurate answers. A score under 60 seconds will earn you the full £3. A score over 135 seconds will earn you zero.

Please note that your reward will not go below zero, even if your score is over 135 seconds.

Practice round

Before the reward round, you will have the chance to do a practice round in order to get yourself familiar with the task. This round is not eligible to be randomly chosen as your reward round.

A2.2.4 Sample round from culture task

The text below reproduces a sample of one of the six rounds from the culture task.

Heidi Thomas is an English screenwriter and playwright. Thomas gained national attention when her play, *Shamrocks And Crocodiles*, won the John Whiting Award in 1985. The drama concerns a pair of siblings from Dublin who investigate their Irish father’s shady past, after his sudden death. The play was praised for its lyrical richness. Following this success, her play *Indigo* was performed by the Royal Shakespeare Company in their 1987—1988 season. Thomas has also authored many screenplays, including the feature film *I Capture the Castle*, and the BBC screen adaptation of Flaubert’s novel *Madame Bovary*. Thomas’ adaptation of Ibsen’s *The Lady and the Sea* was presented in London and at the National Theatre of Norway in Oslo.

David Edgar is a British playwright and writer who has had more than sixty of his plays published and performed on stage, radio and television around the world. Edgar began writing plays from the age of five, to be performed in his family’s garden shed. Edgar first gained attention in the early 1970s with a series of plays that mixed formal innovation with

sharp political commentary. His long association with the Royal Shakespeare Company resulted in his highly successful adaptation of Dickens' novel *Nicholas Nickleby.* Among his more recent plays is *The Prisoner's Dilemma,* an exploration, set in Eastern Europe, of the game theory problem of the same name. He regularly contributes to The Guardian and The London Review of Books.

Ronald Harwood was a South African-British playwright and screenwriter. After training for the stage at the Royal Academy of Dramatic Art, he joined the Shakespeare Company of Donald Wolfit. From 1953 to 1958, Harwood was Wolfit's personal dresser, an experience that inspired his play *The Dresser.* The Dresser was subsequently adapted into two films: a 1983 version with Albert Finney and Tom Courtenay; and a 2015 version starring Anthony Hopkins and Ian McKellen. Harwood won an Oscar for his work on the script of *The Pianist*. The film, winner of the main prize at the Cannes film festival, tells the story of how the Polish-Jewish pianist and composer Władysław Szpilman survived the Holocaust. Harwood was nominated for another Oscar for his adaptation of Jean-Dominque Bauby's memoir *The Diving Bell and the Butterfly*. Harwood died in September 2020.

Questions

1 Who won a major award for their script about a Polish Jew who survives the Holocaust?

2 Who wrote a television adaptation of a 19th century French novel?

3 Who won the John Whiting Award

4 Whose adaptation was performed in Oslo?

5 Whose adaptation was performed by the Royal Shakespeare Company?

Appendix A2.3 Instructions for task evaluators

Below are the instructions given to task evaluators for evaluating the navigation task. They are essentially identical to the instructions given for evaluating the culture task, which are admitted due to space constraints.

Now you will evaluate the performance of players in a navigation task, with a chance to win up to £2.

The navigation task is a maze, which the players had to navigate from start to finish. The players played several rounds of this task. The players' score in the task is equal to the time in seconds that it took for them to complete the maze. Lower scores are better – the lower the score in seconds, the less time it took the player to navigate the maze.

Your job is to examine the profiles of these players, and decide how likely it is that Player A will win against Player B in a head-to-head competition. How the head-to-head round works is explained further below.

Every turn you are presented with a pair of profiles, which you must evaluate. There are nine turns.

The profiles include biographical information on the players.

The profiles will also include information on the players' average score over a number of rounds of the maze task.

This is some important information about the average score:

- The average score is the total score of the player from one or several rounds of play, divided by that number of rounds. For example, Player A may have taken 50 seconds to complete Round 1, 55 seconds to complete Round 2, and 40 seconds to complete Round 3. Their average score over three rounds is:
  50 + 55 + 40 / 3 = 50 seconds.

- The rounds which make up the average score are randomly chosen from the rounds played by the players. They do not necessarily include, for example, the first or last rounds played by them.

- The number of rounds used to make the average score can be between 1 and 5 rounds. The maze players played 6 rounds overall, but you will never see an average score based on all 6 rounds.

Evaluation

In every turn you are given a budget of 5 points, which you must invest in at least one of the players. You should base your investment on which player is more likely to win in the head-to-head competition. The head-to-head competition works like this: a random round will be

selected from the playing history of the two players. This is a round which has **not** been used to make the average score you see in the player profiles.

The winner of the head-to-head competition is the player who has performed better in their random round — that is, the player with the *lower* score in seconds. If it is a tied score, then one of the players is randomly chosen as the winner.

The points you earn in that round are equal to the number of points invested in the winner. This is explained further below.

In order to invest the points, you will have to adjust a slider beneath the player profiles. The slider represents how many points you wish to invest *in Player A*. For example, if you wish to invest 3 points in Player A, adjust the slider to 3. The rest of your points are invested automatically in Player B. If you invest 3 points in player A, you will automatically invest the remaining 2 points in player B.

Reward

For evaluating performance in the navigation task, you will receive a reward of up to £2.

Your reward is based on the number of points that you invest in the players in each round, and which player wins the head-to-head competition. If, for any given round, you invest 3 points in Player A and Player A wins then you will earn 3 points for that round. If, however, you invest 3 points in Player A and Player B wins then you will only earn 2 points for that round.

Since you can earn up to 5 points per round, and there are 9 rounds, you can earn up to 45 points. Every point is worth £0.044 — or 4.4 pence.

When you have finished the task, your points total will be calculated and converted into money.

If you earn the maximum of 35 points, you will earn £2.

Procedure

Before you begin the nine turns, you will first play a round of the maze task, to familiarise yourself with the nature of the task performed by the players you will be evaluating.

You will then play one practice turn of the evaluation task to familiarise yourself with the job of evaluating the players. You can't earn any points for this round. You will then play nine real turns of the evaluation task.

Appendix A2.4 Heterogeneities by even and uneven scores

A2.4.1 Even-score subsample

Table A2.2 shows the regression output for the analysis of the "even-score" subsample for Model A1 (culture task) and Model A2 (navigation task). These models are identical to models 1 and 2 in the main paper, but the sample is restricted to profile pairs with even scores.

Table A2.2 Multilevel estimates for "even-score" subsample.

	(MA1) Culture task	(MA2) Navigation task
Low Info (ref. cat.)		
Medium Info	-0.11	-0.007
	(0.11)	(0.12)
High Info	-0.15	-0.021
	(0.11)	(0.12)
Constant	0.73***	-0.025
	(0.08)	(0.09)
Participant-level error (std dev)	0.49***	0.49***
	(0.07)	(0.09)
Observation-level error (std dev)	0.88*	0.99
	(0.04)	(0.05)
Observations	483	469
Number of participants	303	304

Standard errors in parentheses.
Two-sided p-values: *** p<0.001, ** p<0.01, * p<0.05

A2.4.2 Uneven-score subsample

Table A2.3 shows the regression output for the analysis of the "uneven-score" subsample for models 4 and 5 in the main paper, which are here labelled models A3 and A4.

Table A2.3. Multilevel estimates for "uneven-score" subsample.

	(MA3) Culture task	(MA4) Navigation task
Better performing class/gender	0.60***	0.23**
	(0.08)	(0.07)
Low Info (ref. cat.)		
Medium Info	-0.017	-0.030
	(0.08)	(0.07)
Medium Info * Better performing class/gender	0.098	-0.038
	(0.12)	(0.10)
High Info	-0.16	0.18*
	(0.08)	(0.07)
High Info * Better performing class/gender	0.29**	-0.22*
	(0.11)	(0.10)
Score difference	-0.16***	-0.27***
	(0.01)	(0.03)
Constant	-0.021	0.31***
	(0.09)	(0.08)
Participant-level error (std dev)	0.54***	0.53***
	(0.03)	(0.03)
Observation-level error (std dev)	0.97	0.87***
	(0.02)	(0.02)
Observations	1923	1937
Number of participants	401	401

Standard errors in parentheses.
Two-sided p-values: *** p<0.001, ** p<0.01, * p<0.05

Appendix A2.5 Expected performance gaps by gender and class

The task evaluator's reward structure incentivized them to invest more points in whichever task performer they thought would perform better in a head-to-head round. In the head-to-had round, a round was randomly drawn from the playing histories of Player A and Player B. This randomly-drawn round *did not* come from any of the rounds which were averaged into the performance score shown to the task evaluator for that given pair. For example, if a profile pair was in the medium information condition, the performance scores would be an average of three rounds. The head-to-head round would be randomly drawn from the "remaining rounds" – the rounds *not* included in the performance score.

Task evaluators should choose whichever candidate they thought would do better in expectation. If there were statistical differences in performance by gender and class, then it would be "rational" to favour men / higher class people. Profile pairs were evenly matched on performance scores. As information increases, the chances that the task performers in a pair really are similar in ability increases and class / gender should play less of a role in evaluation.

In this appendix we present marginal effects from a simple bivariate model which uses OLS to estimate the expected gap in performance between genders / classes as a function of the information condition. The observations come from the universe of profile pairs for each task. The dependent variable is the difference in average score in the remining round between the higher and the lower class player in the pair (for the culture task), and the male and the female player in the pair (for the navigation task). The independent variable is the information condition. Robust standard errors are estimated.

Model A5 – culture task

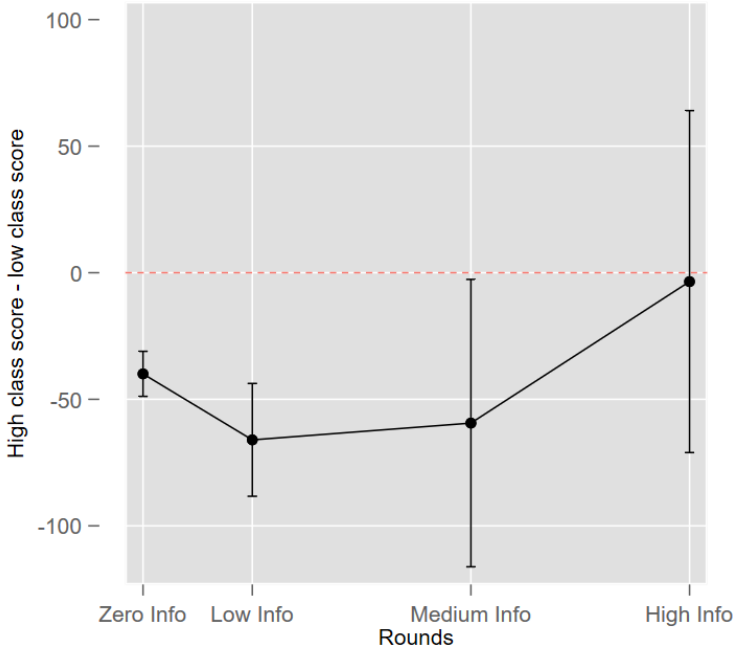
$$ExpectedScoreDifferenceClass_i = \beta_0 + \beta_1 LowInfo_i + \beta_2 MediumInfo_i + \beta_3 HighInfo_i + \varepsilon_i$$

Model A6 – navigation task

$$ExpectedScoreDifferenceGender_i = \beta_0 + \beta_1 LowInfo_i + \beta_2 MediumInfo_i + \beta_3 HighInfo_i + \varepsilon_i$$

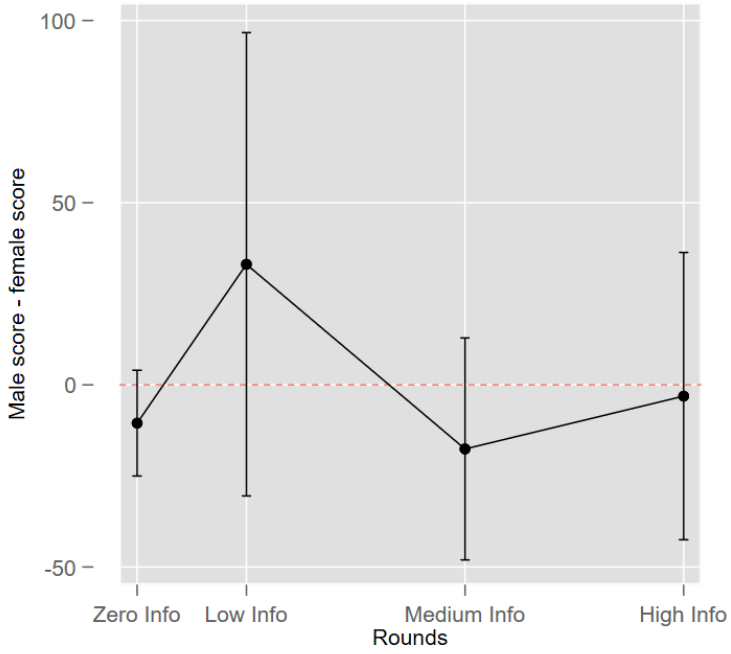
The average marginal effects from each model are graphed in Figures A2.2 and A2.3. For both tasks, lower scores represented better performance (less time spent on a solving a round). Hence a negative value means that the higher class (men) outperformed the lower class (women) on average in the remaining rounds. As can be seen in Figure A2.2, from a statistical discrimination perspective it was rational for task evaluators to discriminate against the lower class in the zero, low, and medium information conditions. But there were no expected performance differentials in the high information condition – so the task evaluators' persistent bias in this condition cost them money. For the gender task there is never a (statistically significant) difference in performance.

Figure A2.2. Expected difference in score in remaining rounds, culture task



Zero info condition = 0 rounds. Low info condition = 1 round.
Med info condition = 3 rounds. High info condition = 5 rounds.

Figure A2.3. Expected difference in score in remaining rounds, navigation task



Zero info condition = 0 rounds. Low info condition = 1 round.
Med info condition = 3 rounds. High info condition = 5 rounds.

References

Arrow, Kenneth J. 1971. *Some Models of Racial Discrimination in the Labour Market*. Santa Monica, CA.: RAND Corporation.

Auspurg, Katrin, Thomas Hinz, and Carsten Sauer. 2017. 'Why Should Women Get Less? Evidence on the Gender Pay Gap from Multifactorial Survey Experiments'. *American Sociological Review* 82(1):179–210.

Azmat, Ghazala, and Rosa Ferrer. 2017. 'Gender Gaps in Performance: Evidence from Young Lawyers'. *Journal of Political Economy* 125(5):1306–55.

Baert, Stijn. 2018. 'Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments since 2005'. *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* 63–77.

Battaglia, Frank, Samad Shah, Sabeena Jalal, Kiran Khurshid, Nupur Verma, Savvas Nicolaou, Sravanthi Reddy, Susan John, and Faisal Khosa. 2018. 'Gender Disparity in Academic Emergency Radiology'. *Emergency Radiology* 26(1):21–28.

Belzil, Christian, and Jorgen Hansen. 2002. 'Unobserved Ability and the Return to Schooling'. *Econometrica* 70(5):2075–91.

Berger, Joseph. 2014. 'The Standardized Experimental Situation in Expectation States Research: Notes on History, Uses, and Special Features'. Pp. 269–93 in *Labouratory experiments in the social sciences*, edited by M. Webster Jr. and J. Sell. Cambridge, MA: Elsevier Academic Press.

Berger, Joseph, Bernard P. Cohen, and Morris Zelditch Jr. 1972. 'Status Characteristics and Social Interaction'. *American Sociological Review* 241–55.

Blackburn, McKinsey, and David Neumark. 1992. 'Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials'. *The Quarterly Journal of Economics* 107(4):1421–36.

Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgement of Taste*. Cambridge, MA: Harvard University Press.

Bukodi, Erzsébet, and John H. Goldthorpe. 2018. *Social Mobility and Education in Britain*. Cambridge: Cambridge University Press.

Chan, Tak Wing, and John H. Goldthorpe. 2007. ‘Class and Status: The Conceptual Distinction and Its Empirical Relevance’. *American Sociological Review* 72(4):512–32.

Clancy, Patrick, and Sara O’Sullivan. 2020. ‘Gender Parity in Higher Education Enrolments: Trends and Paradoxes’. *Irish Educational Studies* 39(3):337–54.

Clifford, Scott, and Jennifer Jerit. 2015. ‘Do Attempts to Improve Respondent Attention Increase Social Desirability Bias?’ *Public Opinion Quarterly* 79(3):790–802.

Correll, Shelley J., and Stephen Benard. 2006. ‘Biased Estimators? Comparing Status and Statistical Theories of Gender Discrimination’. *Advances in Group Processes* 23:89–116.

Correll, Shelley J., Stephen Benard, and In Paik. 2007. ‘Getting a Job: Is There a Motherhood Penalty?’ *American Journal of Sociology* 112(5):1297–1339.

Correll, Shelley J., and Cecilia L. Ridgeway. 2006. ‘Expectation States Theory’. Pp. 29–51 in *Handbook of social psychology*. Springer.

Durlauf, Steven N., and Ananth Seshadri. 2018. ‘Understanding the Great Gatsby Curve’. *NBER Macroeconomics Annual* 32(1):333–93.

Eagly, Alice H., Christa Nater, David I. Miller, Michèle Kaufmann, and Sabine Sczesny. 2020. ‘Gender Stereotypes Have Changed: A Cross-Temporal Meta-Analysis of US Public Opinion Polls from 1946 to 2018.’ *American Psychologist* 75(3):301.

England, Paula. 2010. ‘The Gender Revolution: Uneven and Stalled’. *Gender & Society* 24(2):149–66.

Erikson, R., J. H. Goldthorpe, and L. Portocarero. 1979. ‘Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden’. *The British Journal of Sociology* 30(4):415–41.

Eyal, Peer, Rothschild David, Gordon Andrew, Evernden Zak, and Damer Ekaterina. 2021. ‘Data Quality of Platforms and Panels for Online Behavioural Research’. *Behaviour Research Methods*.

Fernandez, Roberto M., Emilio J. Castilla, and Paul Moore. 2000. ‘Social Capital at Work: Networks and Employment at a Phone Center’. *American Journal of Sociology* 105(5):1288–1356.

Foschi, Martha. 2000. ‘Double Standards for Competence: Theory and Research’. *Annual Review of Sociology* 26(1):21–42.

Friedman, Sam. 2016. ‘Habitus Clivé and the Emotional Imprint of Social Mobility”’. *Journal of Conflict Resolution* 64(1):1340–76.

Friedman, Sam, and Daniel Laurison. 2020. ‘The Class Ceiling: Why It Pays to Be Privileged’.

Gallen, Yana. 2018. ‘Motherhood and the Gender Productivity Gap’. in *Becker Friedman Institute for Research in Economics Working Paper*.

Ganzeboom, Harry B. G., Paul M. De Graaf, and Donald J. Treiman. 1992. ‘A Standard International Socio-Economic Index of Occupational Status’. *Social Science Research* 21(1):1–56.

Goda, Thomas, and Alejandro Torres García. 2016. ‘The Rising Tide of Absolute Global Income Inequality During 1850–2010: Is It Driven by Inequality Within or Between Countries?’ *Social Indicators Research* 130(3):1051–72.

Goldin, Claudia. 2014. ‘A Grand Gender Convergence: Its Last Chapter’. *American Economic Review* 104(4):1091–1119.

Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. 2006. ‘The Homecoming of American College Women: The Reversal of the College Gender Gap’. *Journal of Economic Perspectives* 20(4):133–56.

Gough, Margaret, and Mary Noonan. 2013. ‘A Review of the Motherhood Wage Penalty in the United States’. *Sociology Compass* 7(4):328–42.

Goulstone, J. 2000. ‘The Working-class Origins of Modern Football’. *The International Journal of the History of Sport* 17(1):135–43.

Hauser, David J., and Norbert Schwarz. 2015. ‘It’s a Trap! Instructional Manipulation Checks Prompt Systematic Thinking on “Tricky” Tasks’. *SAGE Open* 5(2):2158244015584617.

Hausmann, Markus, Daniela Schoofs, Harriet E. S. Rosenthal, and Kirsten Jordan. 2009. 'Interactive Effects of Sex Hormones and Gender Stereotypes on Cognitive Sex Differences—A Psychobiosocial Approach'. *Psychoneuroendocrinology* 34(3):389–401.

Hufe, Paul, Ravi Kanbur, and Andreas Peichl. 2022. 'Measuring Unfair Inequality: Reconciling Equality of Opportunity and Freedom from Poverty'. *The Review of Economic Studies* rdab101.

Jackson, Michelle. 2009. 'Disadvantaged through Discrimination? The Role of Employers in Social Stratification1'. *The British Journal of Sociology* 60(4):669–92.

Jerrim, J., and L. Macmillan. 2015. 'Income Inequality, Intergenerational Mobility, and the Great Gatsby Curve: Is Education the Key?' *Social Forces* 94(2):505–33.

Kalkhoff, Will, David Melamed, Josh Pollock, Brennan Miller, Jon Overton, and Matthew Pfeiffer. 2020. 'Cracking the Black Box: Capturing the Role of Expectation States in Status Processes'. *Social Psychology Quarterly* 83(1):26–48.

Knight, Carly R., and Mary C. Brinton. 2017. 'One Egalitarianism or Several? Two Decades of Gender-Role Attitude Change in Europe'. *American Journal of Sociology* 122(5):1485–1532.

Lahey, Joanna N., and Douglas R. Oxley. 2018. *Discrimination at the Intersection of Age, Race, and Gender: Evidence from a Lab-in-the-Field Experiment*. *Working Paper*. 25357. National Bureau of Economic Research.

Markus, Hazel R., and Susan T. Fiske. 2012. *Introduction: A Wide-Angle Lens on the Psychology of Social Class". Pp1-11 in Facing Social Class: How Societal Rank Influences Interaction*. edited by H.R. Markus and S.T. Fiske. New York: Russel Sage.

Marx, K. 1965. *Capital*. London: Lawrence and Wishart.

Mayer, Sabrina J., and Justus M. K. Rathmann. 2018. 'How Does Research Productivity Relate to Gender? Analyzing Gender Differences for Multiple Publication Dimensions'. *Scientometrics* 117(3):1663–93.

McCall, Leslie, and Christine Percheski. 2010. 'Income Inequality: New Trends and Research Directions'. *Annual Review of Sociology* 36(1):329–47.

Mijs, Jonathan J. B. 2021. ‘The Paradox of Inequality: Income Inequality and Belief in Meritocracy Go Hand in Hand’. *Socio-Economic Review* 19(1):7–35.

Neumark, David. 2018. ‘Experimental Research on Labour Market Discrimination’. *Journal of Economic Literature* 56(3):799–866.

Oesch, Daniel. 2006. ‘Coming to Grips with a Changing Class Structure: An Analysis of Employment Stratification in Britain, Germany, Sweden and Switzerland’. *International Sociology* 21(2):263–88.

Pager, Devah, and Hana Shepherd. 2008. ‘The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets’. *Annual Review of Sociology* 34(1):181–209.

Palan, Stefan, and Christian Schitter. 2018. ‘Prolific.Ac—A Subject Pool for Online Experiments’. *Journal of Behavioural and Experimental Finance* 17:22–27.

Peer, Eyal, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. ‘Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioural Research’. *Journal of Experimental Social Psychology* 70:153–63.

Phelps, Edmund S. 1972. ‘The Statistical Theory of Racism and Sexism’. *The American Economic Review* 62(4):659–61.

Rauh, Christopher. 2017. ‘Voting, Education, and the Great Gatsby Curve’. *Journal of Public Economics* 146:1–14.

Reardon, Sean F., Joseph P. Robinson-Cimpian, and Ericka S. Weathers. 2014. ‘Patterns and Trends in Racial/Ethnic and Socioeconomic Academic Achievement Gaps’. Pp. 507–25 in *Handbook of research in education finance and policy*, edited by H. F. Ladd and M. E. Goertz. London: Routledge.

Ridgeway, Cecilia L. 2019. *Status*. New York: Russell Sage.

Ridgeway, Cecilia L., and Susan R. Fisk. 2012. ‘Class Rules, Status Dynamics, and “Gateway” Interactions’. Pp. 131–51 in *Facing social class: How societal rank influences interaction*. New York, NY, US: Russell Sage Foundation.

Rissing, Ben A., and Emilio J. Castilla. 2014. ‘House of Green Cards’. *American Sociological Review* 79(6):1226–55.

Rivera, Lauren A. 2020. ‘Employer Decision Making’. *Annual Review of Sociology* 46(1):215–32.

Rivera, Lauren A., and András Tilcsik. 2016. ‘Class Advantage, Commitment Penalty: The Gendered Effect of Social Class Signals in an Elite Labour Market’. *American Sociological Review* 81(6):1097–1131.

Rosenthal, Harriet E. S., Liam Norman, Shamus P. Smith, and Anthony McGregor. 2012. ‘Gender-Based Navigation Stereotype Improves Men’s Search for a Hidden Goal’. *Sex Roles* 67(11):682–95.

Santos, Adailson Soares, Mário Teixeira Reis Neto, and Ernst Verwaal. 2018. ‘Does Cultural Capital Matter for Individual Job Performance? A Large-Scale Survey of the Impact of Cultural, Social and Psychological Capital on Individual Performance in Brazil’. *International Journal of Productivity and Performance Management* 67(8):1352–70.

Savage, Michael. 2015. *Social Class in the 21st Century*. London: Pelican.

Scarborough, William J., Ray Sin, and Barbara Risman. 2019. ‘Attitudes and the Stalled Gender Revolution: Egalitarianism, Traditionalism, and Ambivalence from 1977 through 2016’. *Gender & Society* 33(2):173–200.

Shurchkov, Olga. 2012. ‘Under Pressure: Gender Differences In Output Quality And Quantity Under Competition And Time Constraints Author’. *Journal of the European Economic Association* 10(5):1189–1213.

Silva, Elizabeth. 2008. ‘Cultural Capital and Visual Art in the Contemporary UK’. *Cultural Trends* 17(4):267–87.

Stockhammer, Engelbert. 2013. ‘Rising Inequality As a Cause of the Present Crisis’. *Cambridge Journal of Economics* 39(3):935–58.

Taber, Christopher R. 2001. ‘The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?’ *Review of Economic Studies* 68(3):665–91.

Thévenon, Olivier. 2013. ‘Drivers of Female Labour Force Participation in the OECD"’. *OECD Social, Employment and Migration Working Papers* (145).

Thewissen, Stefan, Lane Kenworthy, Brian Nolan, Max Roser, and Tim Smeeding. 2018. ‘Rising Income Inequality and Living Standards in OECD Countries: How Does the Middle Fare?’ *Journal of Income Distribution* 1–23.

Thijssen, Lex, Marcel Coenders, and Bram Lancee. 2021. ‘Is There Evidence for Statistical Discrimination against Ethnic Minorities in Hiring? Evidence from a Cross-National Field Experiment’. *Social Science Research* 93:102482.

Thomas, Kyla. 2018. ‘The Labour Market Value of Taste: An Experimental Study of Class Bias in U.S. Employment’. *Sociological Science* 5:562–95.

Vannette, Dave. 2017. ‘Using Attention Checks in Your Surveys May Harm Data Quality’.

Vincent-Lancrin, Stéphan. 2008. ‘The Reversal of Gender Inequalities in Higher Education: An On-Going Trend’.

Wagner, David G., and Joseph Berger. 1997. ‘Gender and Interpersonal Task Behaviours: Status Expectation Accounts’. *Sociological Perspectives* 40(1):1–32.

Chapter 3. To whom do you refer? Studying the effect of status on educational inequality.

Abstract

The literature on educational inequality has established that lower class students are less likely to attend university than their higher class counterparts, even when adjusting for differences in academic performance. I argue that working class students internalise status beliefs about their inferior academic competence and this is an important factor in their decision not to attend university. However, studying the effect of status beliefs is difficult, since it is hard to identify contexts where status beliefs vary and yet the objective characteristics determining status (e.g. income, occupational background) do not. I make use of the insight that subjective social class is influenced by an individual's reference group to identify an educational setting where status may vary but objective class characteristics don't. I argue that when a lower class student is placed in a higher class classroom, they are seen, by themselves as others, as subjectively more lower class than they would be if placed in a lower class classroom. Since they are seen as lower class, they internalise negative status beliefs about lower class people. Hence, being placed in a higher class setting reduces expected academic ability, and hence reduces the lower class students' likelihood of attending university – a prediction at odds with the positive peer effect mechanism. I present two studies, one applying a quasi-experimental design to observational data, and the other applying a vignette method, that provide complementary evidence that status is a consequential factor in determining educational inequality.

3.1 Introduction

The empirical results of Chapter 2 demonstrated that there are undoubtedly status beliefs about social class. This shows that the "class = competence" belief leads to biased third party assessments of performance. In this chapter I address RQ2: Do internalisations of the "class = competence" belief substantively contribute to inequality? In particular, I study how internalisations of the belief lead to inequalities in university attainment between classes.

It is very much apparent that there are big payoffs to a university education. Those who complete a college degree can expect to benefit from a substantial – and growing – wage premium (Crivellaro 2016; van der Velden and Bijlsma 2016). University qualifications grant access to esteemed and respected occupations (Ganzeboom, De Graaf, and Treiman 1992; Treiman 1977). But despite the potential reward of attending university, working class students who are as academically capable as their middle class peers are less likely to attend university (Bernardi and Triventi 2020; Breen and Jonsson 2000; Jackson and Jonsson 2013). What is stopping them?

Researchers have argued that internalised stereotypes play an important role in educational inequality between classes (Durante and Fiske 2017; Gorski 2012; Jury et al. 2017). In the sociological literature, stereotypes about the competence of social groups are also termed "status beliefs" (Ridgeway 2019). It is plausible that working class students internalise negative stereotypes, or status beliefs, about the lower academic competence of working class people. The negative status beliefs make them think they aren't good enough to go to college, and so they don't go. Though the argument is intuitive, it is very hard to empirically test. The essential problem is that internalised status beliefs are attached to objective class positions. It is therefore difficult to separate the effect of the status belief from the objective class characteristics – such as financial resources, parental strategies, opportunity costs – associated with particular class positions.

The solution is to find a context in which status beliefs vary exogenously with respect to objective class characteristics. The first step is to recognise that status beliefs are attached to *perceived* class position: if an individual perceives themselves to belong to a certain class, they will internalise the status belief pertaining to that class. One potential approach then is to identify a situation where perceived class varies exogenously. In this case, the ambiguity that has plagued conceptual and empirical analyses of class is actually the researcher's friend.

Because individual class position is ambiguous, people make use of heuristics to estimate their own and others' class position. Hence, exogeneous variation in the heuristic will lead to variation in perceived social class without affecting objective class position at all.

The heuristic I identify is the reference group. Previous research has established that people commonly use their reference group as a heuristic to estimate their own and others' class position (Evans and Kelley 2004). In the field of educational inequality there is a very consequential and relevant reference group: a student's classmates (McVicar and Polanski 2014; Wouters et al. 2013). I argue that students take their classmates as a reference group, and estimate their own and others' social class in terms of the individual location within the classroom socioeconomic distribution. When a student is placed in a higher class classroom, their perceived social class will be lower than if they are placed in a lower class context because they will occupy a relatively lower position in the classroom socioeconomic distribution. Since they are perceived – by themselves and others – as relatively lower class they will internalise status beliefs about lower class people, which will negatively affect their educational attainment. This leads to the conclusion – at odds with the prevailing literature (van Ewijk and Sleegers 2010) – that working class students who are placed in classrooms with higher class students will have lower educational attainment.

This chapter pursues the argument over two complementary empirical studies, one based on observational data with a plausibly causal design, and the other on a vignette experiment. The complementary results provide evidence that status beliefs contribute to inequality of university attainment between classes.

3.2 Theoretical approach

3.2.1 Status and educational inequality

There are two channels through status beliefs affect educational inequality between social classes. First, there is the "direct effect" – status beliefs affect perceived ability, which also influences a student decision as to whether they are "good enough" to go to university. When a student is considering the decision to attend university, they are likely aware that the academic tasks they face at university will be qualitatively different to those they face in school, as will the learning environment. Hence, status beliefs can help them determine whether someone belonging to their social group would normally excel at university. Correll (2001; 2004) has found that, controlling for actual observed ability, women are likely to have

lower self-assessed ability in mathematical areas when compared to men, which makes them less likely to pursue educational paths through the STEM disciplines. Status beliefs about class may work analogously. Lower class students may internalise status beliefs that they are academically incompetent and decide not to continue their education after secondary school.

Second, there is the "indirect effect" – status beliefs can directly influence school grades. Students will use academic achievement in secondary school as a predictor for likely achievement at university. Many university systems also impose academic requirements for entry. Status beliefs operate as a form of self-fulfilling prophecy which can depress or boost performance (Ridgeway 2019). Those who come from high-status groups internalise the belief that they are competent, which increases confidence in their abilities, which means they actually do perform better than they would absent status beliefs. Through the same mechanism, lower-status groups perform worse. There is evidence that when socioeconomic identity is made salient, working class students perform worse on tests (Croizet et al. 2001; Spencer and Castano 2007), suggesting that prevalent beliefs about the competence of classes can directly effect performance.

Status beliefs are both "first-order" and "second-order" beliefs (Correll et al. 2017). First-order beliefs are an individual's own beliefs about themselves. Second-order beliefs are an individual's beliefs about what others in their environment believe about them. In the classroom, status beliefs can operate through first- and second- order channels. First-order beliefs can form when a given student estimates their own social class in comparison to that of their classmates. Second-order beliefs can form when a given student internalises their classmates' estimates of the student's own social class. Research has shown that second-order beliefs may dominate first-order ones (Correll et al. 2017; Troyer and Younts 1997). Second-order beliefs are an important component of the "internalisation" mechanism – for individuals to internalise a belief about themselves they must believe others hold it. In this paper, Study 1 investigates the existence of first-order status beliefs, and Study 2 investigates the existence of second-order status beliefs.

3.2.2 The ambiguity of class

Social class has proven notoriously tricky to define; one reason why it does not neatly fit into the framework developed by status theorists (Ridgeway and Fisk 2012; Rivera 2015). Indeed, differences over the definition of "class" go back to the roots of sociology (see Wright (2015)

for a comparative discussion of Marxian and Weberian approaches, among others). Epistemically speaking, class has something in common with the American Supreme Court Justice Potter Stewart's definition of pornography. Admitting he could never define the concept in an absolutely satisfying way, the judge nonetheless claimed "I know it when I see it" (Gewirtz 1996). As I will argue, social class, as a concept, is something that people readily use in everyday life, even if they experience some ambiguity over what exactly the concept means and where they fit into the class structure. This combination of ambiguity and tractability is essential for the empirical strategy of this paper.

Many surveys have shown that people have little difficulty in sorting themselves into class positions when asked to, and their self-categorisation tends to strongly correlate with their income, education and occupation (Curtis 2016; Dineen, Robbins, and Simonsen 2019; Sosnaud, Brady, and Frenk 2013). This suggests that when people are asked to self-describe their social class, they regard the question as sufficiently meaningful to give an answer, and that they use the concept in a fairly reliable and consistent way. But, despite the general reliability of the concept, it seems that there is some degree of disagreement, or imprecision, about where exactly people are placed in the class hierarchy. For example, Sosnaud, Brady, and Frenk (2013) find, using American survey data, that while most people with "objectively" working class occupations identify as working class, about 37 percent do not.

Conceptual ambiguity no doubt contributes to this fuzziness. But another important factor is that the attributes associated with social class, such as income, occupation, and education, are more-or-less continuously distributed – and it is hard for people to locate themselves precisely in a continuous distribution. In the case of income, people do not have very precise information on the income distribution in their country, and their perception of where they are in that distribution can vary substantially (Cansunar 2021; Senik 2004). While education may be conceived in categorical terms, distinctions can also be made in terms of fine-grained differences such as the quality of school or college attended (Boliver 2013; Rivera 2011). Occupation may also be conceptualised continuously. Work on occupational prestige scales shows that people are able to make fairly fine-grained distinctions in the hierarchical position of occupations (Treiman 1977). As Oesch and Vigna put it: "even high-status people consider many acquaintances to be above them (family doctors looking up to medical school professors) and low-status people see others even lower (factory workers looking down on sweepers)" (2021: 5).

In short, people face both conceptual and empirical uncertainty about their class position. The combination of conceptual ambiguity and imperfectly known, continuously distributed signals of social class helps to explain why people's class position, in their own eyes as well as those of others, can vary across individuals with objectively identical socioeconomic attributes. This is not to say that there are extreme variations – as pointed out, self-identification correlates strongly enough with objective indicators. But there is sufficient variation for some, potentially substantial, uncertainty about subjective identification.

3.2.3 The role of reference groups in determining perceived class position

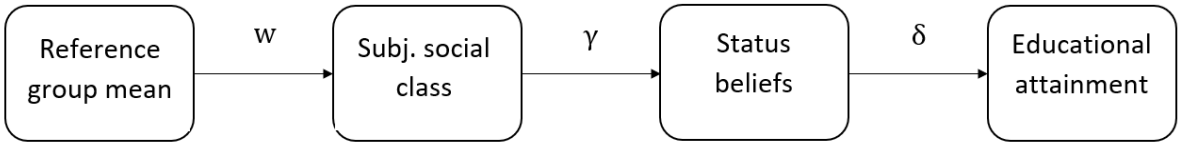
Since people are uncertain about their own and others' class position, they are likely to resort to a heuristic, or rule of thumb, to estimate it. One such heuristic is the reference group – the set of people an individual compares themselves to (Merton 1968). Numerous works have demonstrated that a person's perceived position within some vertical hierarchy (be it of status, success, or some other vertically scalable good or bad) can be strongly influenced by who they are compared to (Crocker et al. 1987; Friend and Gilbert 1973; Morse and Gergen 1970).

Evans and Kelley (2004) argue that an individual's reference group plays an important role in determining their subjective class position. They argue that people find it difficult to fix their exact place in the class hierarchy so they use their position within their reference group of acquaintances as a (partial) proxy. In data drawn from 21 countries, they show that people tend to place themselves towards the middle of the distribution, a finding which they present as evidence that people's position within their reference group partially determines their subjective social class position. Since processes of homophily, segregation, selection, and closure tend to mean that people's reference groups are composed of people like them – with a few people above them and a few below – people, they argue, tend to see themselves as being somewhere in the centre of the distribution. Cansunar (2021) and Dineen, Robbins and Simonsen (2019) advance similar arguments with supporting data.

If the socioeconomic distribution of one's reference group partially determines one's subjective social class identification, then changes in the reference group should change that identification. Hence, I argue that we can use shifts in reference group composition to examine shifts in status. For example, someone who originally attended a predominantly lower class school and who enters a school where most students are from higher class

background may begin to feel themselves to be "more" lower class than they previously felt themselves to be. Likewise, their new acquaintances will see them as further down the class ladder than their old schoolfriends did. So, both first-order and second-order beliefs will place the student in a lower class position. Though their objective class position has not changed at all, their perceived or subjective class position has changed. They will internalise negative status beliefs about their subjectively lower class position. These beliefs will make them feel insecure about their chances of successfully attaining a university qualification. The same argument applies in reverse, of course. If a student moves to a lower class environment they will benefit from positive status beliefs.

Figure 3.1. Graphical representation of the status-reference group model.



This argument is summarised in Figure 3.1, which graphically summarises my proposed status-reference group model in terms of four variables: (i) (mean social class of) reference group; (ii) subjective social class; (iii) status beliefs; (iv) expected educational attainment. Each edge is labelled with a coefficient (w, γ, δ), that summarises the strength of relationship between the variables.

The coefficient w determines the influence of the reference group mean on subjective social class – if this were 0, then comparisons to the reference group would have zero influence. The coefficient γ indexes the strength of relationship between perceived class position and applicable status beliefs. And the coefficient δ indexes the strength of the relationship between status beliefs and educational attainment.

In appendix A3.1, I convert the verbal argument into a simple algebraic model which outputs the graph in Figure 3.1. Given the algebraic model, we can summarise the effect of a change in the mean social class position of the reference group, $\overline{x_k}$, on the probability an individual expects to university, π_i, with the following differential equation, where *k* indexes the reference group and *i* indexes the individual:

$$\frac{d\pi_i}{d\overline{x_k}} = -\delta_i \gamma_i w_k < 0 \qquad (1)$$

This equation allows us to test implications of the model, if we have access to a measure, or a proxy, for any of the coefficients. The core implication of the model is:

Implication 1. The higher the mean social class position of the reference group, the lower the probability an individual expects to attend university.

Given the data that is available, in this chapter I will test a second implication, based on the w_k parameter, the weight given in comparisons by the individual to the reference group mean. The core of this implication is that people do not attach equal weight to comparisons with all (potential) members of their reference group (Hyman 1960; Shibutani 1955). Applying this intuition to the specific case of the paper, it may be that some individuals may give more weight to comparisons with a specific reference group than others do, and hence the socioeconomic composition of that group will have more influence in the formation of their subjective social class position.

Implication 2 The effect of reference group composition will be moderated by the weight assigned by the individual to comparisons to that group.

Implication 2 is tested in Study 2.

3.3 Study 1: Reference groups and first-order beliefs

3.3.1 Empirical strategy and context

This study uses data from the first wave of the CILS4EU study, carried out in England, the Netherlands, Sweden, and Germany in 2010 and 2011. (Dutch and English data are excluded from the analytic sample for reasons of data availability). This wave of the study interviewed secondary school students who were around 14 years old. This study was chosen because it provides information on which classroom within a school the student was located in, information which is not provided by the PISA study for example.

The primary independent variable in this study is the socioeconomic composition of a student's classmates, and the dependent variable is whether the student expects to go to university. I follow Legewie and DiPrete (2012) and operationalise socioeconomic position using the International Socio-Economic Index (ISEI) of occupational status (Ganzeboom et al. 1992). Theoretically, the ISEI approach conceives of occupations as social positions which convert education into income. The scale is designed to assign scores to occupations so that the indirect effect of education on income, mediated by occupation, is maximised, and

the direct effect of education on income, "bypassing" occupation, is minimised. The socioeconomic composition of a student's classmates is operationalised by the mean of the student's classmates' parents' ISEI.

Of course, one cannot assume that the socioeconomic composition of a school population is exogeneous. There are likely to be a number of important confounders. For starters, schools with a higher class socioeconomic composition are likely to provide higher quality education (Hochschild 2021). Secondly, schools with a higher class socioeconomic composition are likely to be more selective, so the working class children who end up there may tend to be high achievers (Robert 2010).

To exclude endogeneity, this study makes use of a design previously implemented by McEwan (2003), Ammermüller and Pischke (2006) and Legewie and DiPrete (2012). These previous studies argued that, given certain assumptions, assignment to classrooms *within schools* could be treated as as-if random. Using fixed effects at a school-level to soak up endogeneity from selection and confounding on unobservables, we can treat classroom-level attributes – such as the socioeconomic composition of the students in the classroom – as plausibly exogeneous. (By classroom it is meant a stable unit of students who usually take their lessons together at the same time from the same teacher. This is more properly called a "class", but the word classroom is used to avoid confusion with "social class"). Hence, we can treat a student's classmates as their reference group and the mean socioeconomic composition of this group as our exogeneous variable of interest.

But is assignment to classrooms is really exogeneous in the data? The most important threat to identification is the fact that schools sometimes make use of "streaming": stratification of students into classes by ability, so that students of the highest ability are in a given class, students of the lowest ability in another, and so on. Including schools which stream in the data would induce selection bias. Since class correlates with academic ability, working class students selected for their ability into the higher-achieving classrooms would also be selected into classrooms with a higher mean socioeconomic position. Hence, the effect of classroom socioeconomic status would be confounded by (unobserved) student ability. Rather than estimating β, the unbiased coefficient on the independent variable $\overline{x_k}$, the mean socioeconomic position of the reference group, we would estimate $\tilde{\beta} = \beta + \gamma\delta$ where γ represents the relationship between unobserved ability and the dependent variable, university

attendance, and δ represents the relationship between unobserved ability and $\overline{x_k}$. Since both of these parameters will be positive, the estimated coefficient will be positively biased, attenuating or reversing the hypothesised negative status effect.

Hence, we want to exclude schools which stream. Unfortunately, there is no variable in the dataset which reports the school's official policy on classroom assignment[12]. In practice this means I have to rely on the reports of the students themselves, who were asked if their school implemented a streaming policy. These reports are not always reliable – in some schools a proportion of students between 0 and 1 reported streaming, meaning that there was not total agreement among students about the policy. Ideally we would exclude all schools from the analytic sample where a proportion greater than zero report streaming, but this is costly in terms of statistical power. Hence I opted to include schools where twenty percent or fewer of students reported streaming, on the assumption that a small portion students may be unreliable informants regarding school policy and hence may give a mistaken answer.

However, the gain in statistical power by including these extra schools can only be justified if the randomness of assignment is not compromised. I follow Legewie and diPrete (2012) in using simulations in order to increase confidence in the randomness of assignment in the analytical sample. I carried out randomisation tests on three subsamples: those where 0% of students reported streaming ("0% sample"); those where up to 20% reported streaming ("20% sample"); and those where 80% or more reported streaming ("80% sample").

The goal is to see whether the pattern of assignment to classrooms in the subsamples differs from the pattern we would expect to observe under randomness. In this case, I am interested in whether the assignment to treatment (classroom) is correlated with the socioeconomic composition of the classroom, which I measure using the classroom mean parental ISEI. The randomisation tests check whether the mean parental ISEI is unequally distributed between classrooms within schools. If it is, this would suggest assignment is correlated with treatment. Since the details of this test are rather technical, a more detailed description of the tests and presentation of the results is available in appendix A3.2.

[12] I had access to the reduced version of the dataset only. The full version of the dataset includes a survey of teachers which would provide more reliable information on school policy. But access to the full version requires on-site presence in Cologne, Germany, which was not feasible.

In brief, the randomisation tests randomly reassign students to "counterfactual" classrooms. While in the observed data, a particular student is in classroom 1, in the randomised data they may ben randomly assigned to classroom 1c or 2c (where "c" indicates "counterfactual"). The test then estimates a counterfactual statistic for the average gap in parental ISEI between the counterfactual classrooms within schools. The tests perform this operation for 10,000 iterations to generate a null distribution for the size of the counterfactual ISEI gap between classrooms if assignment to classrooms were truly random. The actually-observed gap in parental ISEI between classrooms within schools can then be compared to the counterfactual distribution, and a p-value generated by estimating the proportion of counterfactual gaps more extreme than the observed gap.

The p-value for the 0 percent sample was 0.29, and for the 20 percent sample was 0.14, hence the null hypothesis of random assignment was not rejected for either subsample. As might be expected however, the p-value for the 80 percent sample was small (0.01) and would lead to rejection of the null hypothesis. In sum, the tests indicate that usage of the 20 percent sample is justified. It should also be borne in mind that any selection effects will tend to reduce the magnitude of the negative effect of classroom composition predicted here. For robustness, appendix 3.3 replicates the main analysis in study 1, using the 0 percent sample. The direction of the effects is the same, but some covariates have become statistically insignificant as might be expected with a smaller sample.

3.3.2 The mitigating role of peer effects

While the reference group composition may be exogeneous, it does not follow that only the "status effect" is identified. In fact, while the status effect would predict a negative effect of a higher socioeconomic composition of the reference group, a long tradition going back to the "Coleman report" (Coleman et al. 1966) has asserted that having higher class peers has a positive effect on educational achievement. Theoretically, there is good reason to suppose there is a positive peer effect. Students from higher class backgrounds have been socialised to exhibit conscientiousness, discipline, and active educational engagement, behaviours which can lead to positive spillover effects for their classmates by generating a classroom culture more conducive to learning (Legewie and DiPrete 2012; Redford, Johnson, and Honnold 2009; Rumberger and Palardy 2005).

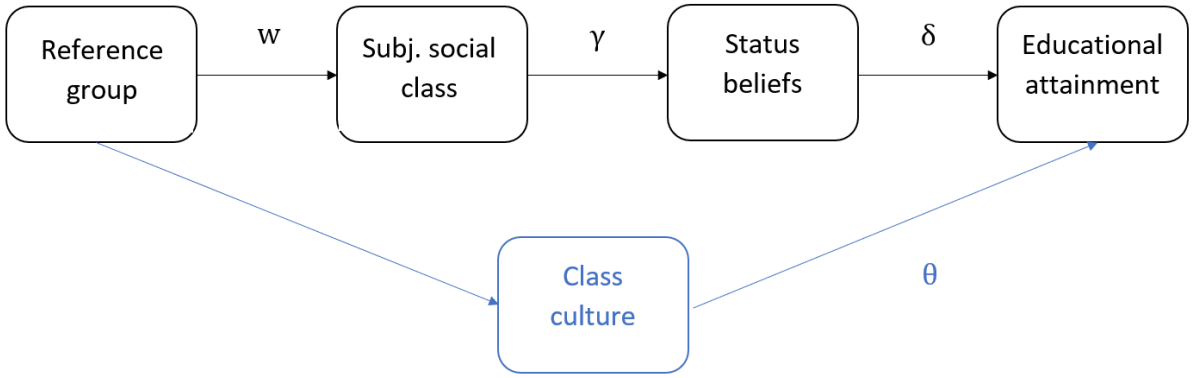
In a meta-review of studies on peer socioeconomic status, van Ewijk and Sleegers (2010) find an average weighted effect size of 0.32 standard deviations across all studies. However, there was considerable variance in the distribution of the effect size, with some studies exhibiting a null effect. As discussed above, the effect of schoolmate socioeconomic composition on educational achievement is likely to be biased upwards by selection effects. To exclude endogeneity, several studies have made use of the within-school classroom assignment strategy utilised here. McEwan (2003) and Ammermueller and Pischke (2009) find a positive effect of average classroom socioeconomic position on student achievement. Legewie and diPrete (2012) adopt a similar approach, interacting classroom socioeconomic composition with gender. They find a positive effect for boys, but a null effect for girls.

Given the likely presence of positive peer effects, the model I presented above must be modified. I introduce the parameter θ to represent what I term the "class culture" mechanism, an umbrella term which groups together the various positive spillover effects generated by higher class students. I modify equation (1) to account for this new parameter (also derived in appendix 3.1):

$$\frac{d\pi_i}{d\bar{x}_k} = \theta - \delta_i \gamma_i w_k \qquad (2)$$

Figure 3.2 modifies the graph in Figure 3.1 to take account of this new parameter.

Figure 3.2. Graphical representation of the status-reference group model, with class culture mechanism added.



It is now unclear whether the quantity represented in equation (2) will be positive or negative, since it depends on whether the negative status effect or positive class culture effect

predominates. The literature suggests that the net effect should be positive. However, there is good reason to suppose that the effect is heterogeneous, dependent on the individual student's social class, because a lower class students resolve to go to university is likely to be less resilient to negative shocks. The default lifeplan for higher class students is to go to university, and their parents invest substantial efforts and resources in ensuring their children reproduce their own class position (Becker and Hecken 2009; Breen and Goldthorpe 1997; Redford, Johnson, and Honnold 2009). They may receive a boost to their confidence from being in class with lower class students, but they are already going to college anyway so their decision won't be materially affected. On the other hand, being placed in a setting with other higher class students is hardly likely to make them feel so inferior on the class dimension that they decide not to go. Hence, for higher class students, the class culture mechanism may predominate.

Lower class families tend to invest less in planning and preparation for a university education, to overestimate the chances of failure at university, and to experience more uncertainty over whether it is the right decision to go (Barone et al. 2018). Given that even well-performing lower class students may be "on the fence", the status effect may be enough to push them off despite the rewards of university attainment. Hence, for them, the status effect may dominate the class culture mechanism.

3.3.3 Hypotheses

For Study 1, the first hypothesis spells out the basic claim of this study:

Hypothesis 1.1 The higher the average socioeconomic position of a student's classmates, the lower their expectation of attending university.

The second hypothesis captures the expected heterogeneity of the effect by individual class background.

Hypothesis 1.2 The effect of reference group composition will be weaker for higher class students.

3.3.3 Data and methods

Data comes from the German, and Swedish samples of wave 1 of the Children of Immigrants Longitudinal Study in Four European Countries (CILS4EU) survey. The dependent variable is binary: whether or not the student expects to go to university. This variable was recoded

from Question 14 of the main student survey, which asked students what the highest level of education they expected to achieve was. Answering "university" was coded as 1. Answering a lower level of education was coded as 0. Answering "don't know" was also coded as 0. Appendix A3.4 shows that the results are substantively the same when excluding participants who answered "don't know".

The main independent variable of interest is the mean classroom parental ISEI – i.e. the mean position of the student's classmates' parents on the International Socio-Economic Index (ISEI) of occupational status (Ganzeboom et al. 1992). The higher the value of this variable, the more "high class" a student's classmates are. Parental ISEI is measured as the average of both parents' individual ISEI; if one parent has a missing value then it is measured as the ISEI for the non-missing parent.

I interact the classroom average ISEI with the student's social class. The latter is also measured by parental ISEI. I dichotomise this variable by splitting it at the median, to create a binary variable HigherClassStudent which takes the value of 0 if the student has a parental ISEI below the median and 1 otherwise. I estimate the following regression model with the subscript *i* indicating the individual, *j* the classroom, and *k* the school:

$$y_i = \beta_0 + \beta_1 HigherClassStudent_i + \beta_2 ClassroomISEI_j + \beta_3 HigherClassStudent_i * ClassroomISEI_j + X_k + \epsilon_{ik} \quad (M1)$$

Here y_i is the probability that the student expects they will attend university after secondary school, and X_k is a vector of school fixed effects. Errors are clustered at the school level. The model is estimated using OLS, making it a linear probability model.

3.3.3 Results

Table 3.2 summarises the results. Since regression output tables for interaction effects are often difficult to interpret, the average marginal effects are graphically summarised in Figure 3.3. Two lines are shown: a black line for the effect of student's social class when students are in a classroom with a mean ISEI score in the 25th percentile (low average social class); and a blue line for the effect of ISEI quartile when students are in a classroom with a mean ISEI score in the 75th percentile (high average social class).

For reference, in the analytic sample the 25th percentile corresponds to a mean classroom ISEI of 36 while the 75th percentile corresponds to a mean classroom ISEI of 50. Occupations

at about 50 on the ISEI include receptionists, clerk, photographer, medical technician. Occupations at about 36 include salesperson, post deliverer, plumber. Hence there is not a huge gulf in class position across the interquartile range. Even if we look only at the top quartile of classrooms in terms of mean ISEI, 24% of students are still "lower class" (i.e have a parental ISEI below the median). A child of plumbers and post deliverers could easily end up in a classroom with the children of receptionists and clerks, making the comparison socially meaningful.

Table 3.2. Effect of classroom socioeconomic composition on expecting to attend university

Higher class student	-0.253**
	(0.081)
Classroom ISEI	-0.005*
	(0.002)
Higher class student * Classroom ISEI	0.008***
	(0.002)
Constant	0.412***
	(0.078)
Observations	3,224
R-squared	0.314

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

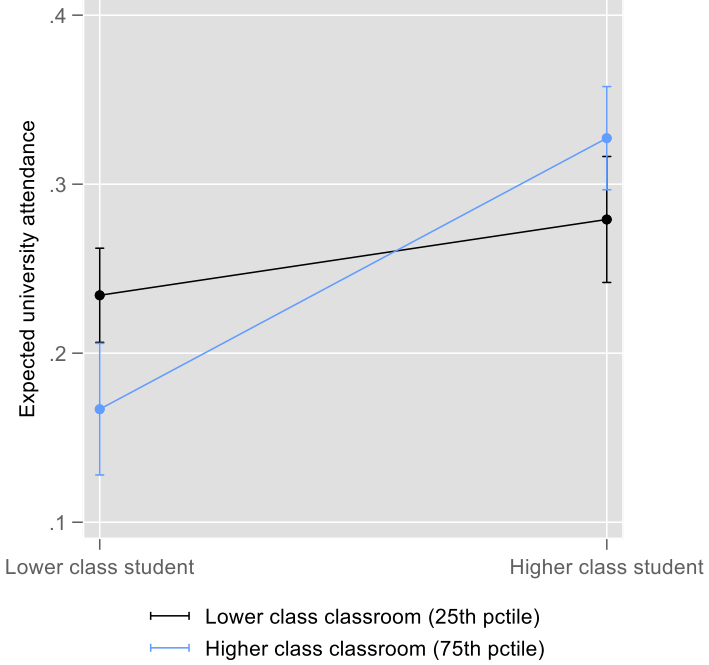
In the theory section I suggested that classroom socioeconomic composition should have both a positive effect through the class culture mechanism and a negative effect through the status mechanism, meaning that the overall direction of effect was theoretically ambiguous. The results indicate that the overall effect of classroom socioeconomic composition is dependent on the social class of the individual student. Hypothesis 1.1 predicted that the effect of higher classroom ISEI would be negative – which it is for lower class students: the coefficient on β_2, which measures the effect of classroom mean ISEI on a lower class student, is negative and statistically significant. Hence, hypothesis 1.1 is partially or conditionally supported: for

lower class students, the negative status effect dominates the positive class culture mechanism: $\theta_i < \delta_i \gamma_i w_k$, in the terms of equation 2.

Hypothesis 1.2 predicted that the status effect would be weaker for higher class students. This appears to be the case, since the positive effect of the class culture mechanism seems to outweigh the negative effect of the status mechanism for higher class students: $\beta_3 - \beta_2 = 0.008 - 0.005 > 0$, and β_3 is statistically significant. Figure 3.3 shows the contrasting direction of effects for each class, though the difference between the 25th and 75th percentile is not statistically significant for higher class students.

The effect for lower class students is quite substantial. When we look at the marginal effects for the analytic sample, lower class students have a marginal probability of 0.167 to expect to attend university when placed in a classroom where the mean ISEI level is high (at the 75th percentile). This marginal probability rises to 0.234 when students are placed in a classroom where the mean parental ISEI is low (at the 25th percentile), yielding an odds ratio of 1.5 and a gap of 7 percentage points. Considering that this effect is the difference between being in a classroom with the children of plumbers and post deliverers versus clerks and technicians, the effect would appear to be quite large and a testament to the consequentiality of status as a mechanism for reproducing inequality.

Figure 3.3. Average marginal effect of classroom ISEI, for a lower and higher class student



3.4 Study 2: Reference groups and second-order beliefs

The first study provided evidence that being placed in a higher social class reference group makes working class students less likely to expect to attend university. Study 1 measured first-order beliefs – a student's own beliefs about whether they will attend university. However, status theory predicts that status beliefs will also operate through second-order beliefs: beliefs about what others believe. Study 2 provides evidence that reference group composition affects second-order beliefs too.

In the context of this study, second-order beliefs are what an individual student believes their classmates believe. Establishing the existence of second-order beliefs is crucial to the story presented here: individuals *internalise* beliefs about the competence of social classes. Internalising a belief means that people are generally aware that others hold the belief. Second-order beliefs give status beliefs something of the character of descriptive norms: rules that people follow in expectation that others expect that they will follow them (Bicchieri 2017). Following descriptive norms facilitates the coordination with others that is required in everyday social interaction. As is well known, the effect of norms can be powerful. Researchers have found that second-order status beliefs reinforce, and even dominate, first-order status beliefs (Correll et al. 2017; Troyer and Younts 1997).

3.4.1 Experimental design and context

Study 2 is a vignette experiment. Eighty-three participants, secondary school students from an all-girls school in the south-east of Ireland, took part in the study. The participants were about 16 years of age.

The vignettes were profiles of hypothetical secondary school-graduates, students who had just finished their final school-leaving exams (known as the "leaving certificate" in Ireland) and were deciding what to do with their careers. The profiles were presented in a concise Q&A format, with the school-graduate answering questions such as "what is your name?", and "what do your parents do?". In order to elicit second-order beliefs, the experimental participants were asked what they thought *their classmates thought* the school-graduates would do after secondary school. As an incentive, they were told that the "closer" their answers were to the average answer of the sample, the more likely they would be to win one of three prizes: 100 euro for 1st place, 75 euro for 2nd place, and 50 euro for 3rd place.

Strictly speaking, the incentive does not necessarily elicit second-order beliefs – it is a coordination game. However, the second order belief is a highly salient coordination point, and it is implausible that higher-order reasoning would flip a participant's strategy. Consider a coordination game with symmetric payoffs where participants are rewarded for guessing whether most of their classmates think that homework is fun. Assume a subject thinks most participants think homework is not fun. If they only naively answer based on what they think most participants think, they will guess homework is not fun. If their reasoning becomes more sophisticated and they think most participants think most participants think homework is not fun, there does not appear to be any further reason to deviate from the answer they give in the naïve scenario. This conclusion follows no matter how reflexive the chain of reasoning becomes.

The participants were given four options for what the school-graduate could do: begin work immediately, do a one-year post-leaving certificate course, attend an institute of technology or attend university. For the purpose of this study, these four options are collapsed into a binary: attend university or not. If a participant responded that they thought their classmates would expect the student to begin work immediately or attend a post-leaving certificate course this was coded as 0, meaning "not attend university." If a participant responded that they thought their classmates would expect the student to attend an institute of technology or a university this was coded as 1, meaning "attend university." Appendix A3.5 provides a substantive and empirical rationale for this simplification. In short, both institutes of technology are functionally identical to universities, whereas post-leaving certificate qualifications provide short courses offering practical training. To achieve consonance with the terminology of Study 1 and of the literature at large, I shall use "university" to encompass both "institutes of technology" and universities proper.

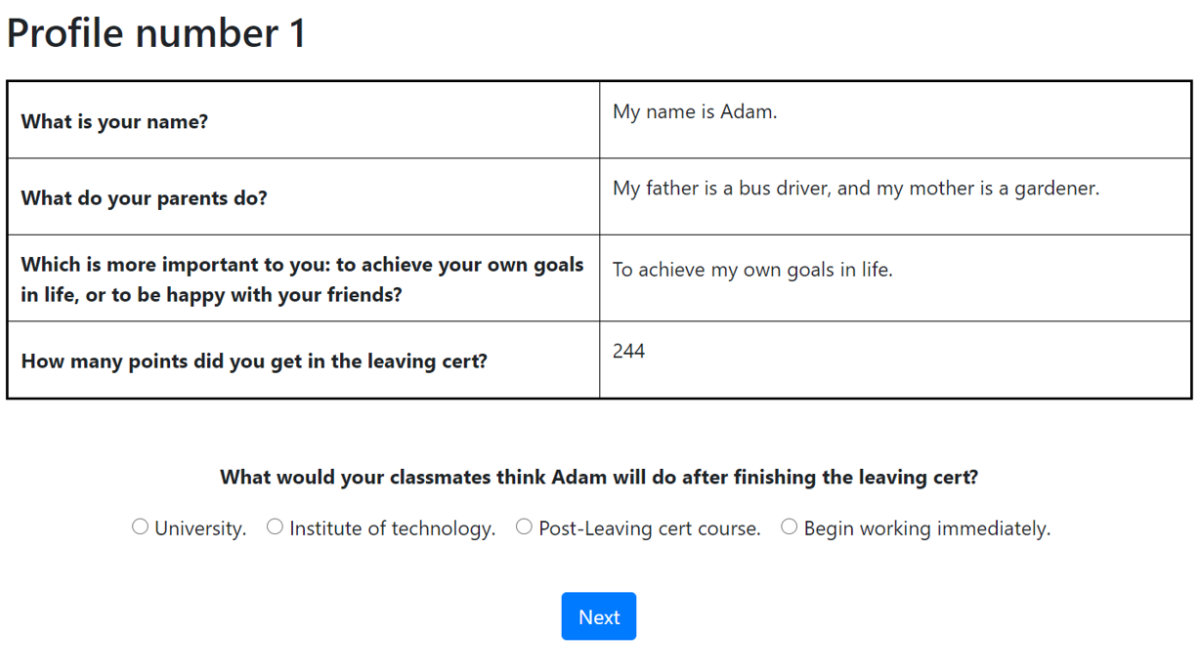
The vignettes featured four different dimensions: 1. The school graduate's name; 2. The school graduate's parents' occupations; 3. The school graduate's self-construal; 4. The grade (number of points) that the school-graduate achieved in the leaving certificate. An example of one such vignette is presented in Figure 3.4.

The parents' occupation is supposed to signal the school-graduate's social class background. While class is a multidimensional variable, occupation is usually the core conceptual dimension (Goldthorpe and McKnight 2006; Erik O. Wright 2015). Occupations were chosen

so that, for lower class backgrounds, they had a score of less than 31 on the International Socio-economic Index (ISEI) of occupational status (Ganzeboom et al. 1992), and higher class jobs had a score of greater than 70 on the same index.

The number of points achieved in the leaving cert reflects the students' final grades in the school-leaving examination. Results are codified in a "points" system, where the maximum number of points attainable is 600[13]. In this vignette design, the points dimension took three different levels: low (240 points), mid (350 points), and high (450 points). These three points correspond approximately to the 25th, 50th, and 75th percentile of the 2019 points distribution[14]. In order to make the profiles more natural, the actual value of the points was perturbed slightly – so that a school-graduate could have 238 points, 241 points, 244 points, et cetera.

Figure 3.4. Example of vignette



The number of points is the sole determinant of entry into bachelor's degree programmes[15]. Higher points are required for entry into more selective universities and more prestigious

[13] Or 625, if the student attempts the honours maths exam.
[14] http://www2.cao.ie/app_scoring/points_stats/lc19pts.pdf
[15] With the exception of some practical programmes such as art, or music, where demonstration of skills is also a requirement, and of medicine, where prospective students must sit an aptitude test.

disciplines. Hence, the number of points is a highly efficient summary of a student's academic performance and their capability of getting into university.

Names were to give the vignettes a more realistic tinge, and were gender-balanced in the vignette universe. The self-construal dimension pertained to hypotheses not relevant to the present study. This dimension was balanced across vignettes, and, while it was correlated with the outcome variable, the correlation was not dependent on the social class of the school-graduate. Hence it's omission is not problematic.

The students responded to a sequence of nine vignettes, plus one duplicate vignette placed at the end to check reliability. The experimental design and original hypotheses were pre-registered and approved by the university ethics committee[16]. However, the experiment was not originally designed to answer the research question addressed by this chapter, so the hypotheses presented here were not pre-registered.

3.4.2 The role of reference groups

The purpose of Study 2 is to see whether the reference group mechanism affects second-order status beliefs. When evaluating a given vignette, the participants are likely to be influenced by status beliefs concerning the social class of the school-graduate featured in the vignette. If they perceive the school-graduate to be lower class, they will use information encoded in status beliefs to form the expectation that their fellow classmates will think the school-graduate won't attend college. Symmetrically, when they perceive the school-graduate to be higher class, they will form the expectation that their fellow classmates will think the school-graduate won't attend college.

For the reasons already given above, they may experience some uncertainty over the social class of the school-graduate. Students will likely have an idea about approximately where another student's parents are located in the class ladder based on their occupation, but they may be unsure about exactly how lower or higher class the family is. Given this uncertainty, I argue that when a student is evaluating a given profile they will make use of a reference group heuristic, and that (part of) the reference group will consist of the school-graduate profiles shown *previous* to the profile the student is currently evaluating. As has been well-

[16] The pre-analysis plan is available here: https://osf.io/jvf2d/

established, when estimating the position of a data point in a distribution, individuals very often form the distribution from examples which are easily available to them – a phenomenon termed the "availability heuristic" (Folkes 1988). In this experiment, when a participant is engaged in reviewing a given profile, information from previous profiles will be fresh in their mind and easy to access. Therefore, I argue that previous profiles should partly constitute the reference group for a given profile.

Of course, the participants may include other examples (such as the occupation of their friends' parents) in the reference group. This is fine, as long as the composition of the previous profiles is random. Because the vignettes are randomly sampled[17], the characteristics of a given vignette are not confounded with the preceding profile, or the ones before it. So the treatment can be argued to be random.

Let the sequence of school-graduate profiles presented to participants be the set *S*, where s_i is an element of the set and the subscript *i* indexes the order of profiles[18], so that s_3 is the third profile shown, s_7 is the seventh, and so on. Let s_i be a school-graduate from a working class background and s_{i-1} be a school-graduate from a middle class background. Likewise, let s_j be a school-graduate also from a working class background and s_{j-1} be a school-graduate also from a working class background. Then the status-reference group model would predict s_i to be seen as lower class than s_j, even when they have the same objective class position, because s_{i-1} is middle class and s_{j-1} working class. In accordance with the status-reference group model, the variation in reference group should affect the (second-order) status belief, and s_i should be seen as less likely to go university. Note that, unlike in Study 1, shifts in the reference group do not confound the status effect with a "class culture" effect since the school-graduates are hypothetical and do not get a boost to academic achievement by being placed in higher social class company.

[17] Actually there is some dependency due to the sampling procedure. The vignettes are sampled without replacement from the universe, meaning there will be a type of order effect where, conditional on, e.g. seeing two higher class in a row, it is more likely that the next will be lower class. Still, the overall randomness should mean these order effects balance out – i.e. letting "H" indicate higher class and "L" indicate lower class, observing the string HHL is as likely as observing LLL.

[18] Bear in mind the participants where shown a sequence of 10 vignettes, with the 10th being a duplicate. Hence only the second to the ninth are analysed. The first is excluded since it has no preceding school-graduate profile to serve as a reference group.

Hypothesis 2.1 The social class of the preceding profile, s_{i-1}, will moderate the expectation of university attendance. If s_{i-1} is lower class, then s_i will be more likely to be expected to attend university than if s_{i-1} is higher class.

University attendance is very high in Ireland – approximately two thirds of students go onto some form of higher education[19]. Given the association between university attendance and socioeconomic background, the vast majority of Irish high class students go to university (Byrne and McCoy 2017). Hence, even if the reference group effect makes a high class school-leaver seem relatively more lower class, it is unlikely to knock them so far down the perceptual class ladder that they fall below the odds of them going to university fall below even. On the other hand, a lower class student may be hovering around this very threshold, and so will be more vulnerable to even small changes in their perceived class. This is the second-order complement to the state of affairs suggested in the first study: that students from higher up the class ladder are conditioned into a default life plan of attending university, and so relatively less vulnerable to being perceived as lower class. Hence, while the processes are different in the first and second-order cases, they are hypothesised to reinforce each other.

Hypothesis 2.2 The effect of the social class of the preceding profile, s_{i-1}, will be weaker if s_i is higher class.

The participants see a sequence of vignettes, and other profiles besides the preceding one may serve as a reference group for s_i. In fact, we can take advantage of the sequential nature of the vignette experiment in order to test Implication 2. Implication 2 theorised that the reference group effect would be moderated by the weight placed on comparisons with the particular reference group. In this case, we can think of each preceding profile as constituting its own reference "group" (like a set of one element).

Research showing that recent observations and experiences more heavily influence decision-making than less recent ones – a phenomenon termed "recency bias" (Arnold et al. 2000; Kalm and Norris 2018). Hence, I expect that when evaluating a given profile, s_i, participants

[19] https://www.irishtimes.com/news/education/sharp-rise-in-number-of-third-level-students-in-2020-report-shows-1.4764943

place more weight on comparisons with s_{i-1} than s_{i-2}. In other words, they will place more weight on comparisons to the profile they have more recently evaluated.

Hypothesis 2.3 The effect of the social class of the preceding profile, s_{i-1}, will be greater than the effect of s_{i-2}.

In order to test theses hypotheses, I run the following models:

$$y_i = \beta_0 + \beta_1 HigherClass_i + \beta_2 HigherClass_{i-1} + \beta_3 HigherClass_i * HigherClass_{i-1} + \beta_4 Points_i + \varepsilon_{ij} \quad (M2.1)$$

$$y_i = \beta_0 + \beta_1 HigherClass_i + \beta_2 HigherClass_{i-2} + \beta_3 HigherClass_i * HigherClass_{i-2} + \beta_4 Points_i + \varepsilon_{ij} \quad (M2.2)$$

The *i* subscript refers to the school-graduate in the profile, and the *j* subscript refers to the participant, within whom standard errors are clustered. Here y_i refers to the participant's belief about what their classmates expect the hypothetical school-graduate in profile *i* to do. $HigherClass$ is a binary variable that takes a value of 1 if the hypothetical school-graduate in the profile is higher class and 0 if not. *Points* is a variable that takes three levels – low, medium, and high. The model is estimated using OLS. Note that the sample sizes will not be the same for each model. In model 2.1, the first observation is discarded since it has no preceding profile. In model 2.2, the first two observations are discarded since neither has a profile that precedes its preceding profile.

3.4.3 Results

The results from models 2.1 and 2.2 are presented in Table 3.3, with marginal effects graphed in Figures 3.5 and 3.6. "Higher class" refers to the class of the profile in the observation, = 1 if higher class and = 0 otherwise. "Reference higher class" refers to the class of the reference profile. In model 2.1, the reference profile is the previous profile, s_{i-1}, and in model 2.2, the reference profile is the profile previous to the previous profile, s_{i-2}.

The analysis will be clearest if we look at the figures. Figure 3.5 shows the average marginal effects for Model 2.1. The x-axis graphs the social class of the profile being evaluated; the blue line graphs the effect when the reference profile is higher class, and the black line graphs the effect when the reference profile is lower class. As in Study 1, we find a statistically significant and negative status effect for lower class school-graduates. Unlike in Study 1, the direction of the effect is also negative for higher class school-graduates, though it is smaller,

and hence statistically significant. In Study 2 there is no countervailing, positive class culture effect to potentially counteract a negative status effect which explains why the direction of effect might be negative for higher class students as well. Hypothesis 2.1 is partially supported.

However, the difference in effect between higher and lower class school-graduates does not reach statistical significance (β_3 is not statistically significant), hence we cannot firmly conclude the evidence supports Hypothesis 2.2 (which stated that the effect of the social class of the preceding profile, s_{i-1}, will be weaker if s_i is higher class). Nonetheless, it seems that the same pattern emerges as in Study 1: a clear, negative status effect for lower class students, and a smaller status effect for higher class students which is statistically indistinguishable from zero. The negative effect for lower class individuals is also substantive: having a lower class reference group raises the probability of being expected to go to university from 0.42 to 0.57, an odds ratio of 1.8.

Let's turn to Implication 2: The effect of reference group composition will be moderated by the weight assigned by the individual to comparisons to that group. As I have already argued, given recency bias I expect that the status-reference group effect will be weaker when the reference profile is the profile *before* the previous profile, i.e the second-last profile; notationally: s_{i-2} not s_{i-1}. In Table 3.3, the coefficient in the Reference Higher Class cell in the M2.2 column shows a statistically insignificant coefficient.

Turning to Figure 3.6, the null effect is unambiguous. Confidence intervals overlap and the direction of effect has even swapped when compared to Figure 3.5. The conclusion is clear: participants put more weight on comparisons to the previous profile, s_{i-1} compared to the profile previous to the previous profile, s_{i-2}. In fact, the latter profile had no effect at all. We therefore find support for Implication 2. Appendix A3.6 additionally shows that the results are robust to the inclusion of round fixed effects.

Table 3.3. Effect of social class of reference group on (second-order) expectations of attending university.

	M2.1	M2.2
Points	0.33***	0.34***
	(0.02)	(0.02)
Higher class	0.11*	0.19***
	(0.04)	(0.04)
Reference Higher Class	-0.15***	0.05
	(0.04)	(0.05)
High class * Reference Higher Class	0.09	-0.06
	(0.07)	(0.06)
Constant	0.23***	0.13**
	(0.04)	(0.04)
Observations	664	581
R-squared	0.358	0.339

Robust standard errors in parentheses.
*** p<0.001, ** p<0.01, * p<0.05

Figure 3.5. Reference group: previous profile.

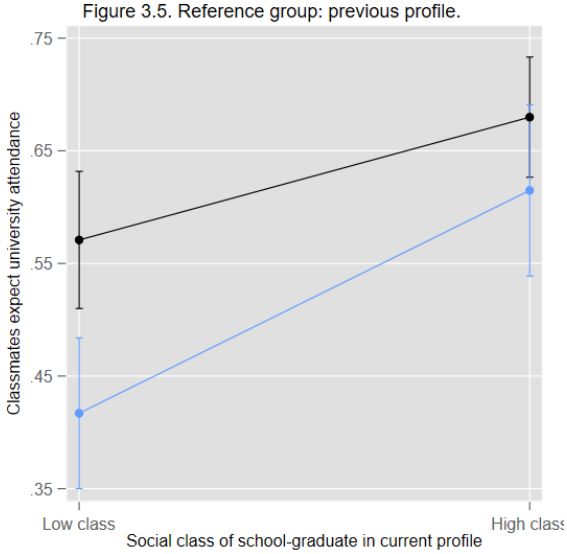
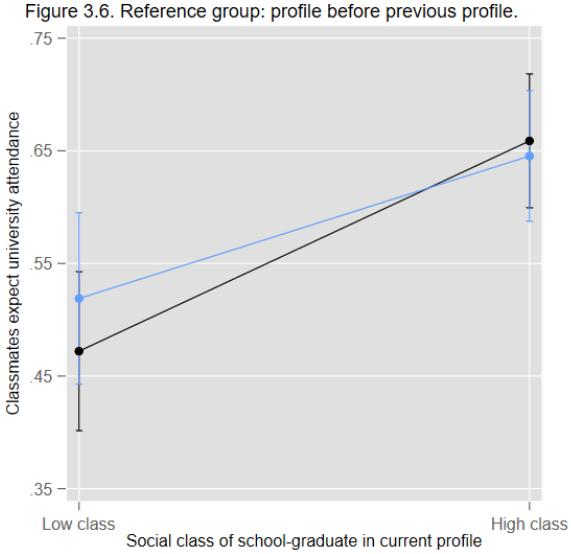
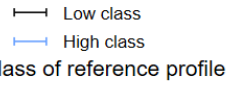


Figure 3.6. Reference group: profile before previous profile.





Social class of reference profile

3.5 Discussion and conclusion

The aim of this chapter was to see if there was evidence that status beliefs about class contribute to educational inequality. The scientific problem is that it is hard to identify variation in status beliefs because they are "attached" to objective characteristics. If both status beliefs and objective characteristics vary together, we will have a confounding problem. I argued, following (Evans and Kelley 2004), that, because class position is ambiguous, people estimate their own and others' social class by comparisons to a reference group. Hence, exogeneous variation in the reference group should cause perceived social class to vary while objective social class remains constant. Variation in perceived class should also cause variation in status beliefs: someone who is "seen" as higher class will then benefit from status-based assumptions that they are more competent; and someone seen as lower class will be disadvantaged by status-based beliefs that they are less competent. Meanwhile, since only the reference group has changed, objective characteristics remained the same.

Study 1 showed that when lower class students are placed in higher class classrooms, they are less likely to expect to go to university, as the reference group model anticipates. There is no such effect for higher class students – they do not get much of an "advantage" being placed in

lower class classrooms – perhaps because they benefit (more) from the positive "class culture" mechanism. I interpreted this clear negative effect for lower class students as the result of a status mechanism. Lower class students in higher class classrooms felt *even more* lower class because they were comparing themselves, and internalising others' comparisons, to higher class classmates.

Status beliefs are both first- and second-order beliefs. Study 1 demonstrated the existence of first-order beliefs. Study 2 provided evidence of second-order status beliefs. When a lower class school-graduate is evaluated after another lower class school-graduate, they are more likely to be expected to go to university than when the previous school-graduate is higher class. Though the direction of effect was the same for higher class school-graduates, it was smaller and not statistically significant. The natural explanation is that the previous profile is, due to availability bias, treated as a reference group for the following one. Availability bias should diminish as the exemplar becomes less available to the mind. Indeed, the social class of school-graduates two profiles previous had no effect.

The two studies complement each other. The pattern of effects is the same: lower class are disadvantaged and higher class are not, or are less so. The second study provides the greater internal validity of an experimental design, and the first study provides greater external validity. This chapter as a whole makes at least two scientific contributions. First, it contributes to the measurement of stereotypes about competence and demonstrates they have a consequential effect on inequality between classes.

Secondly, it contributes to the literature on peer effects. While the estimated "peer effect" of socioeconomic status has tended to be *either* null or positive (van Ewijk and Sleegers 2010), the theoretical assumption has tended, to be, with good reason, that the direction of effect would not be negative. Since it is probably the case that higher class students benefit from having higher class classmates, it was logical to infer that lower class students would benefit too. However, such conclusions were likely biased by the fact that most students who have higher class classmates are also themselves higher class. There was no guarantee that lower class students would benefit equally. Rather, as status theory would predict, and as the evidence presented here seems to demonstrate, lower class students are disadvantaged by being placed in higher class classrooms.

This chapter inevitably has several limitations. Neither study directly measured perceived social class; it would be beneficial to replicate the general strategy of this paper on data that included such a measure. Study 1 had to rely on certain assumptions about random assignment to classrooms which may not be fully robust. Meanwhile, Study 2 was not originally designed for the research question addressed here. Future research could implement experiments which are consciously designed to capture the effect of reference groups. For example, participants could be asked to evaluate a number of different profiles at the same time rather than a sequence of vignettes.

It could also be pointed out that status may vary for reasons other than ambiguity. Status theorists have pointed out that a given attribute – female gender, darker skin – is not always high or low status, but will vary depending on the attributes of others in the context (Correll and Ridgeway 2006). We could also imagine that, even with zero ambiguity about class position, a middle class person will be seen as of higher status when they are in a context where every else is lower class than when they are in a context where every else is higher class because status is not about absolute position – it is about relative position, not shifting perceptions. This argument could also potentially explain the results in this chapter, and would ultimately still cohere with the broad point that status matters for class inequality.

Another alternative explanation for the results of Study 1 is that they are driven by the "big fish little pond" effect (Aparicio Fenoll 2021). Class is associated with academic ability. Lower class students who are placed with higher class students will likely have a lower relative academic performance. This will harm their confidence in their academic ability, which can then explain the effect that is found. So, what I have argued to be about class may really be about ability (which is confounded with class). This is a plausible alternative explanation, and motivates the inclusion of Study 2. The status effect in Study 2 cannot be explained by a "big fish little pond" effect. If we accept Study 2 as evidence for a general stereotype about educational competence, it is then plausible to assume that at least part of the effect identified in Study 1 is driven by a status effect. Hence in the effect identified in Study 1 can be considered as an upper bound.

Despite the various limitations, I think this chapter has provided important evidence of a previously hidden channel of educational inequality.

A3.1 Algebraic model

The following simple algebraic model derives equations 1 and 2, as presented in the main paper.

Let an individual's subjective social class, S_i, be a weighted average of their objective socioeconomic location, κ_i, and their position within their reference group, ρ_i. Assume that both κ and ρ are continuous variables, and are standardised so that $\kappa_i \sim N[0,1]$, $\rho \sim [0,1]$. We can write:

$$S_i = \alpha_i \kappa_i + (1 - \alpha_i)\rho_i \qquad (A1)$$

Different individuals may give more or less weight to their objective position versus their reference group position, and this is captured by the fraction $\alpha_i \in [0,1]$.

We can write the individual's position in their reference group as:

$$\rho_i = \kappa_i - \mu_i \qquad (A2)$$

Letting $\mu_i = \frac{1}{n}\sum_{j \neq i}^{n} \kappa_j$ be the mean socioeconomic location of the individual's reference group, where *n* is the size of the reference group and κ_j indexes the objective social class of the reference group constituents. (Since ρ_i is standardised the right hand side should be divided by the standard deviation of the socioeconomic distribution of the individual's reference group. But in order to simply the notation we can assume this quantity equals 1).

Now we can represent the effect, for the individual, of being placed in a context with a new reference group whose mean socioeconomic location is $\overline{x_k}$. Before being placed in this context, the individual will have a prior reference group. For example, before going to university, a student will have a reference group consisting their secondary school friends etc. That prior group will have a mean socioeconomic location, μ_{0i}. When they go to college, they may compare themselves to a new group of college mates as well. Hence the whole reference group of old (secondary school) friends and new (college) friends will have a posterior mean, μ_{1i} which includes the new reference subgroup mean $\overline{x_k}$.

However, different people may attach differing weight to comparisons with the new reference group. At one extreme, Person A may not care about the new group at all (maybe they do not make any friends at college, and only attend online courses). Hence their posterior reference group mean socioeconomic location will be identical to their prior, i.e. $\mu_{1i} = \mu_{0i}$. At another

extreme, Person B may now only care about the new reference group and will no longer compare themselves to people outside of it. In this case $\mu_{1i} = \overline{x_k}$. Letting $w_k \in [0,1]$ be the weight placed on comparisons with the reference subgroup we can write:

$$\mu_{1i} = (1 - w_k)\mu_{0i} + (w_k)\overline{x_k} \tag{A3}$$

Then, subbing (A3) into (A2) and the resulting equation into (1) we get:

$$S_i = \kappa_i - (1 - \alpha_i)[(1 - w_k)\mu_{0i} + w_k\overline{x_k}] \tag{$A1'$}$$

Now suppose that an individual's subjective assessment of their own probability of success at university, π_i, is linearly related to their status, τ_i, which has a mean of 0 in the population, and where positive values indicate positive status beliefs and negative values indicate negative status beliefs, and the absolute value determines the strength of the beliefs. We express π_i in terms of τ_i and a vector of other factors χ_i:

$$\pi_i = \delta_i\tau_i + X_i \tag{A4}$$

Here $\delta_i > 0$ governs the strength of relationship between status and probability of success at university. We can define "probability of success at university" as the probability of achieving a degree. Now, without loss of generality, assume status is determined solely by subjective social class so that $\tau_i = \gamma_i S_i$, where $\gamma_i > 0$ governs the strength of relationship between subjective social class and status. Substituting (1') into this expression for status, and then substituting the resulting equation into (A4) we get:

$$\pi_i = \delta_i\gamma_i(\kappa_i - (1 - \alpha_i)[(1 - w_k)\mu_{oi} + w_k\overline{x_k}]) + \chi_i \tag{$A4'$}$$

This is not a pleasant-looking equation but the implied effect of an increase in $\overline{x_\iota}$ on π_i is straightforward and can be expressed by taking the derivative:

$$\frac{d\pi_i}{d\overline{x_k}} = -\delta_i\gamma_i(1 - \alpha_i)w_k < 0 \tag{A5}$$

That is to say, increasing the mean socioeconomic location of the reference subgroup decreases the subjective probability of success in university, and this relationship is moderated by δ_i, the strength of association between subjective status and subjective probability of success in university; by γ_i, the strength of association between subjective socioeconomic location and subjective status; by $(1 - \alpha_i)$, the weight given to reference

group location in forming subjective socioeconomic location; and by w_k, the weight given in comparisons by the individual to the reference subgroup.

Equation A5 is almost equivalent to equation (1) in the main paper. The difference is that there is an additional term, $(1 - \alpha_i)$, in equation A5 which is removed from the main paper for clarity's sake. The removal can be seen as equivalent to setting $\alpha_i = 1$ which, while unrealistic, does not affect the generalisability of the model.

In the first study we also consider that the mean socioeconomic location of the classroom reference group could have a positive effect.

Returning to the series of equations presented above, let $C_i = \theta_i \overline{x_k}$ be the relationship between class culture, C_i, and the mean socioeconomic position of the classroom $\overline{x_k}$, where $\theta_i > 0$ is the parameter governing the strength of the relationship. We can then insert C_i into equation A4'' to yield:

$$\pi_i = \theta_i \bar{x_i} + \delta_i \gamma_i (\kappa_i - (1 - \alpha_i)[(1 - w_k)\mu_{oi} + w_k \overline{x_k}]) + \chi_i \qquad (A4'')$$

Taking the derivative of $\overline{x_k}$ we get a new expression for (A5):

$$\frac{d\pi_i}{d\overline{x_k}} = \theta_i - \delta_i \gamma_i (1 - \alpha_i) w_k \qquad (A5')$$

This is almost identical to equation 2 in the main paper, the only difference being the inclusion of the $(1 - \alpha_i)$ term.

A3.2 Randomisation tests

I carried out randomisations test on three subsamples: those where 0% of students reported streaming ("0% sample"); those where up to 20% reported streaming ("20% sample"); and those where 80% or more reported streaming ("80% sample"). The subsamples were drawn from the English, German, and Swedish data. The Dutch data is excluded since the question about streaming was not asked in that country survey. Also, no English schools had 20% or less of students reporting streaming hence there are no English schools in the 20% or 0% subsamples (or the analytic sample).

Randomisation tests work by generating a "null distribution" of the test statistic when the key independent variable has been randomly reshuffled for each observation. By generating a probability distribution of the parameter of interest under conditions of randomness, this can

tell us how likely a test statistic is to be generated by chance. In our case we are interested in whether the assignment to treatment (classroom) is correlated with the socioeconomic composition of the classroom. In this study, socioeconomic composition is measured by the classroom's mean parental ISEI. One way to check whether assignment is correlated is to look at the distribution of ISEI between classrooms *within* schools. Take School A, which has Classroom X and Classroom Y. The classroom mean parental ISEI in Classroom X is μ_{ax} and the classroom mean parental ISEI in Classroom Y is μ_{ay}. The absolute difference between these two quantities is $\delta_a = |\mu_{ax} - \mu_{ay}|$. For a given (sub)sample *s* we can calculate the average of this quantity across all schools, which we denote by $\overline{\delta_s}$. This is our "test statistic". If this quantity tended to be very large within our sample we might suspect that there was some correlation between socioeconomic composition and classroom assignment where there was a tendency to select high class students into one classroom and low class students into another, rendering the quasi-experimental aspect of the design invalid.

But how large is "very large"? We need to have a null distribution to compare it to, and we can use the randomisation test method to generate one. It works by randomly shuffling the treatment variable – classroom assignment – and estimating $\overline{\delta_{spi}}$, the average absolute difference between-classrooms-within-schools in mean parental classroom ISEI, where the subscript *s* stands for the subsample, *p* for the fact that the quantity is attained through a randomisation, and *i* indexes the iteration of the randomisation. We repeat this process for n iterations, yielding a distribution of $\overline{\delta_{spi}}$. We can then compare our test-statistic, $\overline{\delta_s}$, to this distribution. If the test statistic lies in the outer tails of the distribution it is unlikely to have been generated by a random process. Indeed, we can apply the standard p-value approach and see if it is more extreme than 95% of the distribution, and if it is we reject the null hypothesis of randomness.

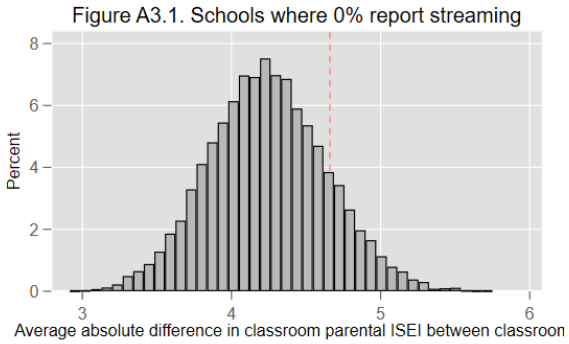
The procedure was done for the 0 percent, 20 percent, and 80 percent samples, with each randomisation test being run for 10,000 iterations. Table A3.1 below summarises, for each sample, the test statistic $\overline{\delta_s}$, the mean of the null distribution of permuted values $\overline{\delta_{spn}} = \frac{1}{n}\sum_{i=1}^{n=10,000} \overline{\delta_{spi}}$, and the proportion of permuted values greater *in absolute distance from the mean* than the absolute distance of the test statistic from the mean of the distribution – we denote this quantity by π_s, where the *s* subscript indexes the subsample. This quantity, π_s, is a measure of how extreme the test statistic is compared to the elements of the null

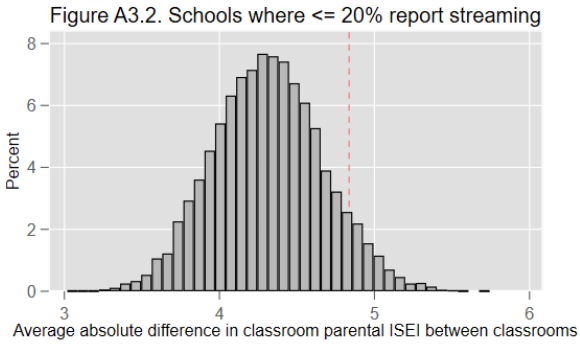
distribution, and is the equivalent of the p-value. Figures A3.1, A3.2, and A3.3 graph these distributions with the test statistics indicated by the dashed red line.

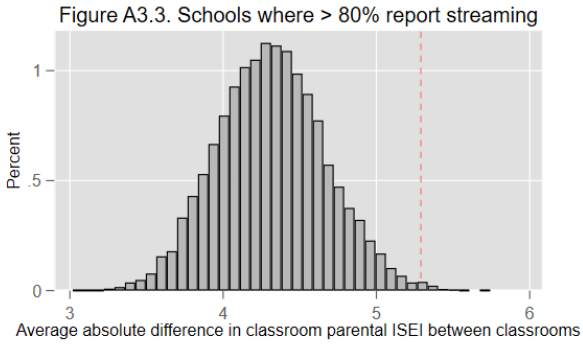
All test statistics lie to the right of their computed $\overline{\delta_{spn}}$. $\overline{\delta_0}$ is the least extreme: π_0 = 0.29. $\overline{\delta_{20}}$ is the more extreme: π_{20} = 014. Nonetheless, at conventional levels of statistical significance we would not reject the null hypotheses of randomness for either quantity. $\overline{\delta_{80}}$ is the most extreme: π_{80} = 0.01. We clearly reject the null hypothesis here: there is very likely some sort of sorting going on in this subsample.

Table A3.1 Results from the randomisation tests

Subsample	$\bar{\delta}_s$, **Test statistic,**	$\overline{\delta_{spn}}$, **Mean of permuted values**	π_s, **Proportion of permuted values more extreme than test statistic**
0 percent	4.66	4.24	0.29
20 percent	4.83	4.31	0.14
80 percent	5.29	4.6	0.01







A3.3 Results when restricting the sample to zero streaming

Table A3.2 shows the results from model 1.1 when the sample is restricted to schools where no student reported streaming. For clarity this models is referred to as A1.1.

Table A3.2. Effect of classroom socioeconomic composition on expecting to attend university, zero streaming sample.

	A1.1
Higher class student	-0.262**
	(0.088)
Mean Classroom ISEI	-0.001
	(0.002)
Higher class student* Mean Classroom ISEI	0.008**
	(0.002)
Constant	0.183
	(0.093)

Observations	2,470
R-squared	0.294

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

The results using this smaller sample size are in the same direction as in the models reported for the 20% streaming sample in the main paper, though the "main effect" of classroom ISEI (i.e. the effect when Higher class student = 0, so when the student is lower class) is not statistically significant.

A3.4 Results when excluding "don't knows"

Table A3.3 shows the results from models 1.1 when excluding those who answered "don't know" when asked what the highest level of education they expected to achieve was. For clarity this models is referred to as A1.2. The results are the same in direction as those reported in the main paper and the same pattern of statistical significance is observed.

Table A3.3. Regression output for each model when excluding "don't knows"

	A1.1
Higher class student	-0.269**
	(0.086)
Mean Classroom ISEI	-0.004*
	(0.002)
Higher class student* Mean Classroom ISEI	0.009***
	(0.002)
Constant	0.416***
	(0.078)
Observations	2,941
R-squared	0.368

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

A3.5 Rationale for making vignette outcome binary

In Study 2, the outcome variable is collapsed from four values to two. In the vignette experiment, the participants were given four options for what the school-graduate could do: begin work immediately, do a one-year post-leaving certificate course, attend an institute of technology or attend university. For the purpose of this study, these four options are collapsed into a binary: attend university or not. If a participant responded that they thought their classmates would expect the student to begin work immediately or attend a post-leaving certificate course this was coded as 0, meaning "not attend university." If a participant responded that they thought their classmates would expect the student to attend an institute of technology or a university this was coded as 1, meaning "attend university."

Substantively, there is a bigger gulf in nature of qualification between the two binary categories than there is within them. Post-leaving cert courses are generally practical and strictly work-orientated and lead to qualifications below the level of a bachelor's degree and at the same level, or just above, the leaving certificate qualification. They are in many respects an alternative, or complement, to on-the-job training rather and not geared towards providing more the abstract and theoretical knowledge provided by bachelor's degrees.

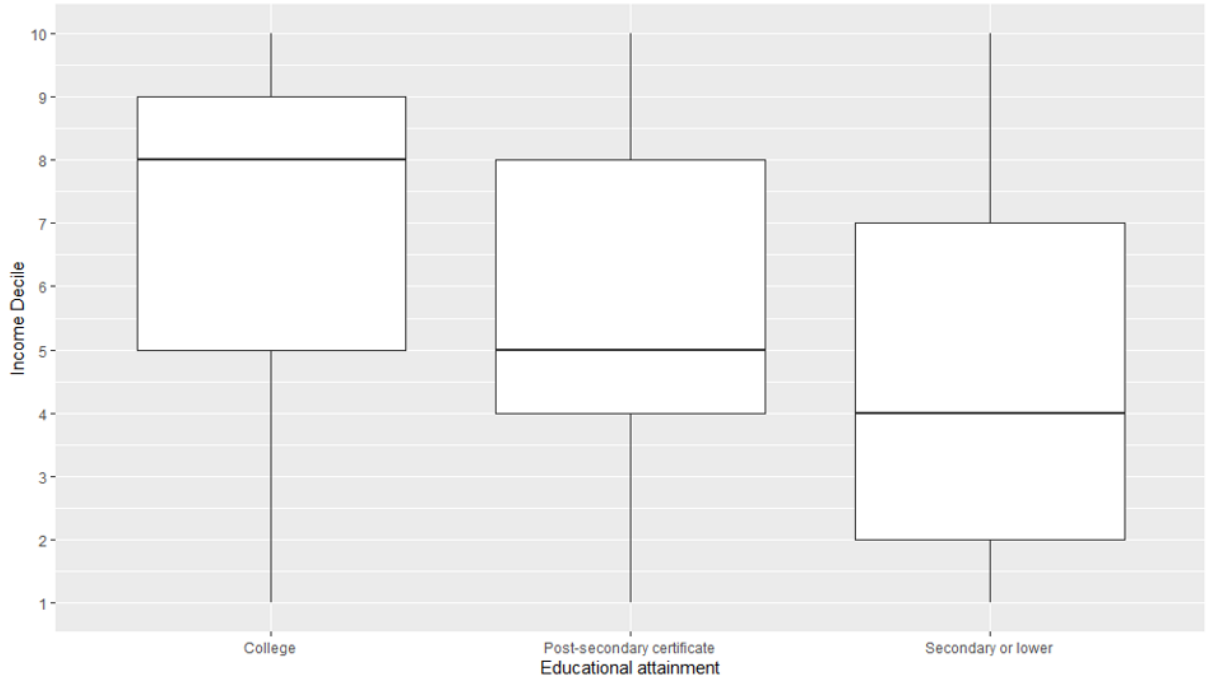
Universities and institutes of technology both offer graduate and postgraduate-level qualifications. The distinction is mainly one of historical pedigree and priority, and hence of putative prestige. Historically in an Irish context, "university" referred to an older, more selective degree-granting institution whereas institute of technology referred to a newer, less selective kind of degree-granting institution. Indeed, institutes of technology in Ireland are currently undergoing the process of being renamed as "technological universities". They are analogous to the British "red brick" universities, which often developed from polytechnical institutes to full-blown universities, granting degrees in both the sciences and humanities.

On an empirical level, there is a significant bunching of "life chances" within the binary categories. Figure A1 displays data from the Irish component of the 2018 round of the European Social Survey, with a boxplot of household income decile by educational category. There is no information on whether the respondent attended either a university versus institute of technology; just on their level of educational qualification. Hence all those with bachelor's or higher degrees are categorised under "college". I use this agnostic term to avoid confusion with the specific meaning of university in an Irish context.

The data is from a sample restricted to under-45s, to account for confounding of age with cohort probability of attaining tertiary education (much fewer people went to college in the 1980s and before). As can be seen, those with a college degree (undergraduate or higher) earn much more than those with a post-leaving cert qualification ("post-secondary certificate") or those who only have a secondary education or lower, and the median deciles for the latter two are much closer to each other than they are for to the median decile of the college-educated category. Hence, uniting the post-secondary and secondary-or-lower categories appears reasonably justified, at least in terms of expected income.

Unfortunately we cannot distinguish between institute of technology and "university" graduates in this data. Data from the Irish Higher Education Authority indicates that institute of technology graduates do earn about 14 percent less than university graduates four years after graduation[20]. This is a significant though not particularly large gap, indicating that expected income for institute of technology graduates is still much closer to university graduates than to post-secondary graduates.

Figure A4. Boxplot of income decile for educational category. (ESS round 9, Irish data).



[20] See page 10 here: https://hea.ie/assets/uploads/2019/12/Higher-Education-Earnings-Report-Dec-19.pdf

A3.6 Adjusting vignette results for rounds

Table A3.4. Regression output from Models 2.1 and 2.2 with fixed effects for round

	A2.1	**A2.2**
Points	0.334***	0.338***
	(0.022)	(0.024)
Higher class	0.113*	0.195***
	(0.043)	(0.046)
Reference Higher Class	-0.155***	0.049
	(0.044)	(0.052)
High class * Reference Higher Class	0.091	-0.067
	(0.066)	(0.061)
Round number 3	0.074	
	(0.065)	
Round number 4	-0.044	-0.123
	(0.065)	(0.068)
Round number 5	0.052	-0.034
	(0.060)	(0.059)
Round number 6	0.066	-0.008
	(0.055)	(0.058)
Round number 7	0.138*	0.052
	(0.062)	(0.065)
Round number 8	0.094	0.024
	(0.065)	(0.058)
Round number 9	0.077	-0.004
	(0.058)	(0.067)
Constant	0.174**	0.141*
	(0.053)	(0.062)
Observations	664	581
R-squared	0.370	0.350

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Table A3.5 presents results from the a modified version of models 2.1 and 2.2 from the main paper, including fixed effects for the round in which the participant made their choice. As can be seen, the results remain substantively the same. In model A2.1, Round number 2 is the reference category and hence is omitted. In model A2.2, Round number 3 is the reference category (since round number 2 is not in the subsample) and hence is omitted.

References

Ammermüller, Andreas, and Jorn-Steffen Pischke. 2006. *Peer Effects in European Primary Schools: Evidence from PIRLS. ZEW Discussion Paper*. 06–027. ZEW - Leibniz Centre for European Economic Research.

Aparicio Fenoll, Ainoa. 2021. 'The Best in the Class'. *Economics of Education Review* 84:102168. doi: 10.1016/j.econedurev.2021.102168.

Arnold, Vicky, Philip A. Collier, Stewart A. Leech, and Steve G. Sutton. 2000. 'The Effect of Experience and Complexity on Order and Recency Bias in Decision Making by Professional Accountants'. *Accounting & Finance* 40(2):109–34. doi: 10.1111/1467-629X.00039.

Barone, Carlo, Giulia Assirelli, Giovanni Abbiati, Gianluca Argentin, and Deborah De Luca. 2018. 'Social Origins, Relative Risk Aversion and Track Choice: A Field Experiment on the Role of Information Biases'. *Acta Sociologica* 61(4):441–59. doi: 10.1177/0001699317729872.

Becker, Rolf, and Anna E. Hecken. 2009. 'Why Are Working-Class Children Diverted from Universities?—An Empirical Assessment of the Diversion Thesis'. *European Sociological Review* 25(2):233–50. doi: 10.1093/esr/jcn039.

Bicchieri, Cristina. 2017. *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. New York: Oxford University Press.

Boliver, Vikki. 2013. 'How Fair Is Access to More Prestigious UK Universities?' *British Journal of Sociology* 64(2):334–64.

Breen, Richard, and John H. Goldthorpe. 1997. 'Explaining Educational Differentials: Towards a Formal Rational Action Theory'. *Rationality and Society* 9(3):275–305. doi: 10.1177/104346397009003002.

Byrne, Delma, and Selina McCoy. 2017. 'Effectively Maintained Inequality in Educational Transitions in the Republic of Ireland'. *American Behavioral Scientist* 61(1):49–73. doi: 10.1177/0002764216682991.

Cansunar, Asli. 2021. 'Who Is High Income, Anyway? Social Comparison, Subjective Group Identification, and Preferences over Progressive Taxation'. *The Journal of Politics* 83(4):1292–1306. doi: 10.1086/711627.

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederick D. Weinfeld, and Robert L. York. 1966. *Equality of Educational Opportunity*. Washington, DC.: US Department of Health, Education, and Welfare.

Correll, Shelley J. 2001. 'Gender and the Career Choice Process: The Role of Biased Self-Assessments'. *American Journal of Sociology* 106(6):1691–1730. doi: 10.1086/321299.

Correll, Shelley J. 2004. 'Constraints into Preferences: Gender, Status, and Emerging Career Aspirations'. *American Sociological Review* 69(1):93–113. doi: 10.1177/000312240406900106.

Correll, Shelley J., and Cecilia L. Ridgeway. 2006. 'Expectation States Theory'. Pp. 29–51 in *Handbook of social psychology*. Springer.

Correll, Shelley J., Cecilia L. Ridgeway, Ezra W. Zuckerman, Sharon Jank, Sara Jordan-Bloch, and Sandra Nakagawa. 2017. 'It's the Conventional Thought That Counts: How Third-Order Inference Produces Status Advantage'. *American Sociological Review* 82(2):297–327. doi: 10.1177/0003122417691503.

Crocker, Jennifer, Leigh L. Thompson, Kathleen M. McGraw, and Cindy Ingerman. 1987. 'Downward Comparison, Prejudice, and Evaluations of Others: Effects of Self-Esteem and Threat.' *Journal of Personality and Social Psychology* 52(5):907–16.

Croizet, Jean-Claude, Michel Désert, Marion Dutrévis, and Jacques-Philippe Leyens. 2001. 'Stereotype Threat, Social Class, Gender, and Academic under-Achievement: When Our Reputation Catches up to Us and Takes Over'. *Social Psychology of Education: An International Journal* 4(3–4):295–310. doi: 10.1023/A:1011336821053.

Curtis, Josh. 2016. 'Social Mobility and Class Identity: The Role of Economic Conditions in 33 Societies, 1999–2009'. *European Sociological Review* 32(1):108–21. doi: 10.1093/esr/jcv077.

Dineen, Jennifer, Mark D. Robbins, and Bill Simonsen. 2019. 'Social Class: Perception and Reality'. *International Journal of Public Administration* 42(1):55–65. doi: 10.1080/01900692.2017.1399419.

Durante, Federica, and Susan T. Fiske. 2017. 'How Social-Class Stereotypes Maintain Inequality'. *Current Opinion in Psychology* 18:43–48. doi: 10.1016/j.copsyc.2017.07.033.

Evans, Mariah D. R., and Jonathan Kelley. 2004. ‘Subjective Social Location: Data From 21 Nations’. *International Journal of Public Opinion Research* 16(1):3–38. doi: 10.1093/ijpor/16.1.3.

van Ewijk, Reyn, and Peter Sleegers. 2010. ‘The Effect of Peer Socioeconomic Status on Student Achievement: A Meta-Analysis’. *Educational Research Review* 5(2):134–50. doi: 10.1016/j.edurev.2010.02.001.

Folkes, Valerie S. 1988. ‘The Availability Heuristic and Perceived Risk’. *Journal of Consumer Research* 15(1):13–23.

Friend, Ronald M., and Joel Gilbert. 1973. ‘Threat and Fear of Negative Evaluation as Determinants of Locus of Social Comparison.’ *Journal of Personality*.

Ganzeboom, Harry B. G., Paul M. De Graaf, and Donald J. Treiman. 1992. ‘A Standard International Socio-Economic Index of Occupational Status’. *Social Science Research* 21(1):1–56. doi: 10.1016/0049-089X(92)90017-B.

Gewirtz, Paul. 1996. ‘On “I Know It When I See It”’. *The Yale Law Journal* 105(4):1023–47. doi: 10.2307/797245.

Goldthorpe, John H., and Abigail McKnight. 2006. ‘The Economic Basis of Social Class’. *Mobility and Inequality: Frontiers of Research in Sociology and Economics* 109–36.

Gorski, Paul C. 2012. ‘Perceiving the Problem of Poverty and Schooling: Deconstructing the Class Stereotypes That Mis-Shape Education Practice and Policy’. *Equity & Excellence in Education* 45(2):302–19. doi: 10.1080/10665684.2012.666934.

Hochschild, Jennifer L. 2021. ‘Social Class in Public Schools’. *Journal of Social Issues* 59(4).

Hyman, Herbert H. 1960. ‘Reflections on Reference Groups’. *Public Opinion Quarterly* 24(3):383–96. doi: 10.1086/266959.

Jury, Mickaël, Annique Smeding, Nicole M. Stephens, Jessica E. Nelson, Cristina Aelenei, and Céline Darnon. 2017. ‘The Experience of Low-SES Students in Higher Education: Psychological Barriers to Success and Interventions to Reduce Social-Class Inequality’. *Journal of Social Issues* 73(1):23–41. doi: 10.1111/josi.12202.

Kalm, Kristjan, and Dennis Norris. 2018. ‘Visual Recency Bias Is Explained by a Mixture Model of Internal Representations’. *Journal of Vision* 18(7):1–1.

Legewie, Joscha, and Thomas A. DiPrete. 2012. ‘School Context and the Gender Gap in Educational Achievement’. *American Sociological Review* 77(3):463–85. doi: 10.1177/0003122412440802.

McEwan, Patrick J. 2003. ‘Peer Effects on Student Achievement: Evidence from Chile’. *Economics of Education Review* 22(2):131–41.

McVicar, Duncan, and Arnold Polanski. 2014. ‘Peer Effects in UK Adolescent Substance Use: Never Mind the Classmates?’ *Oxford Bulletin of Economics and Statistics* 76(4):589–604. doi: 10.1111/obes.12030.

Merton, Robert K. 1968. *Social Theory and Social Structure*. New York: Free Press.

Morse, Stan, and Kenneth J. Gergen. 1970. ‘Social Comparison, Self-Consistency, and the Concept of Self.’ *Journal of Personality and Social Psychology* 16(1):148–56.

Oesch, Daniel, and Nathalie Vigna. 2021. ‘A Decline in the Social Status of the Working Class? Conflicting Evidence for 8 Western Countries, 1987–2017’. *Comparative Political Studies* 00104140211047400.

Redford, Jeremy, Jennifer A. Johnson, and Julie Honnold. 2009. ‘Parenting Practices, Cultural Capital and Educational Outcomes: The Effects of Concerted Cultivation on Academic Achievement’. *Race, Gender & Class* 16(1/2):25–44.

Ridgeway, Cecilia L. 2019. *Status*. New York: Russell Sage.

Ridgeway, Cecilia L., and Susan R. Fisk. 2012. ‘Class Rules, Status Dynamics, and “Gateway” Interactions’. Pp. 131–51 in *Facing social class: How societal rank influences interaction*. New York, NY, US: Russell Sage Foundation.

Rivera, Lauren A. 2011. ‘Ivies, Extracurriculars, and Exclusion: Elite Employers’ Use of Educational Credentials’. *Research in Social Stratification and Mobility* 29(1):71–90. doi: 10.1016/j.rssm.2010.12.001.

Rivera, Lauren A. 2015. ‘Class, Cognition, and Face-to-Face Interaction’. Pp. 1–15 in *Emerging Trends in the Social and Behavioral Sciences*, edited by R. A. Scott and S. M. Kosslyn. New York: John Wiley & Sons, Ltd.

Robert, Peter. 2010. ‘Social Origin, School Choice, and Student Performance’. *Educational Research and Evaluation* 16(2):107–29.

Rumberger, Russel W., and Gregory J. Palardy. 2005. ‘Does Segregation Still Matter? The Impact of Student Composition on Academic Achievement in High Schoo’. *Teachers College Record* 107:1999–2045.

Senik, Claudia. 2004. ‘When Information Dominates Comparison: Learning from Russian Subjective Panel Data’. *Journal of Public Economics* 88(9–10):2099–2123.

Shibutani, Tamotsu. 1955. ‘Reference Groups as Perspectives’. *American Journal of Sociology* 60:562–69. doi: 10.1086/221630.

Sosnaud, Benjamin, David Brady, and Steven M. Frenk. 2013. ‘Class in Name Only: Subjective Class Identity, Objective Class Position, and Vote Choice in American Presidential Elections.’ *Social Problems* 60(1):81–99.

Spencer, Bettina, and Emanuele Castano. 2007. ‘Social Class Is Dead. Long Live Social Class! Stereotype Threat among Low Socioeconomic Status Individuals’. *Social Justice Research* 20(4):418–32. doi: 10.1007/s11211-007-0047-7.

Treiman, Donald J. 1977. *Occupational Prestige in Comparative Perspective*. New Yprk: Free Press.

Troyer, Lisa, and C. Wesley Younts. 1997. ‘Whose Expectations Matter? The Relative Power of First- and Second-Order Expectations in Determining Social Influence’. *American Journal of Sociology* 103(3):692–732. doi: 10.1086/231253.

Wouters, Sofie, Hilde Colpin, Jan Van Damme, Steven De Laet, and Karine Verschueren. 2013. ‘Early Adolescents’ Academic Self-Concept Formation: Do Classmates or Friends Matter Most?’ *Learning and Individual Differences* 27:193–200. doi: 10.1016/j.lindif.2013.09.002.

Wright, Erik O. 2015. *Understanding Class*. New York: Verso.

Chapter 4. Worse things happen at sea. Testing a cognitive dissonance theory of belief in meritocracy.

Abstract

People are more likely to view income inequality as legitimate when it is based on differences of merit. And in recent years, even as income inequality has increased in developed countries, so has the proportion of people who believe income is distributed meritocratically. But people have only imperfect information over the distribution of income in their society. Moreover, belief in meritocracy is higher in countries which are less socially mobile. This suggests belief in meritocracy is not a straightforward recognition of objective social processes, but may in fact be the result of a cognitive bias. I formulate an argument, based on the work of Jon Elster, that belief in meritocracy is a coping mechanism for reducing cognitive dissonance. I theorise that people can follow two strategies to reduce cognitive dissonance arising from inequality. If they have control over the income distribution, they can take action to reduce inequality. If they lack control, they can try to convince themselves the distribution is fair (because it is meritocratic). To test this argument, I design and implement an online experiment (n = 351) where participants are exposed to varying degrees of inequality, and are given varying degrees of control over the income distribution. Analysis of quantitative and qualitative data does not support the cognitive dissonance mechanism as an explanation for belief in meritocracy. However, the results show that when lacking agency over the income distribution, people are more likely to regard the distribution as fair.

4.1 Introduction

This chapter addresses RQ3: How does the "class = competence" belief arise?

Expressing the view that "class = competence" entails the belief that income is distributed meritocratically – as high class people, who earn more, are also more competent. Most people in developed countries believe that their societies are meritocratic, and the proportion who believe this has grown over the past few decades (Mijs 2021). The growing belief in meritocracy helps explain why growing income inequality in developed countries (Nolan, Richiardi, and Valenzuela 2019) has largely been tolerated, since people tend to view economic inequality as legitimate when they think it is based on distinctions of merit (Ahrens 2020; Alesina and Angeletos 2005; Hufe et al. 2022).

This widespread belief that society is meritocratic is puzzling for a number of reasons. First, expressing the view that their society is meritocratic requires that a person make a claim about the correlation between the distribution of income and the distribution of merit. But, as is well known, people have very imperfect knowledge of the income distribution in their country (Cansunar 2021); so how can they reliably estimate this correlation? Second, belief in meritocracy is greatest where inequality is highest (Mijs 2021), but high inequality is also correlated with higher structural barriers to individual mobility (Hufe et al. 2022). Hence the cross-country pattern of belief in meritocracy appears to vary inversely with the actual degree of meritocracy. Thirdly, endorsing the belief also requires that working class people "accept" they are incompetent and deserve less – why would people voluntarily endorse a belief, on the basis of imperfect evidence, that derogates themselves?

I suggest that the story about inequality and meritocracy has got things the wrong way round. It is not that people observe the degree of meritocracy in society and are duly satisfied with inequality if they judge the rich to be sufficiently more meritorious than the poor. Rather, I argue, people wish to be satisfied about the degree of inequality, and so, under certain circumstances, they convince themselves, whatever the evidence, that society is meritocratic. Following Elster (1985), I argue that the desire to reduce cognitive dissonance motivates people to believe they live in a meritocracy. Cognitive dissonance arises when an individual's observation that their society is economically unequal clashes with the desire to believe that their society is just. When people cannot change the distribution of income, they reduce their

dissonance by engaging in a process of wishful thinking: they convince themselves that their society is meritocratic, and therefore the degree of inequality they observe is legitimate.

I design and implement an online experiment to test this theory. The results do not support the cognitive dissonance mechanism as an explanation for belief in meritocracy. However, the results show that when lacking agency over the income distribution, people are more likely to regard the distribution as fair. Analysis of qualitative data suggests that when people are deprived of agency they refrain from moral evaluations and adopt a fatalistic attitude towards human behaviour.

4.2 The tolerable tyranny of merit?

The concept of a "meritocracy" refers to a society where the distribution of economic rewards is based on individual merit. Young, who coined the term "meritocracy" (1961), defined merit as "IQ + effort". In contemporary usage, merit is usually defined as a combination of talent, skill and hard work (Ahrens 2020; Heiserman and Simpson 2017; Mijs 2021). Hence, if someone believes that their society is meritocratic, they believe that higher class people are generally more hard working and more talented than lower class people; i.e they are expressing the view that "class = competence". As Mijs reports (2021), a (large) majority of people in developed countries report this view, and this proportion has increased over time.

The growing popularity of this belief may help explain why growing inequality has also been largely tolerated – people tend to view inequality as legitimate when it is based on distinctions of merit (Ahrens 2020; Alesina and Angeletos 2005; Hufe et al. 2022). This result might have surprised Young himself, who assumed working class people would resent the idea that they "deserve" less because they are inherently inferior. The narrator of his book warns that: "For more than half a century, the lower classes have been harbouring resentments which they could not make articulate, until the present day" (Young 1961:188).

Young's intuition provides us with one reason why widespread belief in meritocracy is puzzling. For a working class person to express the view that society is meritocratic, they must therefore endorse the view that (i) they are relatively incompetent; and (ii) their relative lack of income and wealth – and associated difficulties, burdens, and insecurities – are justified by their incompetence. Why should people endorse a view that derogates themselves, and why should such a derogatory view calm potential anger at economic inequality?

It is not just Young who tended to believe meritocracy would generate dissatisfaction. Some of the most prominent social theorists who lived through the birth pangs of capitalism and liberal democracy, tended to believe that the lower classes would chafe at the idea of meritocracy. Durkheim believed that "the very development of industry and the almost infinite extension of the market" had accompanied a dissolution of the traditional norms which had placed "natural" limits on the desire to accumulate wealth (Durkheim 1952: 216). In absence of these norms, he doubted whether the lower classes would be satisfied by the idea that the rich deserved to have more than them on the basis of superior ability. Tocqueville, in his famous analysis of American democracy, expressed similar concerns (de Tocqueville 2007). He argued that while democracy was predicated on the notion of formal equality between all, the operations of the capitalist market would lay bare inequality of ability to an intolerable degree. For the lower classes "there is no kind of superiority, however legitimate it may be, which is not irksome in their sight" (de Tocqueville 2007).

Maybe working class people are simply acknowledging, however reluctantly, an obvious truth that the distribution of income indeed corresponds to the distribution of merit? However, this view is doubtful for at least two reasons. Firstly, it is unlikely that people possess the information required to accurately make such a judgment. It is well-known that people have only imperfect knowledge of the distribution of income in their society (Bredemeier 2014; Cansunar 2021). It also stands to reason that people have only imperfect information on the distribution of "merit" in their society. Economists, for example, have long pointed out that "true" ability is very difficult to fully measure (Belzil and Hansen 2002; Blackburn and Neumark 1992; Taber 2001). It therefore seems unlikely that most individuals are able to make an accurate assessment of the correlation between competence and income.

Secondly, the pattern of cross-country support for the view that society is meritocratic is at odds with objective cross-country measures of meritocracy. Meritocracy implies that every career is "open to talent", and that occupational attainment should be based on ability and effort not on social origin (Brown 2013; Whelan and Layte 2004). Hence, if belief in meritocracy merely tracks the objective degree of meritocracy in society, we should expect that belief in meritocracy is highest in countries that have the least structural barriers to individual social mobility. But this is not the case. Mijs (2021) finds that support for meritocracy is greater in more unequal countries, even though more unequal countries present greater structural barriers to individual mobility (Durlauf and Seshadri 2018; Hufe et al.

2022; Jerrim and Macmillan 2015). Greater belief in meritocracy does not mirror objective evidence for where meritocracy is greatest.

In short, it is unlikely that belief in meritocracy is simply an accurate observation of social reality. We cannot accept a simple story that people first neutrally observe the degree of meritocracy in their country and, on that basis, decide how satisfied they are with the degree of inequality. As others have already suggested (Heiserman and Simpson 2017), the process of inference may run in reverse. In this chapter, I argue that people wish to feel satisfied about the degree of inequality and, under certain conditions, they convince themselves that income is distributed meritocratically to achieve that satisfaction. In other words, belief in meritocracy is a type of wishful thinking motivated by the need to reduce the dissonance caused by inequality.

4.3 Wishful thinking and cognitive dissonance

Elster has adapted the fable of the Fox and the Sour Grapes to explain how wishful thinking can lead oppressed people to autonomously legitimate their own oppression (1985). The Sour Grapes fable tells the story of a fox who wishes to eat a bunch of grapes off a vine. But the vine's branches are tall, and no matter how high the fox jumps it can't reach the grapes. Frustrated and dismayed, the fox declares to itself that the grapes were no doubt unripe, and so it would not have enjoyed eating them anyway.

Elster argues that the "wishful thinking" the fox displays in convincing itself the grapes were sour could explain why disadvantaged social groups can convince themselves the structure which disadvantages them is legitimate, even without any external ideological or coercive pressure. The fox engages in wishful thinking because it is unable to reach the grapes. Likewise, when oppressed people cannot change the structure that oppresses them, they may soothe the pang of injustice if they convince themselves the structure is actually just after all. The existence of extant ideological frames (such as meritocracy or, in an earlier epoch perhaps, divine sanction) facilitates this process by providing readymade conceptual schemata for the oppressed person to adapt into their worldview. The soothing effect of these frames has been termed the "palliative function of ideology" (Jost and Hunyady 2003).

In formulating this explanation, Elster explicitly utilises the concept of "cognitive dissonance" to summarise the sense of frustration, discomfort, and unease which individuals feel when confronted with an unjust situation. He suggests that the desire to reduce cognitive

dissonance provides an explanation for why oppressed or disadvantaged people might wish to justify their own oppression – an explanation also put forward by system justification theorists (Jost and Hunyady 2003).

The term "cognitive dissonance" was coined by the social psychologist Leon Festinger (Festinger 1962). People experience cognitive dissonance when they hold two beliefs which are in contradiction with each other. For example, let us say that I believe P: "My mother is in Kilkenny". Then, one day I see a woman in Florence who bears a striking resemblance to my mother, and gives rise to the belief Q: "I see a woman in the Mercato Centrale who looks like my mother." Q implies not-P: "My mother is not Kilkenny", and therefore contradicts P. Hence, I experience dissonance.

Using the term "belief" suggests a binary attitude towards a proposition: I either do or do not believe that my mother is in Florence. However, our attitude towards a proposition is often not one of either total belief or unbelief. This is an important point because uncertainty about beliefs plays a crucial role in the theory presented here. Philosophers use the term "credence" to capture the fact that we do not need to express absolute belief or disbelief in a proposition, but can express varying degrees of confidence in its truth (Jackson 2020). In Bayesian epistemology, credences place a person's attitude towards a proposition on the [0,1] interval scale, where 0 indicates certainty that the proposition is false and 1 indicates certainty that it is true. Credences can be treated as a function, $c(.)$, which takes a proposition as an argument and maps it onto [0,1]. For example, if R = "It will rain tomorrow", and I am very, though not absolutely, confident that it will rain, my credence can be represented as $c(R) = 0.9$.

Credences can help us neatly summarise cognitive dissonance in a simple formula. Let P = "My mother is in Kilkenny" and Q = "My mother is in Florence". Since I am pretty confident of P, we can say $c(P) = 0.9$, and since I am not so sure the woman in the Mercato Centrale is my mother, we can say $c(Q) = 0.6$. Now let us represent my cognitive dissonance, δ, by the product of both credences: $\delta = c(P) * c(Q)$. The value of δ will lie on the [0,1] interval, with 0 representing minimum dissonance and 1 representing maximum dissonance. In the case of my mother, $\delta = 0.9 * 0.6 = 0.54$.

4.4 Dissonance reduction strategies

Festinger (1962) reasoned that, since dissonance generates negative affect, people will adopt strategies to mitigate it. If $\delta = c(P) * c(Q) > 0$, an individual would ideally like to reduce

either $c(P)$ or $c(Q)$ to zero. I assume that there are broadly two ways in which they can do this: (i) an "action-based" dissonance reduction strategy; and (ii) a "belief-based" dissonance reduction strategy. In the first approach, a person suffering cognitive dissonance can take action to change the state of affairs which gave rise to the dissonance. For example, I could approach the woman in Mercato Centrale and verify whether she really is, as she appears, my mother. If she is, then I set $c(P)$ to zero, and if not I set $c(Q)$ to zero. Either way, my dissonance is reduced to zero.

In the second approach, a person suffering cognitive dissonance can engage in a process of motivated reasoning to change their beliefs. This is the strategy chosen by the fox in the sour grapes fable. I convince myself that the woman I saw must be someone else… after all, I only saw her for a moment and she was wearing a hat that partially concealed her face…

That my credence in the dissonance-inducing proposition is neither 0 nor 1 is a necessary condition for wishful thinking. Elster points out that "wishful thinking" should be distinguished from "self-deception" (1985). While self-deception involves a necessary contradiction between what one professes to believe and the belief one is privately compelled to hold by the evidence, wishful thinking can lead to support for propositions which are at least consistent with, or (partially) supported by the evidence. Because I do *not* fully believe in P, there must be some data or argument which mitigates against full credence. I can therefore seize upon and expand this counter-rationale while ignoring the evidence in favour of P. In essence, wishful thinking is a subspecies of motivated reasoning: a cognitive process which is selective or one-sided, arbitrarily seeking out and preferring evidence in favour of the outcome that the subject wishes to be true (Ellis 2022; Epley and Gilovich 2016).

If uncertainty is the first necessary condition for wishful thinking to take place, the second is emotion. Wishful thinking functions metabolically, to convert a negative affect (dissonance) into a positive one (relief, or satisfaction). Emotions are commonly present in the reasoning process – people talking about "gut instinct" or choosing what "feels right" – and it has even been argued that they are a prerequisite for forming beliefs under uncertainty, since without emotion people would remain paralysed between alternatives with uncertain outcomes, unable to decide what to believe and what to do (Damasio 1995).

But while emotions can generate the epistemic propulsion necessary to take action under uncertainty, they can also distort the reasoning process so that the individual arrives at the

answer that feels satisfactory rather than the one that is justified (Ellis 2022; Epley and Gilovich 2016). The sense of being allotted an unfair amount, or of being looked down upon, are two of the primary negative affects which can be engendered by inequality. When you factor in uncertainty over the causes of inequality, the situation is ripe for wishful thinking.

4.5 Dissonance reduction and meritocracy

When observing unequal outcomes in their society, working class people may form the following belief U = "Income is unjustly distributed (in a way that disadvantages me)". It is likely that people do not have full credence towards this statement – i.e. $0 < c(U) < 1$. Uncertainty towards U can arise because individuals do not have perfect information on the distribution of income, or on the distribution of merit, and they may also not have a fully developed normative schema which would allow them to confidently evaluate the justness or unjustness of the distribution.

A positive credence in proposition U may then give rise to cognitive dissonance if it contradicts another positive credence such as the proposition J = "Society is just". Lerner (1980) provided the foundational evidence that people have a strong desire to believe they live in a just society, or "just world" as he put it. Subsequently, belief in a just world has been shown to motivate victim blaming, even in the case of serious crime (Correia and Vala 2003; Russell and Hand 2017), and to motivate low status groups to ignore discrimination against themselves (Choma et al. 2012; Lipkus and Siegler 1993). I argue that the belief that society is just is strong, so that $c(J)$, should be fairly close to 1. There are many reasons why people have strong motivations to believe their society is just. In essence, believing that the world is just enhances well-being and self-esteem, which is explains why it is associated with subjective well-being even when adjusting for subjective health, social contacts, and education (Dzuka and Dalbert 2006)

Believing that society is just dampen. s negative emotions that result from negative outcomes (Hafer and Correy 1999), and it may serve to encourage effort and self-belief in the contexts of uncertainty concerning the return to effort (Bénabou and Tirole 2006). Most relevant to the argument presented here, members of disadvantaged groups who believe, despite their objective position, that society is just tend to benefit from higher self-esteem and lower negative affect than those of them who believe they are unjustly treated (Dalbert 2002; Liu et al. 2021; Major et al. 2007; Schaafsma 2013). Hence, oppressed groups are incentivised to

believe the world is just in order to buffer their self-esteem. Nonetheless, the belief that society is just is likely to vary across populations – in areas where society is more obviously brutal, oppressive, and exploitative, such a belief will be less tenable. The population sampled in this data comes from Britain which, while hardly a paradise, is relatively free of overt oppression, brutality, and exploitation by the state or economics elites.

U and J seem to contradict each other, and hence can give rise to cognitive dissonance: $\delta = c(U) * c(J) > 0$. Such dissonance cannot easily be dismissed if people attach importance to general considerations of justness in society, as well as particular considerations as to whether they have been given a "fair deal". Furthermore, in societies that experience higher inequality $c(U)$ should be closer to 1, and hence δ should be greater – an implication which suggests cognitive dissonance could explain the positive association between belief in meritocracy and income inequality (Mijs 2021).

If $\delta = c(U) * c(J) > 0$, then individuals will seek to reduce their dissonance by reducing their credence in either of the beliefs. I assume that people follow the path of least resistance and aim to reduce their credence in the proposition they already have least credence in. I have already argued that $c(J)$ should be fairly close to 1. On the other hand, $c(U)$ is likely to be lower since people tend to have fairly imperfect information on the distribution of income in society, not to mention they may experience some uncertainty over the normative aspects of proposition U. Hence, assuming that $c(J) > c(U)$, people will try to reduce their credence in U.

As I have already suggested, people may follow an "action-based" dissonance-reduction strategy or a "belief-based" dissonance-reduction strategy. First considering the action-based strategy, we can further distinguish between individualistic and collectivistic approaches. An individualistic approach could consist in a concerted effort to achieve individual social mobility. After arriving at a position at the top of the income distribution, an individual may be less worried about income inequality and hence experience less dissonance. A collectivistic solution could be achieved either through trade union or political activity. Trade union activism would entail agitation for higher wages in one's own workplace or industry; political activity, including the simple act of voting, would entail voting for a party that promises to redistribute income.

However, an action-based strategy might not be feasible. For example, the individual may live in a rigid, socially immobile society which would make it difficult to climb up the ladder. On the collective end of things, it may be the case that union density is low and none of the main political parties support redistribution. In this case, the individual would have recourse only to a belief-based strategy. They would engage in a process of motivated reasoning, identifying some plausible rationale to convince themselves $c(U)$ is at or close to 0. To this end, it would be natural for them to seize on the prevalent cultural schema of meritocracy and, without any totally compelling evidence to the contrary, to convince themselves that income correlates strongly with merit. Since people are more likely to view economic inequality as legitimate if derived from inequality in merit (Ahrens 2020; Alesina and Angeletos 2005; Hufe et al. 2022), believing that society is meritocratic would push $c(U)$ towards 0 and minimise $\delta = c(U) * c(J)$.

In sum, a non-rich person living in an unequal society will, at some point in time, tend to have a positive credence towards two conflicting propositions: U = "It appears that income is unjustly distributed (in a way that disadvantages me)", and J = "Society is just". They will therefore experience dissonance in proportion to the product of their credence in both propositions, $\delta = c(U) * c(J)$. If $c(J)$ is stronger, they will focus their attention on trying to reduce $c(U)$. In principle they have two dissonance-reduction strategies available to them: (i) an action-based strategy; and (ii) a belief-based strategy. If (i) is unfeasible, the individual is likely to resort to (ii). They will engage in a process of motivated reasoning, convincing themselves that the distribution of income is just because it is meritocratic. Therefore, their credence in U is reduced towards zero. In the following section I outline an experimental design for testing this mechanism.

4.6 Experimental design

The basic structure of the experiment is based on the ultimatum / proposer design. One set of participants (the "Receivers") receive an unequal, and potentially unjust division of money from another set (the "Proposers"). The Proposers have either earned the money by doing a real effort task or were simply given it, depending on a lottery whose probabilities are known to the Receivers. The Receivers then have to guess whether their own Proposer earned the money or not. Half of the receivers are in a "Choice" condition, where they can reject the

Proposer's offer, giving both themselves and the Proposer zero, and half are in a "No-choice condition" where they do not have the option to reject.

The general prediction is that those Receivers who get an unequal offer experience cognitive dissonance. If they are in the "No-choice" condition, they do not have recourse to an "action-based" dissonance reduction strategy. They must instead resort to a "belief-based" strategy: to convince themselves the Proposer had earned the money and hence deserved to keep most of it. Hence, a desire to reduce dissonance elicits belief that income is meritocratically distributed. The experiment was coded using the python-based oTree platform, and the code has been made available on GitHub[21].

How apt is this design – which is, in essence, the classic ultimatum/dictator design – to capturing people's approach to inequality in the "real world"? It might be objected that the experience of inequality refers to an individual's exposure to a whole society rather than a dyadic offer context. However, much of an individual's experience of inequality surely does occur at a micro level – for example, in the context of being offered a wage by an employer. Hence, the design might be seen as most aptly extrapolating to micro- and meso-level contexts such as firms and other bargaining scenarios between a relatively more powerful and less powerful actor.

4.6.1 Proposers

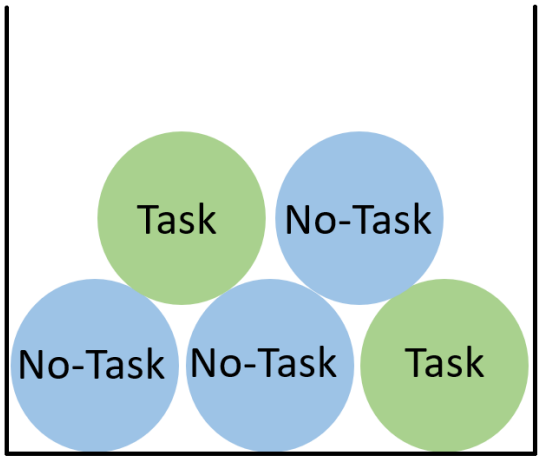
The first set of participants hired were the Proposers (N = 20), who did their part on the morning of the 28th of March 2022. They were British adults hired through the Prolific platform. They were paid £2.50 to participate, and had the chance to win an additional £2.70. The median time they took to completion was 194 seconds. The median earning was £4.30.

At the beginning of the experiment, the Proposers were entered into a lottery that would determine whether they did a real effort task to earn some money or whether they would simply be gifted it. They were shown the image in Figure 4.1, and it was explained to them that one of five balls would be chosen from a box. If the chosen ball were a "Task ball" they would have to do a maths task, consisting of four maths questions[22]. Maths questions were

[21] The code for the Proposer component can be found here: https://github.com/LiamOFoghlu/Proposer, and the Receiver component here: https://github.com/LiamOFoghlu/Receiver.
[22] The maths questions are reproduced in Appendix 3.1.

chosen because mathematics is generally perceived as a challenging subject (Brandenberger, Hagenauer, and Hascher 2018; Metje, Frank, and Croft 2007), and it was important that the Receivers believed the task to require some degree of effort and competence to complete. If the Proposer answered at least three questions correctly, they would receive a £3 prize. On the other hand, if one of the "No-task" balls were drawn, they would get to skip the task and simply be gifted the £3 prize. Since there were only two Task balls, and three No-task balls, there was a 0.6 probability of drawing a No-task ball[23].

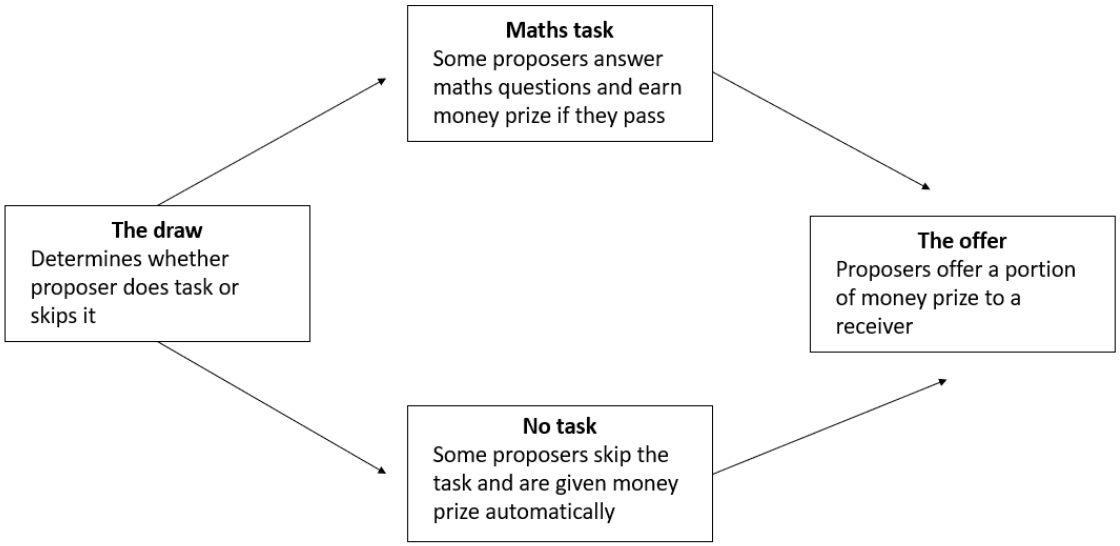
Figure 4.1. Visual illustration of the lottery.



Once the Proposer got the prize, whether by earning it through the task or by being gifted it, they had to offer a portion of it to a Receiver. They had the choice to offer either 40% or 10% of the £3. The offer of 40% represents a low-inequality outcome, with Proposer and Receiver getting almost the same amount. In ultimatum games, 40% is roughly the median offer (Camerer 2011; Tisserand 2014), and about 10% more than the median in dictator games (Engel 2011). An offer of this magnitude is very rarely rejected (Camerer 2011), which suggests that Receivers generally think it is a fair offer. The second offer, 10%, represents a very unequal division. Offers below 20% are rejected about half the time in ultimatum games (Camerer 2011), suggesting that it is generally regarded as an unfair offer.

[23] This probability was chosen because Receivers would be asked a binary question about whether their Proposer earned the money. If the prior probability was already above 0.5, then most Receivers, regardless of the treatment, would guess their Proposer earned the money making it difficult to distinguish between treatments.

Figure 4.2. Proposer Task Outline



They Proposers were also told that the Receiver could be in one of two conditions: a Choice condition or a No-choice condition. As explained, Receivers in the Choice condition could reject the offer, giving both themselves and the Proposer zero. Receivers in the No-choice condition did not have this option. To prevent strategic behaviour by the Proposers, they were not told which condition the Receiver they would be matched with would be in. They were simply told that there was a fifty-fifty chance of being either. Because there were much fewer Proposers than Receivers, a given Proposer would be randomly matched with many Receivers and one of these matches would be randomly selected to determine their payoff. The structure of the Proposer task is summarised in Figure 4.2.

4.3.2 Receivers

After the Proposers did their part, the Receivers (n = 351) were hired that same afternoon. Because I am particularly interested in why working class people espouse a belief in meritocracy, this sample was restricted to participants from a working class background. In effect, this meant restricting the sample to non-students without university degrees, aged 25 and over. The age restriction was enforced to ensure that potential participants who hadn't registered their student status weren't accidentally hired. The sample was evenly balanced between male and female. As with the Proposers, they were also British. They were paid

£2.25 for their participation, and could earn a bonus of up to £2.20. The median time to completion was 317 seconds, the median earning was £3.45.

At the beginning of their component, the Receivers were informed about the nature of the Proposer's task. They were told about the lottery and shown the image in Figure 1 to ensure they understood the odds of drawing a Task or No-task ball. They were then informed whether they were in the Choice or No-choice condition. Following which, they received the Proposer's offer of either 40% or 10% of £3. I refer to the 40% offer as the "Low-inequality" condition, and the 10% offer as the "High-inequality" condition.

Depending on the offer they received and whether they had the option to reject the offer or not, the Receiver ended up in one of four different treatments: (i) High-inequality Choice – they received an offer of 10% which they could reject if they wanted to, giving both themselves and the Proposer zero; (ii) High-inequality No-choice – they received an offer of 10% which they did not have the option to reject; (iii) Low-inequality Choice –they received an offer of 40% which they could reject if they wanted to; (iv) Low-inequality No-choice – they received an offer of 40% which they did not have the option to reject.

Table 4.1. Receiver treatment conditions and sample sizes.

	Choice	**No-choice**	***Total***
High inequality (10% offer)	129	125	254
Low inequality (40% offer)	48	49	97
Total	177	174	351

Table 4.1 summarises the sample size for each treatment. As can be seen, more participants were recruited into the High-inequality condition compared to the Low-inequality condition. The rationale for this was to concentrate statistical power within the most theoretically important comparison. It is within the High-inequality condition that cognitive dissonance may arise; and it is the difference between the High-inequality No-choice and High-

inequality Choice treatments that determines the available dissonance reduction strategies. (See the pre-analysis plan[24] for details on statistical power.)

Following the offer, the Receivers were asked to guess whether the Proposer they were matched with (i) earned the money by doing the task; or (ii) was gifted it. This is the primary dependent variable: whether or not the Receiver thinks the Proposer earned the money. Receivers got £1 if they guessed correctly. Following this guess, they were given a short survey consisting of two open-ended questions and a single binary question. These are the questions they were asked:

Question 1. Please write in the box below a short explanation of why you thought Proposer N drew a Task ball / No-task ball.

Question 2. Proposer N offered you £0.30 / £1.20 and kept £2.70 / £1.80 for themselves. Do you think this was fair? Please write a short explanation in the box below.

Question 3. All in all, do you think Proposer N's offer was fair or unfair?

- Unfair
- Fair

Both questions 1 and 2 are open-ended questions, designed to understand the logic underlying the Proposer's guess. Question 3 is the secondary dependent variable.

4.7 Theoretical expectations and hypotheses

4.7.1 Theoretical expectations

The analytic sample is drawn from the Receivers' data. There are two dependent variables: (i) whether the Receiver thinks their Proposer earned the money; and (ii) whether the Receiver thinks the Proposer's offer was fair. I expect the answer a Proposer gives for each dependent variable to be determined by (i) the Receiver's prior credence that their Proposer earned the money; (ii) what the Proposer's offer signals about the probability they earned the money;

[24] Available here: https://osf.io/e265w/

(iii) whether the Receiver experiences cognitive dissonance; and (iv) what options they have to reduce the dissonance.

Let $c(earner)$ be the Receiver's prior credence that their Proposer earned the money. At the beginning of the experiment, Receivers were told that there was only a 2 in 5 chance the Proposer had to do the task. This information was given to them again when they were asked to guess whether the Proposer was an earner. Hence, Receivers begin with the prior $c(earner) = 0.4$. But the offer the Proposer sends to the Receiver gives a potentially informative signal on the Proposer's identity as an earner or non-earner – *if* the Receiver believes that earners and non-earners have differing propensities to make low or high offers. So, after receiving an offer a Receiver may update to a posterior credence $c(earner|offer)$.

What kind of signal will the offer send? I expect that the Receivers will be aware of an "entitlement norm" – a belief that people who earn money through effort and skill may feel entitled to a larger share of it. The entitlement norm underlies meritocratic reasoning itself, and its existence is supported by experimental evidence (Carr and Mellizo 2013; Engel 2011; Hoffman et al. 1994; Korenok, Millner, and Razzolini 2017; List and Cherry 2000; Ruffle 1998). (Though for a null effect see Demiral and Mollerstrom [2020]). Following the entitlement norm, Receivers may infer that a Proposer is more likely to offer 10% when they have earned the money, and 40% when not. The offer therefore constitutes a signal which may cause the Receiver to update their credence that their Proposer did or did not earn the money. Following the entitlement norm, an offer of 40% should lead to a posterior which is lower than the prior probability: $c(earner|offer = 40\%) < c(earner)$. On the other hand, an offer of 10% should lead to a posterior which is greater than the prior probability: $c(earner \mid offer = 10\%) > c(earner)$.

I expect that a Receiver will guess their Proposer was an earner if $c(earner) > 0.5$. If not, they will guess the Proposer is a non-earner. However, even if the updated posterior, $c(earner|offer = 10\%)$, has a value exceeding 0.4, it does not follow that$c(earner|offer) > 0.5$, since the credence may not be very elastic to the signal. If the receiver is following a Bayesian updating procedure when they receive the offer, their posterior credence will result from the following equation:

$$c(earner \mid offer) = \frac{c(earner) * c(offer \mid earner)}{c(offer)}$$

Since $c(\ earner\ )$ is already exogenously given, the essential term governing the updating process is the likelihood that an earner would give a particular offer, $c(\ offer \mid earner\ )$. More precisely, the most important term is the likelihood ratio:

$$\frac{c(\ offer \mid earner\ )}{c(\ offer \mid non-earner\ )}$$

If this ratio is greater than 1.5, then a Receiver following a Bayesian updating procedure would guess their proposer earned the money. If not, they won't.

Receivers in the Low-inequality Choice and Low-inequality No-Choice treatments will receive an offer of 40%. Their posterior credence $c(earner|offer = 40\%)$ will be lower than 0.5, hence they will guess their Proposer did not earn the money. They should regard the offer as just, since it is the median offer in ultimatum games (Camerer 2011; Tisserand 2014). They will not experience cognitive dissonance.

Receivers in the High-inequality Choice and High-inequality No-Choice treatments will receive an offer of 10%. Since this offer is very low, and tends to be rejected in ultimatum games (Camerer 2011), it is quite likely that these Receivers will regard the offer as unjust. Hence they will form a positive credence towards the proposition U = "I have received an unequal offer", $c(U) > 0$. Next, if the Receiver possesses some positive credence towards the proposition J = "People are just", $c(J) > 0$, as Just World theory predicts they should (Lerner 1980), then they will possess a positive credence towards two mutually contradictory propositions, U and J. This will generate cognitive dissonance proportional to the product of the credences.

Receivers experiencing dissonance will want to reduce it (Festinger 1962). Those in the High-inequality Choice condition can avail of an action-based dissonance reduction strategy: they may reject the unequal offer, thereby restoring some rough justice by giving both them and the Proposer zero. Since the distribution of income is now totally equal (at zero each), $c(U) = 0$. But those in the High-inequality No-choice condition will not have access to an action-based dissonance reduction strategy. Their only recourse is to avail of a belief-based dissonance reduction strategy. They will convince themselves the offer would be fair if the Proposer earned it, and that the Proposer is likely to have earned it because they feel entitled to the lion's share of the prize. Table 4.2 summarises the expected dissonance-reduction strategies for each treatment.

Table 4.2. Dissonance reduction strategies for each treatment.

Treatment condition	Experience dissonance	Dissonance reduction strategy
Low-inequality Choice	No	n/a
Low-inequality No-choice	No	n/a
High-inequality Choice	Yes	Action-based
High-inequality No-choice	Yes	Belief-based

The differential availability of dissonance reduction strategies should therefore affect the updating process. Following the entitlement norm, both High-inequality Choice and High-inequality No-choice Receivers will update their posterior credence $c(earner \mid offer = 40\%)$. But this posterior credence will only be greater than 0.5 if the following inequality holds:

$$\frac{c(\,offer \mid earner\,)}{c(\,offer \mid non-earner\,)} > 1.5$$

Of course there will naturally be some uncertainty about the exact value of this ratio, and it is in these conditions that wishful thinking will flourish. I predict that the Receivers in the High-inequality No-choice condition will be more likely to convince themselves that this ratio is greater than 1.5 because this is their only way to reduce dissonance. The High-inequality No-choice Receivers want to believe the Proposer earned the money because the unequal offer will then, following meritocratic principles, appear fair. High-inequality Choice Receivers do not have to convince themselves the offer is fair – they can reject it if they find it unfair. Hence, I also expect Receivers in the High-inequality No-choice treatment to be more likely than Receivers in the High-inequality Choice treatment to report that they think their offer is fair.

4.7.2 Hypotheses

Each hypothesis is formulated for either: (i) $prop(earner \mid t)$, the proportion of participants in treatment condition *t* who guessed their Proposer was an earner; and (ii) $prop(fair \mid t)$, the proportion of participants in treatment condition *t* who thought their Proposer's offer was fair. The hypotheses consist of comparisons of proportions between Choice and No-choice

conditions within a given inequality condition. All hypotheses are pre-registered[25]. In the notation I abbreviate each treatment with their initials: Low-inequality Choice (LC), Low-inequality No-choice (LN), High-inequality Choice (HC), and High-inequality No-choice (HN).

Dependent variable (i): whether the Proposer is an earner

Low-inequality Choice and Low-inequality No-choice receivers get a more or less fair offer. I do not expect them to experience dissonance. There should be no difference in the proportion of participants guessing their Proposer was an earner. Hence:

H1 $prop(earner \mid LN) = prop(earner \mid LC)$**.**

High-inequality Choice and High-inequality No-choice receivers experience cognitive dissonance. High-inequality Choice Receivers can resolve their dissonance through rejecting the Proposer's offer. High-inequality No-choice Receivers have recourse only to wishful thinking: they convince themselves the Proposer earned the money. Hence:

H2 $prop(earner \mid HN) > prop(earner \mid HC)$.

Dependent variable (ii): whether the Proposer's offer was fair

Following the same logic as for H1, I expect no difference in fairness judgments between Proposers in the Low-inequality No-choice and Low-inequality Choice.

H3 $prop(fair \mid LN) = prop(fair \mid LC)$.

However, High-inequality No-choice Receivers should be more likely than High-inequality Choice receivers to believe their offer was fair because it is their only strategy for dissonance reduction.

H4 $prop(fair \mid HN) > prop(fair \mid HC)$.

[25] Pre-registration available here: https://osf.io/e265w/

4.8 Results

4.8.1 Analysis of hypotheses

This section presents the results for each of the pre-registered hypotheses[26]. In the pre-analysis plan[27] it was stated that the statistical tests would be carried out using a non-parametric method: specifically, a randomisation test. For robustness' sake, I also report p-values from a parametric test of proportions using Pearson's chi-squared test statistic.

The proportion in each condition who guessed their Proposer was an earner is reported in Table 4.3, and in Figure 4.3. In Table 4.3, the parametric and non-parametric p-values at the end of each row correspond to the p-values for a test of difference of proportions between the Choice and Non-choice conditions within each inequality condition (i.e. Low-inequality Choice versus Low-inequality No-choice; and High-inequality Choice versus High-inequality No-choice).

As can be seen from Table 4.3, there is no (statistically significant) difference between the Low-inequality Choice and Low-inequality Choice treatments. **H1** is supported. However, there is also no (statistically significant) difference between the High-inequality Choice and High-inequality Choice treatments. **H2** is not supported. There is no evidence, on the basis of these results, that a cognitive dissonance mechanism underlies belief in meritocracy.

Table 4.3. Proportion in each condition who guessed their Proposer was an earner.

	No-choice	**Choice**	**p-value (non-parametric)**	**p-value (parametric)**
Low inequality	0.29	0.29	0.814	0.948
High inequality	0.42	0.45	0.611	0.681

The proportion in each condition who thought their offer was fair is reported in Table 4.4, and in Figure 4.4. As with Table 4.3, the parametric and non-parametric p-values at the end of each row correspond to the p-values for a test of difference of proportions between the Choice and Non-choice conditions and within each inequality treatment (i.e. Low-inequality

[26] A replication package to reproduce these results is available at https://osf.io/e265w/
[27] Available here: https://osf.io/e265w/

Choice versus Low-inequality No-choice; and High-inequality Choice versus High-inequality No-choice).

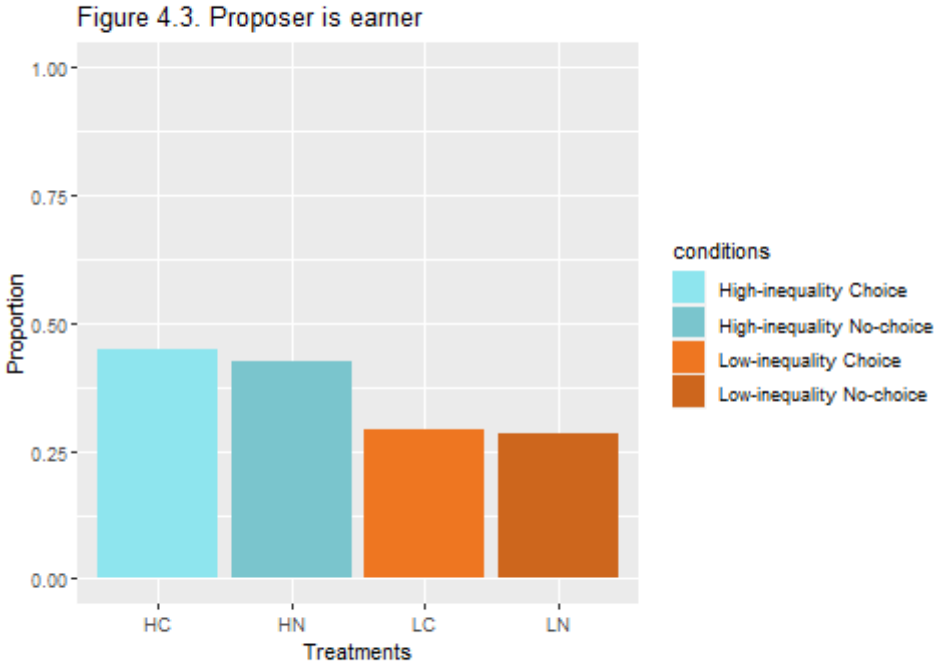
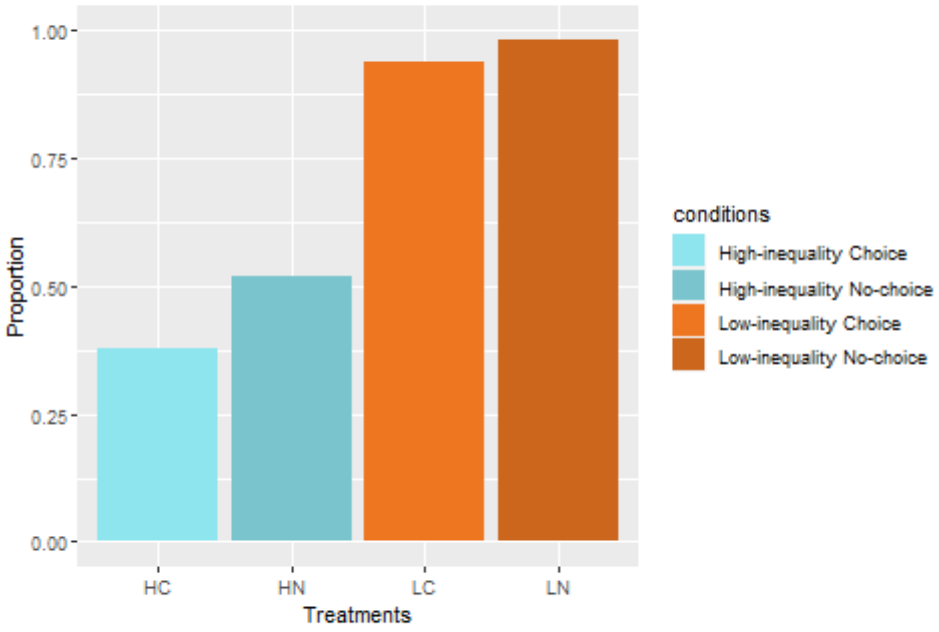


Table 4.4. Proportion in each condition who thought their Proposer's offer was fair.

	No-choice	**Choice**	**p-value (non-parametric)**	**p-value (parametric)**
Low inequality	0.98	0.94	0.118	0.297
High inequality	0.52	0.38	0.023	0.025

As can be seen, there is no (statistically significant) difference between the Low-inequality Choice and Low-inequality Choice treatments. **H3** is supported. However, there *is* a statistically significant difference between the High-inequality Choice and High-inequality No-choice conditions. 52% of Receivers in the High-inequality No-choice condition thought their offer was fair, but only 38% of Receivers in the High-inequality Choice condition thought their offer was fair. **H4** is supported.

Figure 4.4. Proposer's offer is fair



The results are surprising, at least in light of the theoretical discussion above. Even though the Receivers in the High-inequality No-choice condition were no more likely guess their Proposer earned the money than the Receivers in the High-inequality Choice condition, they *were* more likely to judge the offer to be fair. The difference, while not massive, was substantive. It amounts to a proportion of 0.52 – 0.38 = 0.14, which corresponds to a small-to-medium effect size using the Cohen's *h* typology. Given the small stakes and the brevity of the experiment (median time to completion of just over 5 minutes), this is a sizeable effect.

The results suggest that people who are disadvantaged by income inequality, *and* do not have the means to change the income distribution, are more likely to judge the inequality as fair – *even though they are no more likely to think it arises from distinctions of merit.* This is an intriguing result, since previous work has suggested that income inequality was seen as justifiable when it was the result of differences in merit (Ahrens 2020; Alesina and Angeletos 2005; Janmaat 2013), with the implication often present that this is more or less the only

reason people would find inequality to be justified. But the results here suggest that another type of rationale is at play.

In order to discover that rationale, I analyse some qualitative data gathered from open-ended questions presented at the end of the experiment. It goes without saying that these results cannot be treated as anything other than suggestive. They do not derive from pre-registered hypotheses, and the qualitative data itself consists of fairly brief, sometimes-cryptic responses. Nonetheless, I think an interpretative analysis of the data is warranted, both on methodological and substantive grounds.

Methodologically, the interpretation of experimental results all too often relies on the researchers' theoretical assumptions. Analysis of qualitative data complements the deductive strengths of the experimental method, with the inductive insights that can be gained through more interpretative methods. This is particularly valuable in the case of a "null result", where the researcher is often at a loss to explain "what went wrong". Beach and Littvay (2020) have recently called for complementing experimental designs with a "process tracing" perspective; a methodology which allows the researcher to distinguish the "observational traces" that prospective causal mechanisms leave behind as they operate. Substantively, the data contribute to our understanding of the theoretically unanticipated results of this experiment, and allow for the formation of new deductive hypotheses. More broadly, the answers reveal that despite the brevity of the experiment, the anonymity of the interaction and the smallness of the stakes, the Receivers (for the most part) engaged in substantive, often affect-driven, reasoning.

4.8.2 Frames

The Receivers were asked two open-ended questions:

Question 1. Please write in the box below a short explanation of why you thought Proposer N drew a Task ball / No-task ball

Question 2. Proposer N offered you £0.30 / £1.20 and kept £2.70 / £1.80 for themselves. Do you think this was fair? Please write a short explanation in the box below.

In order to interpret the resulting data, I aimed to identify the common rationales used by the Receivers in their answers to the question. First I created a dataset, consisting of the following variables: the participant's ID code (a random string of numbers and letters, e.g.

azi3yv6v); the offer they received (10% or 40%); their guess about whether the proposer was an earner; their fairness judgment (the binary question which asked them to judge whether the offer was "fair" or "unfair"); their answer to Question 1, which I term their "guess rationale"; and their answer to Question 2, which I term their "fairness rationale". I deliberately did not include information on the Choice treatment, which would have allowed me to know whether they were in the High-inequality Choice or High-inequality No-choice conditions. By excluding this information, I thought my answers would be less biased by incipient hypothesising.

After creating the dataset, I simply read through the guess and fairness rationales and noted down common themes. I then organised these themes in terms of "frames" and went through the data again, assigning one or more frames to each answer. "Frame" is a usage borrowed from Bacharach (2006), who developed the concept of the frame as a way to explain "anomalous" results in game theoretic experiments. According to Bacharach, the frame is the lens through which the individual "sees" the problem they are presented with in the game. Players who "see" a prisoner's dilemma in terms of a "we" frame will think the problem is one of achieving co-operation between them and the other player. Those who see the game in terms of a "me" frame, will see the problem in the way it was seen by classical game theory: as a question of maximising personal gain in an interactive context. Hence, judgments and actions often follow "thick", value-laden, often affect-driven cognitive processes, rather than "thin", utilitarian decision-making procedures. The "frame" concept has subsequently been used by Butler, Burbank, and Chisholm (2011) in an interesting qualitative analysis of interviews with participants in ultimatum and dictator games.

4.8.3 Guess rationale

I identified five different frames used by Receivers in their "guess rationale", the reason they gave for why they guessed the Proposer did / didn't earn the money. The frames are summarised in Table 4.4. The "Frame" column gives the name of the frame, the "Logic of frame" column gives the substantive meaning of the frame, and the "Example quote" column gives an example quote from a Receiver, identified by their ID code, who used this frame in their rationale.

Table 4.4. Guess rationale frames

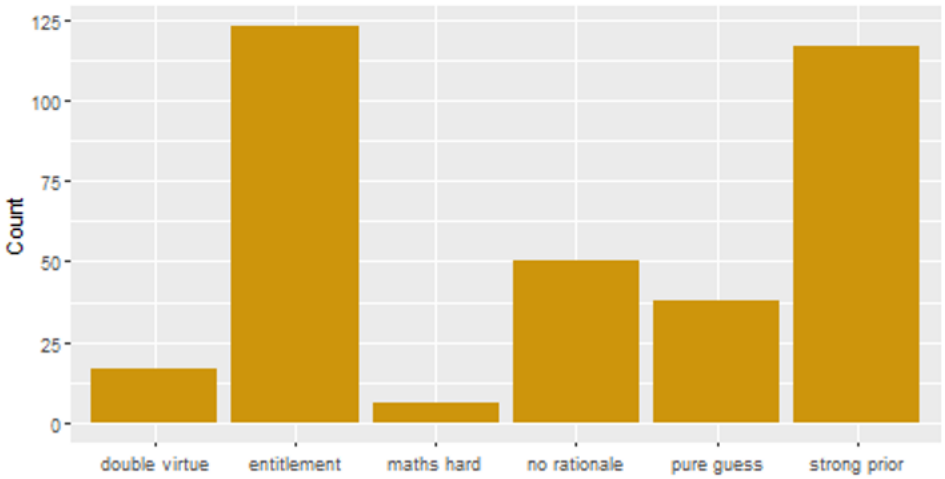
Frame	Logic of frame	Example quote
Strong prior	Since there are more No-task balls than Task balls, it is more likely that a Proposer drew a No-task ball.	"There were 3 no-task balls and 2 task balls therefore the chances of drawing a no-task ball were slightly higher." (Receiver nw5zpcfk)
Entitlement inference	A Proposer who earned the money is more likely to offer 10% than a Proposer who didn't earn the money.	"I think that someone who worked for the task is likely to offer the lower amount." (Receiver acwbmrvn).
Maths hard	The maths questions were hard, so even if a Proposer drew a Task ball there is a good chance they would get less than 3 correct and not get the £3.	"The maths challenges seemed quite difficult so I'm guessing they didn't do any" (Receiver pivk9vxv).
Pure guess	A pure and simple guess. Perhaps based on gut instinct	"I'm not too sure, it was just a total guess." (Receiver azi3yv6v)
Double virtue	Those who are smart enough to complete a maths task are also generous, and hence likely to offer 40%. Conversely, those who offer small amounts are greedy and unintelligent.	Example quote: "They only gifted 10% which made me think they aren't very bright or fair, had they needed to answer a question I'm confident they would not be clever enough to answer it correctly" (Receiver oee1o3kp)

Some Receivers gave answers that were too short to have any substantive content, or simply inscrutable. These answers were assigned the (non-)frame: "No rationale". The absolute number of participants who chose each guess rationale are graphed in the bar chart in Figure 4.5, alongside those whose answer was marked "No rationale".

The guess rationales supplied by the Receivers were, for the most part, unsurprising. We can see in Figure 4.5 that the joint most popular rationales were Entitlement (35% of Receivers) and Strong prior (33% of Receivers). Both rationales were theoretically anticipated. In the "Theoretical expectations and hypotheses" section above, I presumed that the receiver's

guess would depend on whether the signal sent by the offer would sufficiently counterbalance the prior information on the lottery when receivers were forming their posterior probability. Those who gave Strong prior as their rationale ultimately gave more weight to the prior, while those who gave Entitlement as their rationale gave more weight to the signal.

Figure 4.5. Guess rationales



4.8.4 Fairness rationale

I identified eight different frames used by Receivers in their fairness rationale, the reason they gave for why they thought the Proposer's offer was fair or unfair. Not all the frames were totally distinct from each other, and each shared a common moral language with at least one other frame. Recognising this, I categorised each of the eight frames within one of three "metaframes": (i) The Norm metaframe refer to abstract obligations concerning the relationship between individuals; (ii) The Virtue metaframe concerns the moral behaviour or character of particular individuals; and (iii) the Ontology metaframe involves more general statements about the nature of humans and human society. Owing to the relative complexity and nuance of some answers, more than one metaframe was assigned to some of the Receivers. Table 4.5 summarises the fairness frames.

Table 4.5. Fairness rationale frames

Metaframe	Frame	Logic of frame	Example quotes
Norm	Egalitarian	Money should be shared equally, as a matter of principle.	"No, it should be shared out more equally, this is what I would have done. " (Receiver 3ioohqcm). "no i don't think

			it was fair because it wasn't equal" (Receiver xzpff0r3).
Norm	Entitlement	It is fair for the Proposer to claim the larger share of the money *if* they earned it by doing the maths task.	"Fair if they had to complete the task - passing 40% to a Receiver who potentially didnt have to do much for it would not feel reasonable to the Proposer" (Receiver q6im09sd). "If they had to do the task then fair but if they were just given the money they could have split it more evenly" (Receiver ujh3ujk2).
Norm	Proprietarian	Since the Proposer is the one making the offer, it is entirely their decision to make. It is their choice, and they should be allowed to make whatever choice they want.	"I think that the Proposer was looking after his own interests. I think he could have offered more but that was his choice so I suppose to them it was fair. I think he could have offered more." (Receiver q4xsat3b). "i do think this was fair as this was their decision" (Receiver z0qyc8yx).
Virtue	Greed	People shouldn’t claim more than a certain amount of money.	"I don't think it's fair I think it's someone being greedy" (Receiver nslps6lj). "No, which is why I rejected it. I'd rather have nothing than someone get more out of greed." (Receiver 3r656kbm).
Virtue	Generosity	People have given more than morality or justice strictly required.	"No I don't think it was fair they could have been more generous." (Receiver zffrt4rr). "yes, I think it was rather nice of them to do this as they could have had 90% of the money and after all, they were the ones that was (potentially) doing the leg work by answering the maths question!" (Receiver u727y7j4).
Ontology	Fatalism	Humans do not owe general obligations to each other.	"No, it isn't fair, but that's people for you. People

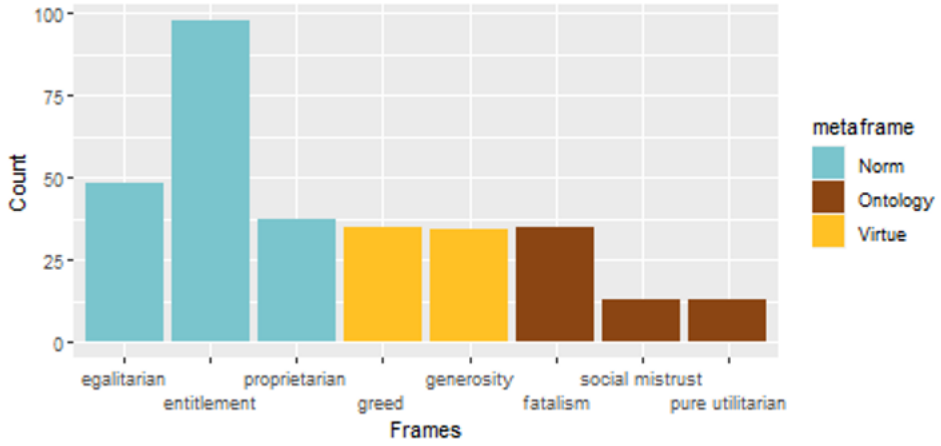
		Each of us will take as much as we can get, even if that means depriving others.	always look after themselves." (Receiver wo8i6md6). "yes it seems fair to me. Not everything is guaranteed and nobody is obligated to offer anything in life really." (Receiver 3ytnpcfd).
Ontology	Social mistrust	There are no obligations to share money with people who you don't know.	"I mean it's not particularly fair, but given they are matching with a total stranger I can completely understand their decision. " (Receiver sjpz3paa). "Obviously I would rather they donate more of their money, but we have no connection to one another so the choice is easy to make. I would make the same choice, and thus I can't really describe it as unfair." (Receiver a9c01yi4).
Ontology	Pure utilitarian	Receiving anything at all is good, and something is better than nothing.	"Slightly unfair but it was the Proposer's choice and given that I had no choice, I still got something." (Receiver rcwk3sqi). "It would have been nice to have received more but something is better than nothing." (Receiver hbjh72qg).

Figure 4.6 shows the total number of times each frame was used. Overall, 314 fairness rationales were coded. Some Receivers, about 9%, included more than one metaframe in their rationale, hence Figure 6 "double counts" some participants. On the other hand, just under 20% of Receivers did not give an answer from which a substantive rationale could be extracted.

Just over half of the frames used (183) belong to the Norm metaframe. The predominance of this frame may owe something to the anonymity of the interaction and the artificiality of the environment, since the frame concerns abstract obligations between individuals rather than

judgments of character. The remaining frames are split almost evenly between the Virtue metaframe and the Ontology metaframe.

Figure 4.6. Fairness frames



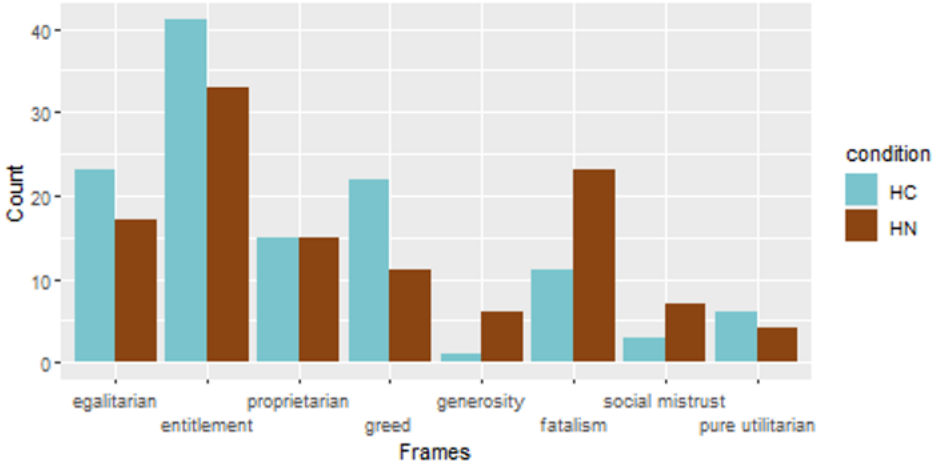
What is striking about the fairness rationales is their sheer diversity, exceeding the scope which had been theoretically anticipated. The only fairness rationale I explicitly discuss in the theoretical section of this paper is the entitlement one. And while this is the modal frame, it was only used by 28% of Receivers. People are a lot more creative and diverse in their moral reasoning than was anticipated. But is there some systematic pattern amongst this variety, which can explain the difference in fairness judgments between the High-inequality Choice and High-inequality No-choice treatments?

4.8.5 Explaining the High-inequality Choice / High-inequality No-choice difference

Figure 4.7 shows the distribution of fairness rationales between the High-inequality No-choice and High-inequality Choice treatments. In order to highlight the differences between the conditions, only the conditions and not the metaframes are colour-coded.

In looking for an explanation of the difference in outcome, we would want to see some frame or frames which: (i) are sufficiently large as a proportion of all frames; and (ii) show a sufficiently unbalanced ratio between the High-inequality Choice and High-inequality No-choice treatments. Eyeballing Figure 4.7, the fatalism and greed columns seem to stand in this regard. Both represent a relatively large proportion of frames and a relatively large ratio between conditions.

Figure 4.7. Fairness frames for HN and HC conditions



But the eyeball can be a crude and biased instrument. Table 4.6 summarises several statistics which allow us to judge criteria (i) and (ii) for each frame. Let σ_{ci} be the sum of Receivers in the High-inequality Choice condition who used frame *i*, and σ_{ni} be the sum of Receivers in the High-inequality No-choice condition who used frame *i.* The second column in Table 6, with the heading "Total count", summarises the total number of usages of each frame across both treatments, i.e $\sigma_{ci} + \sigma_{ni}$. The third column, "Difference", is the difference in total usage of each frame between the High-inequality Choice and High-inequality No-choice treatments, i.e. $\sigma_{ci} - \sigma_{ni}$. The fourth column, "Weighted difference", is an attempt to combine criteria (i) and (ii), through multiplying the Difference parameter by the Total count for that frame divided by the total number of frames used across treatments, i.e. $(\sigma_{ci} - \sigma_{ni}) * [(\sigma_{ci} + \sigma_{ni})/(\sum_i^N \sigma_{ci} + \sigma_{ni})]$, where $N = 8$ is the index of unique frames. The final column shows the natural log of the ratio of the sum of Receivers in the High-inequality Choice condition who used frame *i*, and the sum of Receivers in the High-inequality No-choice condition who used frame *i*, i.e $ln(\sigma_{ci}/\sigma_{ni})$. (The natural log is used so that the difference between an overbalance and an underbalance is symmetrical.)

The Weighted difference measures provides a single statistic summary of both criterion (i) and (ii). For now, let's exclude from our analysis, somewhat arbitrarily, all frames with an *absolute* Weighted difference < 1. (Actually this is not a very arbitrary exclusion threshold since none of the excluded frames have an absolute weighted difference > 0.2.) This leaves us with four frames: egalitarian, entitlement, greed, and fatalism. The first three frames concern moral evaluations of action or character; the last frame is a rejection of moral evaluation per se. Those in the High-inequality Choice condition are more likely to use the first three

frames, and those in the High-inequality No-choice condition are more likely to use the last frame.

What this suggests is that the High-inequality No-choice Receivers were *less* likely to make moral evaluations, either concerning abstract norms or personal virtues, and *more* likely to reject the idea of making moral evaluations at all. Instead, they tended to take a fatalistic view of things: human beings owe each other nothing; hence it is fair enough to claim as much as possible, even if that means less for the other person. They tended to endorse an "anything goes" mentality, often stating that they would also offer the lowest amount if they were in the Proposer's position. Their answers reflected an anomistic perspective on life, where there is an absence of norms governing the division of wealth between people.

Table 4.6. Distribution of frames across High-inequality Choice and High-inequality No-choice conditions.

Fairness frames	**Total count**	**Difference**	**Weighted difference**	**Logged ratio**
egalitarian	40	6	1	0.30
entitlement	74	8	2.5	0.22
proprietarian	30	0	0	0.00
greed	33	11	1.5	0.69
generosity	7	-5	-0.1	-1.79
fatalism	34	-12	-1.7	-0.74
social mistrust	10	-4	-0.2	-0.85
pure utilitarian	10	2	0.1	0.41

The evidence suggests that some level of control is required for a sense of moral agency. Moral judgments seem to be connected with the ability, even if only counterfactual, of taking action. The absence of control appears to smother moral indignation, and cultivate equanimity in the face of (potential) injustice. For a significant portion of working class, an attitude of resigned indifference seems to be the preferred response to inequality.

4.9 Discussion and conclusion

In this paper, I presented and tested a mechanism purporting to explain why people disadvantaged by inequality might be autonomously motivated to legitimise their own disadvantage. The mechanism was derived from Elster's (1985) work on wishful thinking, where he argues that people can be motivated to justify their own disadvantage in order to mitigate cognitive dissonance. I termed this type of motivated reasoning a "belief-based" dissonance reduction strategy, and argued it was more likely to occur when individuals lack agency or control over the income distribution and their place in it. Individuals engaging in the "belief-based" dissonance reduction strategy could make use of the prevalent cultural schema of "meritocracy", and convince themselves that the income distribution is fair because it is based on merit.

To test this mechanism, I designed and administered an experiment, where participants were randomly sorted in a High- or Low-inequality condition, which determined how even their split of a money prize was; and a Choice or No-choice condition, which determined whether they had some measure of control over the distribution of the prize. The results did not support the cognitive dissonance explanation. Receivers in the High-inequality treatment, who did not have control over the distribution, were no more likely to think the Proposers earned the money than those who had control. But they were more likely to think it the division of money was fair.

In order to interpret this finding, I turned to the analysis of qualitative data gathered through two open-ended questions. My interpretation of the data led me to the conclusion that Receivers in the High-inequality No-choice condition reported a more pessimistic perspective on humanity. They espoused the view that people cannot be expected to treat equally with each other, and that each person would try to claim as much as they could even at the expense of others. In consonance with this, they were also less likely to express a moral evaluation of the Proposer's behaviour and more likely to express and attitude of resigned indifference.

One implication of this interpretation is that when individuals are disadvantaged by inequality, and lack the means to rebalance the distribution, they may mentally shield themselves from distress by adopting a fatalistic attitude. In his seminal contribution to cultural studies, The Uses of Literacy (Hoggart 1992), Richard Hoggart reflected on the prevalence of this attitude among the working class community in which he grew up. For the

working class people of his youth, the elements of class society as they knew them acquired the character of "natural laws" to which people simply had to adapt themselves (Hoggart 1992:92). This outlook was reflected in a host of sayings and clichés: "what is to be will be", "that's just the way things are", "it's no good kicking against the pricks", "what can't be mended must be made do with", "worse things 'happen at sea", and so on (Hoggart 1992:92). Hoggart explains that this fatalistic attitude often towards the "natural laws" of their society could be understood as, itself, a type of strategy: "T.S Eliot says somewhere that stoicism can be a kind of arrogance, a refusal to be humble before God: working-class stoicism is rather a self-defence, against being altogether humbled before men" (Hoggart 1992:93).

This strategy for "self-defence" is not *necessarily* a form of wishful thinking. Elster is careful to distinguish the kind of motivated reasoning that the fox in the Sour Grapes fable got up to, from the deliberate cultivation of philosophies of forbearance and self-denial such as Stoicism, Buddhism, and Spinozism (Elster 1985:118). In some cases there really is nothing a person can do to change their fate, and it is more rational to accept the possibility of injustice or tragedy. In an epigraph to the chapter where he introduces the Sour Grapes paradigm, Elster cites the version of Ronald Niebuhr's famous "Serenity Prayer" aphorism, often used by Alcoholics Anonymous: "God grant us the serenity to accept the things we cannot change, courage to change the things we can and wisdom to know the difference." Interestingly, the Alcoholics Anonymous version actually reverses the order of the first two clauses, making a slight but important change in emphasis. In Niebuhr's original version, primacy is given to courage and action, whereas in the AA version primacy is given to the acceptance of the immutable.

The results could also be facilitated within the cognitive dissonance reduction framework if we conceive that Receivers in the High-inequality No-choice condition were reducing their credence in J = "Society is just". In the theoretical framework, I argued that dissonance reduction follows the past of least resistance, with people preferring to reduce their credence in the belief that has the relatively lower credence. I also argued that credence in the proposition "Society is just" was likely to be stronger than belief in the proposition U = "Income is unjustly distributed", both because people have a strong psychological motivation to believe that society is just and because people likely face both empirical and moral uncertainties regarding the justness of the income distribution. The latter uncertainty is replicated in the experiment by the fact that the Receiver does not know whether their

Proposer is an earner. However, it is also possible that my assumption that there is generally high credence in J = "Society is just" was mistaken. Perhaps people do not generally believe that society is just. Hence, it may have been easier for people to revise downwards their credence in J than U. The qualitative evidence would perhaps suggest that people engage in a more complex strategy than increasing or decreasing their belief in a proposition. Receivers who adopted the fatalistic frame arguably maintained a belief in a just society, but re-evaluated their notion of what a just society would be, from one in which people have an obligation to display some minimal level of altruism or generosity to a Hobbesian one in which all may legitimately avail of the advantages available them to maximise their own utility, regardless of the cost to others.

There are limitations to the conclusions and to the overall design and argument of the study. In terms of conclusions, the primary theoretical expectation – that Receivers in the High-inequality No-choice condition would be more likely to think their Proposer earned the money – was not supported. The interpretation for why High-inequality No-choice Receivers were more likely to judge their offer to be fair is based on qualitative data of a fairly limited character, and can only be considered suggestive. Further research could formally test this interpretation.

From a design perspective, it could also be questioned whether the division of money between two individuals can "stand in", experimentally speaking, for the distribution of income between the members of a whole society. This is a fair question. Inequality results from a broad array of factors, not all of which can be attributed to the conscious actions of wealthy individuals to take more for themselves at the expense of poorer ones. But experiments cannot exactly reproduce the conditions of society, nor are they designed to do so. I think it is fair to draw inferences from the results of this study to the extent that: (i) rich people do actually have some control over the degree of inequality in their society; and (ii) people *perceive* that rich people have some control over the degree of inequality in their society.

I think (i) and (ii) are perfectly valid assumptions. Nonetheless, the results may have more external validity if we extend them to meso- or micro-level contexts, such as organisations, small-scale social networks, and dyadic bargaining situations rather than society as a whole. So this study will be of greatest relevance to contexts like the workplace, where the control or

agency of workers may vary due to institutional factors such as the presence of trade unions, and where the workers' own share of revenue depends on how much management withholds as profit. In the experiment presented here, the Proposer and Receiver roles could be seen as analogous to that of management and worker. Indeed, perhaps the most common realworld setting invoked by studies implementing or addressing the ultimatum/dictator paradigm is the wage negotiation between employers and workers, which has included work on fairness perception (Fehr and Schmidt 2001; Pull 1999; Ruffle 1998). Another, arguably closer match, is to rental markets, where tenants usually directly engage with an individual landlord, and where the size of the rent directly impacts the division of income between tenant and landlord. In fact, the landlord/tenant metaphor is often invoked by papers which use the ultimatum/dictator paradigm to explore perceptions of fairness and inequality (for example Kirk, Gollwitzer, and Carnevale 2011; Pillutla and Murnighan 1995; Solnick 2001).

Restricting the external validity to such contexts is not much of a limitation in my view. Observational survey research where micro-level outcomes are regressed on macro-level variables such as GINI or GDP tend to present, implicitly or explicitly, the members of society as being like a pot of boiling potatoes, each homogeneously exposed to the same temperature. But, while people may read about inequality and economic growth in the papers, their actual experience of social forces tends to consist of micro-level interactions in meso-level contexts. Hence, as Tilly argues (1998), studying the internal processes of formal and informal organisations, may well provide the core insights into how inequality arises and persists.

A4.1 Maths questions

Below are reproduced the maths questions which were given to the Receivers who drew a Task ball.

Question one

A shop has an offer: buy 8 kiwis, and every extra kiwi after that is half price. A man goes to the shop and pays £4.50 for some kiwis. The full price of a kiwi is £0.50. How many does he buy?

○ 9 ○ 12 ○ 10 ○ 15

Question two

A hairdresser has an offer: every third visit is free. They charge £48 for a haircut. Last year Sarah paid £144 for a haaircut. How many times did she go?

○ Two times ○ Three times ○ Four times ○ Five times

Question three

A woman walks from the bottom to the top of a hill. She starts at 9.40am and arrives at the top at 10.20 am. She takes a rest for ten minutes. Then she walks back down. On the way down she walks twice as fast as she did on the way up. What time is it when she reaches the bottom of the hill?

○ 11.20 ○ 10.40 ○ 10.50 ○ 11.10

Question four

A trader buys a painting for £120 and sells it for £170. They pay a £10 transaction fee. Their profit expressed as a percentage of total cost is:

○ 50% ○ 60% ○ 80% ○ 33%

References

Ahrens, Leo. 2020. ‘Unfair Inequality and the Demand for Redistribution: Why Not All Inequality Is Equal’. *Socio-Economic Review* mwaa051. doi: 10.1093/ser/mwaa051.

Alesina, Alberto, and George-Marios Angeletos. 2005. ‘Fairness and Redistribution’. *American Economic Review* 95(4):960–80. doi: 10.1257/0002828054825655.

Bacharach, Michael. 2006. *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton, NJ, US: Princeton University Press.

Beach, Derek, and Levi Littvay. 2020. ‘A More Conscious Experimental Design: How More Explicit Use of Process Tracing Can Improve Experiments’. doi: 10.33774/apsa-2020-wfq1m.

Belzil, Christian, and Jorgen Hansen. 2002. ‘Unobserved Ability and the Return to Schooling’. *Econometrica* 70(5):2075–91.

Bénabou, Roland, and Jean Tirole. 2006. ‘Belief in a Just World and Redistributive Politics*’. *The Quarterly Journal of Economics* 121(2):699–746. doi: 10.1162/qjec.2006.121.2.699.

Blackburn, McKinsey, and David Neumark. 1992. ‘Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials’. *The Quarterly Journal of Economics* 107(4):1421–36.

Brandenberger, Claudia C., Gerda Hagenauer, and Tina Hascher. 2018. ‘Promoting Students’ Self-Determined Motivation in Maths: Results of a 1-Year Classroom Intervention’. *European Journal of Psychology of Education* 33(2):295–317. doi: 10.1007/s10212-017-0336-y.

Bredemeier, Christian. 2014. ‘Imperfect Information and the Meltzer-Richard Hypothesis’. *Public Choice* 159(3):561–76. doi: 10.1007/s11127-013-0072-z.

Brown, Phillip. 2013. ‘Education, Opportunity and the Prospects for Social Mobility’. *British Journal of Sociology of Education* 34(5–6):678–700. doi: 10.1080/01425692.2013.816036.

Butler, David J., Victoria K. Burbank, and James S. Chisholm. 2011. ‘The Frames behind the Games: Player’s Perceptions of Prisoners Dilemma, Chicken, Dictator, and Ultimatum Games’. *The Journal of Socio-Economics* 40(2):103–14. doi: 10.1016/j.socec.2010.12.009.

Camerer, Colin F. 2011. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Cansunar, Asli. 2021. 'Who Is High Income, Anyway? Social Comparison, Subjective Group Identification, and Preferences over Progressive Taxation'. *The Journal of Politics* 83(4):1292–1306. doi: 10.1086/711627.

Carr, Michael D., and Phil Mellizo. 2013. 'Entitlement in a Real Effort Ultimatum Game'. *Boston, MA: University of Massachusetts Boston* (2013_01).

Choma, Becky, Carolyn Hafer, Faye Crosby, and Mindi Foster. 2012. 'Perceptions of Personal Sex Discrimination: The Role of Belief in a Just World and Situational Ambiguity'. *The Journal of Social Psychology* 152(5):568–85. doi: 10.1080/00224545.2012.667459.

Correia, Isabel, and Jorge Vala. 2003. 'When Will a Victim Be Secondarily Victimized? The Effect of Observer's Belief in a Just World, Victim's Innocence and Persistence of Suffering'. *Social Justice Research* 16(4):379–400. doi: 10.1023/A:1026313716185.

Dalbert, Claudia. 2002. 'Beliefs in a Just World as a Buffer Against Anger'. *Social Justice Research* 15(2):123–45. doi: 10.1023/A:1019919822628.

Damasio, Antonio R. 1995. *Descartes' Error : Emotion, Reason and the Human Brain*. Picador.

Demiral, Elif E., and Johanna Mollerstrom. 2020. 'The Entitlement Effect in the Ultimatum Game – Does It Even Exist?' *Journal of Economic Behavior & Organization* 175:341–52. doi: 10.1016/j.jebo.2018.08.022.

Durkheim, Émile. 1952. *Suicide.* London: Routledge.

Durlauf, Steven N., and Ananth Seshadri. 2018. 'Understanding the Great Gatsby Curve'. *NBER Macroeconomics Annual* 32(1):333–93. doi: 10.1086/696058.

Dzuka, J., and C. Dalbert. 2006. 'The Belief in a Just World and Subjective Well-Being in Old Age'. *Aging & Mental Health* 10(5):439–44. doi: 10.1080/13607860600637778.

Ellis, Jon. 2022. 'Motivated Reasoning and the Ethics of Belief'. *Philosophy Compass* n/a(n/a):e12828. doi: 10.1111/phc3.12828.

Elster, Jon. 1985. *Sour Grapes : Studies in the Subversion of Rationality*. Cambridge University Press.

Engel, Christoph. 2011. 'Dictator Games: A Meta Study'. *Experimental Economics* 14(4):583–610. doi: 10.1007/s10683-011-9283-7.

Epley, Nicholas, and Thomas Gilovich. 2016. 'The Mechanics of Motivated Reasoning'. *Journal of Economic Perspectives* 30(3):133–40. doi: 10.1257/jep.30.3.133.

Fehr, Ernst, and Klaus M. Schmidt. 2001. 'Theories of Fairness and Reciprocity-Evidence and Economic Applications'. *Available at SSRN 264344*.

Festinger, Leon. 1962. *A Theory of Cognitive Dissonance*. Reissue. Stanford University Press.

Hafer, Carolyn L., and Brenda L. Correy. 1999. 'Mediators of the Relation Between Beliefs in a Just World and Emotional Responses to Negative Outcomes'. *Social Justice Research* 12(3):189–204. doi: 10.1023/A:1022144317302.

Heiserman, Nicholas, and Brent Simpson. 2017. 'Higher Inequality Increases the Gap in the Perceived Merit of the Rich and Poor'. *Social Psychology Quarterly* 80(3):243–53. doi: 10.1177/0190272517711919.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith. 1994. 'Preferences, Property Rights, and Anonymity in Bargaining Games'. *Games and Economic Behavior* 7(3):346–80.

Hoggart, Richard. 1992. *The Uses of Literacy*. London: Penguin Books.

Hufe, Paul, Ravi Kanbur, and Andreas Peichl. 2022. 'Measuring Unfair Inequality: Reconciling Equality of Opportunity and Freedom from Poverty'. *The Review of Economic Studies* rdab101. doi: 10.1093/restud/rdab101.

Jackson, Elizabeth G. 2020. 'The Relationship between Belief and Credence'. *Philosophy Compass* 15(6):e12668.

Janmaat, Jan Germen. 2013. 'Subjective Inequality: A Review of International Comparative Studies on People's Views about Inequality'. *European Journal of Sociology / Archives Européennes de Sociologie* 54(3):357–89. doi: 10.1017/S0003975613000209.

Jerrim, J., and L. Macmillan. 2015. ‘Income Inequality, Intergenerational Mobility, and the Great Gatsby Curve: Is Education the Key?’ *Social Forces* 94(2):505–33.

Jost, John, and Orsolya Hunyady. 2003. ‘The Psychology of System Justification and the Palliative Function of Ideology’. *European Review of Social Psychology* 13(1):111–53. doi: 10.1080/10463280240000046.

Kirk, Dan, Peter Gollwitzer, and Peter Carnevale. 2011. ‘Self-Regulation in Ultimatum Bargaining: Goals and Plans Help Accepting Unfair but Profitable Offers’. *Social Cognition* 29:528–46. doi: 10.1521/soco.2011.29.5.528.

Korenok, Oleg, Edward Millner, and Laura Razzolini. 2017. ‘Feelings of Ownership in Dictator Games’. *Journal of Economic Psychology* 61(C):145–51.

Lerner, Melvin J. 1980. ‘The Belief in a Just World’. Pp. 9–30 in *The Belief in a Just World: A Fundamental Delusion, Perspectives in Social Psychology*, edited by M. J. Lerner. Boston, MA: Springer US.

Lipkus, Isaac M., and Ilene C. Siegler. 1993. ‘The Belief in a Just World and Perceptions of Discrimination’. *The Journal of Psychology* 127(4):465–74. doi: 10.1080/00223980.1993.9915583.

List, John A., and Todd L. Cherry. 2000. ‘Learning to Accept in Ultimatum Games: Evidence from an Experimental Design That Generates Low Offers’. *Experimental Economics* 3(1):11–29. doi: 10.1023/A:1009989907258.

Liu, Xiaoyan, Tong Xie, Wen Li, Yanqiang Tao, Pengwei Liang, Qian Zhao, and Jianping Wang. 2021. ‘The Relationship between Perceived Discrimination and Wellbeing in Impoverished College Students: A Moderated Mediation Model of Self-Esteem and Belief in a Just World’. *Current Psychology*. doi: 10.1007/s12144-021-01981-4.

Major, Brenda, Cheryl R. Kaiser, Laurie T. O’Brien, and Shannon K. McCoy. 2007. ‘Perceived Discrimination as Worldview Threat or Worldview Confirmation: Implications for Self-Esteem’. *Journal of Personality and Social Psychology* 92(6):1068–86. doi: 10.1037/0022-3514.92.6.1068.

Metje, N., H. L. Frank, and P. Croft. 2007. ‘Can’t Do Maths—Understanding Students’ Maths Anxiety’. *Teaching Mathematics and Its Applications: An International Journal of the IMA* 26(2):79–88. doi: 10.1093/teamat/hrl023.

Mijs, Jonathan J. B. 2021. ‘The Paradox of Inequality: Income Inequality and Belief in Meritocracy Go Hand in Hand’. *Socio-Economic Review* 19(1):7–35. doi: 10.1093/ser/mwy051.

Nolan, Brian, Matteo G. Richiardi, and Luis Valenzuela. 2019. ‘The Drivers of Income Inequality in Rich Countries’. *Journal of Economic Surveys* 33(4):1285–1324. doi: 10.1111/joes.12328.

Pillutla, Madan M., and J. Keith Murnighan. 1995. ‘Being Fair or Appearing Fair: Strategic Behavior in Ultimatum Bargaining’. *The Academy of Management Journal* 38(5):1408–26. doi: 10.2307/256863.

Pull, Kerstin. 1999. *What Is the Fair Wage? A Model of as-If-Co-Operation*. Quint-Essenzen.

Ruffle, Bradley J. 1998. ‘More Is Better, but Fair Is Fair: Tipping in Dictator and Ultimatum Games’. *Games and Economic Behavior* 23(2):247–65.

Russell, Kirsten J., and Christopher J. Hand. 2017. ‘Rape Myth Acceptance, Victim Blame Attribution and Just World Beliefs: A Rapid Evidence Assessment’. *Aggression and Violent Behavior* 37:153–60. doi: 10.1016/j.avb.2017.10.008.

Schaafsma, Juliette. 2013. ‘Through the Lens of Justice: Just World Beliefs Mediate Relationships between Perceived Discrimination and Subjective Well-Being’. *International Journal of Intercultural Relations* 37(4):450–58. doi: 10.1016/j.ijintrel.2013.04.002.

Solnick, Sj. 2001. ‘Gender Differences in the Ultimatum Game’. *Economic Inquiry* 39(2):189–200. doi: 10.1111/j.1465-7295.2001.tb00060.x.

Taber, Christopher R. 2001. ‘The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?’ *Review of Economic Studies* 68(3):665–91.

Tilly, Charles. 1998. *Durable Inequality*. Berkeley: University of California Press.

Tisserand, Jean-Christian. 2014. ‘The Ultimatum Game, a Meta-Analysis of 30 Years of Experimental Research’.

de Tocqueville, Alexis. 2007. *Democracy in America*. edited by I. Kramnick. New York: Norton.

Whelan, Christopher T., and Richard Layte. 2004. ‘Economic Change, Social Mobility and Meritocracy: Reflections on the Irish Experience’. *Quarterly Economic Commentary* 3:89–108.

Young, Michael Dunlop. 1961. *The Rise of the Meritocracy, 1870-2033 : An Essay on Education and Equality*. Penguin Books.

Chapter 5. Conclusion

Meditation on chance which led to the meeting of my father and mother is even more salutary than meditation on death.

Simone Weil

5.1 Implications of the research

This thesis was motivated by the gap between what social scientists know, or purport to know, about social class – that it is an important and highly consequential source of inequality in contemporary societies – and what most people seem to believe in developed societies – that they are open and fluid meritocracies in which anyone can "make it" if they have the ability and the right work ethic. Hence, I suggested, social class was like the dark matter of society: a powerful yet invisible force which potently shapes social reality.

I proposed that the explanation for this discrepancy was the widespread belief that "class = competence". That is to say: that higher class people are generally more competent than lower class people when it comes to the skills, faculties, and abilities required to achieve economic success. On a theoretical level, I drew identified this belief as a "status belief". Status theorists had discussed the possibility of status beliefs concerning class (Ridgeway and Fisk 2012), though there has been virtually no empirical work on the matter, because of thorny conceptual and empirical problems.

A status belief about class could be highly consequential for inequality since it works through many several mechanisms. Firstly, it biases third-party assessments of individual ability, meaning that higher class people are more likely to receive rewards, remuneration, and access to valuable opportunities. Secondly, it affects the self-assessed ability of higher class people (positively) and lower class people (negatively), which both affects performance and influences decisions about what path to pursue in life. Thirdly, the belief legitimates what might otherwise be seen as unfair distributions of rewards. The effects of the belief will also be largely invisible because it appears to be merely the common sense claim that those who enjoy economic success have achieved their position on the basis of talent and effort.

The three empirical chapters of the paper were dedicated to addressing each of the three mechanisms identified in the above paragraph. The first chapter addressed Research Question 1: Is there a widely-held belief that "class = competence", controlling for objective indicators

of productivity associated with class position? It was specifically concerned with the question of *discrimination* – is there a cognitive bias which causes people discriminate against lower class people, net of the (close) correlation between class and conventional indicators of productivity? As discussed in that chapter, in order to answer this question it was necessary to design an experiment capable of disentangling status and statistical mechanisms. The results show that there is indeed status discrimination in the case of social class, which proves the existence of status beliefs about class.

The second empirical chapter was designed to answer the Research Question 2: Given that there are status beliefs about class, how consequential are they for inequality? Even if an effect is found to be substantive in an experiment, it is hard to know if it will really matter all that much in reality. But the problem with investigating status "in the wild" is that it is difficult to identify situations where status varies and yet the underlying objective characteristics which generate an individual's status do not. My solution to this conundrum was to point out that status beliefs are attached to perceived social class. And perceived social class is partially determined by an individual's position within the socioeconomic distribution of their reference group. Hence, a lower class child in a higher class classroom will be perceived as "more" lower class than if they are in a lower class classroom. In this chapter I provide evidence that there is a status effect – lower class students in higher class classrooms are less likely to expect to go to university, and this effect dominates the positive effects that are supposed to come from having upper class peers. The effect was large. And since a university qualification is required for the majority of well-paying jobs, the effect of status on unequal university attendance is one pathway in which the "class = competence" belief has consequential, yet invisible, effects on class inequality.

The third paper was designed to answer Research Question 3: How does this the "class = competence" belief arise? Expressing the view that class = competence essentially means that an individual believes they live in a meritocracy – a society in which economically successful people are more talented and hardworking than their poorer counterparts. I argued that the pattern of belief in meritocracy across countries was incompatible with objective evidence on the actual degree of meritocracy in those societies. Following an argument originally developed by Jon Elster (1985), I put forward the argument that belief in meritocracy could be explained as a coping strategy for cognitive dissonance. Essentially, when people perceive that the income distribution is unjust, they experience cognitive dissonance. They can follow

an action-based strategy to reduce cognitive dissonance, by taking action to change the income distribution. Or they can follow a belief-based strategy and convince themselves that the distribution is fair after all. If the first course of action is blocked, they will have to choose the latter. The empirical results did not support this argument. However, they did suggest that people who experience large levels of inequality and can't do anything about it are more likely to judge that inequality as fair – potentially because being deprived of agency makes them adopt an attitude of resigned indifference.

In sum, the research seems to indicate that status beliefs about class make a substantive, yet largely invisible, contribution to class inequality. Is there a prospect of the beliefs reducing, or being vanquished altogether? As Ridgeway has pointed out (Ridgeway 2019), status beliefs ultimately have their origin in material differences between social groups. There is some evidence, for example that status beliefs about gender have diminished in recent time (Eagly et al. 2020b). This may be due to decreasing material inequality between men and women in developed countries: the mass entry of women into the labour market (Fernández 2013; Thévenon 2013), the substantive if incomplete reduction in the gender wage gap (Goldin 2014; Goldin et al. 2006), and the fact that girls now tend to outperform boys academically (Clancy and O'Sullivan 2020; Vincent-Lancrin 2008).

By contrast, income inequality and social mobility are increasing in the developed world (Durlauf and Seshadri 2018; Hufe et al. 2022; Nolan et al. 2019). The material basis for status beliefs concerning class – large gaps in economic resources – has therefore widened. Will the gap shrink again? The traditional basis for working class people to shrink the gap was through trade union agitation, to redistribute the economic surplus towards labour, and supporting parties which would redistribute income through the tax and welfare systems. However, union membership has declined substantially in the past few decades (Schnabel 2013; Waddington 2015). And the political parties which traditionally mobilised working class votership in the pursuit of aggressive redistribution and other pro-labour policies have either declined or adopted more centrist political programmes which have caused them to shed working class support (Bremer 2018; Rennwald and Pontusson 2021).

The results of the final empirical chapter of this thesis would suggest that, in the absence of the traditional institutional machinery which allowed the working class to exert influence over the income distribution, a significant portion of them will adopt an attitude of resigned

indifference to class inequality. What are the prospects for a revival of working class collective agency? History teaches us that in circumstances of deep inequality coupled with a generalised absence of collective struggle, radical millenarian movements may emerge to generate the animating fervour required to mobilise a previously dormant population (Hobsbawm 1971). It remains to be seen whether such forces will emerge, or whether a revival of collective agency may occur due to more prosaic, or more impersonal, processes of social change.

To conclude with the scientific contribution, I believe that, though the field of social inequality research is certainly crowded, the results of this thesis illustrate a powerful yet virtually unexplored mechanism driving inequality between social classes. This opens the possibility for a substantive, fruitful, and varied research agenda. What other areas besides university attainment are affected by status beliefs concerning classes? Can the historical and geographical evolution of class-based status beliefs be measured? And what does really give rise to such beliefs? These are only some of the questions that could be addressed. As such, the results presented here can only be regarded as preliminary and suggestive. If there is virgin soil to be tilled, then the extent and contours of its landscape are as yet only dimly revealed.

5.2 Reflection on methodology

I have seen it said that the product of a PhD is not the empirical results but the candidate themselves. To the extent that this is true, I hope it will be judged permissible to adopt in the following comments a somewhat more personal tone. At the very least, the element of self-reflexivity may add some relevance and freshness to the remarks which I would not otherwise claim to be particularly original.

Through the process of instruction and collaboration which facilitated the writing of these chapters, I have tried to learn how to more efficiently produce and present scientific knowledge. To this end, I found the experience of working with senior colleagues on the paper that became Chapter 2 to be a particularly instructive experience. Thanks to their guidance, through revision after revision, I learned how to formulate an argument clearly, so that conclusions will evidently follow from premises and that the claims will be empirically sound. I learned how to formulate hypotheses and how to test them; how to persuade without covering up; what to put in and what to leave out. Or at least, I tried to learn these things and,

to the extent that I made progress in this regard it is thanks to this collaboration and to the instruction I have received from my supervisor.

As a corollary, I have also striven to adopt an analytic, deductive, parsimonious – and abstract – style of reasoning. While ultimately a "style" rather than an objective technique or method, I believe that the rigour and clarity of such a style has much to recommend to it, and facilitates the type of clear, concise, valid, and sound argumentation which I came to value. I tried, particularly in earlier drafts, to make Chapter 3 conform to the same approach. And while I found such an approach fruitful, I also found myself wondering if it relied too much on assumptions and on a priori reasoning and less on a qualitative exploration of the data which would seem to be tacitly forbidden by the analytical-deductive approach.

My experience of writing Chapter 4 magnified the heft of this intuition. Although I think the theoretical argument in this chapter was reasonably well-constructed, and based, after all, on the work of a great analytic thinker, the data did not support the theorised mechanism. One result does not invalidate a theory, of course, nor does the invalidation of a theory invalidate the whole analytical-deductive approach. Indeed, it is an essential aspect of it. Nonetheless, I think if I had relied purely on the analytical-deductive approach in this final empirical chapter, I would ultimately have shed less light on the topic under investigation. The provision of even a small quantity of qualitative data, consisting of the participants' open-ended rationales for their fairness judgments, revealed a richness that had not been predicted by the theory, and allowed for a tentative alternative explanation where the theory would have provided none. If the analytical-deductive approach has the ability to cast a razor-sharp shaft of light into the murky realm of the unknown, this narrow diameter of this high-powered beam may nonetheless fail to reveal the object of inquiry. The interpretative-inductive approach emits a more diffuse radiance which may not clearly throw into relief the object of inquiry, but may be more likely to illuminate it.

Here one of the core themes of the thesis – the ubiquitous uncertainty of the social world – must reflect back on the researcher and their method. The goal of the social researcher is to unearth and explain what is uncertain and unknown about human society. And as a socially relevant concept, uncertainty has played an important role in each chapter; in Chapter 2, serving as foil for status theory; in Chapter 3, creating the need for reference groups as a heuristic; and in Chapter 4, as a scope condition for wishful thinking. But the researcher

cannot hold themselves exempt from the limitations that uncertainty places on their thinking. They too are ultimately engaged in constructing heuristics to understand a world that is always in flux, and they are themselves vulnerable to wishful thinking about the explanatory power of their theories. They are in the end, no matter how deep their knowledge nor how sophisticated their method, also "naïve scientists".

The epigraph to the Introduction of this thesis is taken from Spinoza's Ethics: "I shall consider human actions and appetites just as if I were considering lines, planes, or bodies." The Ethics is a masterwork of a logical rigour, wherein Spinoza seeks to deduce "geometrically", from a short list of axioms, profound conclusions about metaphysics, human nature, ethics, and free will. His epistemological wager, nicely captured in this aphorism, is that the type of deductive and strictly logical analysis proper to mathematics could equally well be applied to the study of human beings. Modern social scientists are rightly more modest in their epistemological claims. But a (diluted) version of the same epistemology remains in place. The analytical-deductive style requires a "geometric" treatment of human desires, beliefs, and understandings.

While this approach has yielded beautiful theories and deep insights into human nature and activity, it is also vulnerable to the simple, brute, and inescapable fact of uncertainty. As I argue in the Introduction, the objectivity of the social world consists in the aggregate of subjective beliefs and actions of human beings endowed with free will, and it is precisely this creative, diverse, and unexpected ways in which free will is disposed of which make the social world itself uncertain.

Hence, the Simone Weil quote that heads this concluding chapter provides an apt counterpart to the Spinozan epigraph. Meditation on chance, and how it can unravel the tightest knots which logic may weave, *is* salutary. If causation is the holy grail of social science, then it is quite likely to be the analytic-deductive approach that will get us there in the end. But it is worth reflecting on the fact that the adoption of more rigorous, more objective methods can also lead to the destruction of knowledge. This paradox is beautifully illustrated in a paper by the economist Paul Krugman, which never been published in a peer-reviewed venue, but is available, in bare web 1.0 format, on the MIT website (Krugman n.d.).

In this paper, The Fall and Rise of Development Economics, Krugman seeks to explain why A.O Hirschman's theory of economic development failed to win favour in mainstream

economics when it was first put forward in the ‘40s and ‘50s. Krugman’s explanation is that the theory was, at that time, untranslatable into the language of mathematical economics, which had become the sole acceptable expression of economic arguments within the academy. At the end of this paper, Krugman provides a sketch of a formal model to illustrate that the increasing returns mechanism, which Hirschman’s theory relied upon, could now be represented using the standard mathematical techniques which economics had developed in the following decades. But too late for Hirschman, whose insights were ignored at the time. (Though he, of course, made influential contributions in other fields [Hirschman 1970]).

In between, Krugman reflects on the nature of scientific models, taking, as his paradigmatic example, the case of African cartography. This example was inspired by a friend of Krugman’s, a collector of “antique maps”, who had supposedly written a paper entitled "The evolution of European ignorance about Africa". (In a pleasingly Borgesian twist, no such paper actually seems to exist). Krugman suggests that the uninformed reader might expect that as European cartography developed, maps of Africa would become both more accurate *and* more detailed. But this not the case. As techniques for measuring and graphically representing geographical distances improved, the African coastline – the physical feature of the continent most readily accessible to the rapacious Europeans – was mapped with ever-increasing accuracy. By the 18th century, the coastal representation was practically as precise as it is on modern day maps. But at the same time as the coastlines grew more precisely delineated, the interior emptied out. Maps showed a very accurately drawn outline with a large blank interior suggesting total ignorance of the vast majority of the continent.

The newer and more rigorous cartographic methods could only be applied to chart areas that Europeans already had access to. The features of the interior which had been displayed on previous maps had been largely derived from travellers’ tales and the second-hand reports of merchants and ambassadors to the urbanised coastal peripheries. This knowledge was necessarily vague. It did not meet the new standard for scientific evidence. In one respect there was an improvement – old maps showed pictures of monsters and mythical creatures that supposedly inhabited various parts of the continent, besides more plausible geographical features which didn’t actually exist such as Ptolemy’s Lake Zaire. But in other respects, it represented a loss of information. Whole rivers and cities – which really existed – disappeared from the canvas and page. The novel technology of compass and theodolite, and novel techniques of coordinate geometry and trigonometric calculation, led to the creation of

new, more precise, more scientific knowledge. But they also destroyed the rougher, vaguer, and arguably richer type of qualitative knowledge which they could not assimilate. The consequences were not merely scientific; the perception of the blank interior, which erased the peoples of the continent, prepared the way for their physical erasure by Britain and the other colonial powers that wished to paint the white space into neatly demarcated zones of control and exploitation.

Likewise, because mathematical economists had focussed on modelling perfect competition and constant returns, Hirschman's theory, which relied on increasing returns to scale, could not be modelled. Hence, it was nonsense to the rigorous but restrictive language of mainstream economics in the postwar period. Hirschman's theory was based on his own practical experience working with the National Planning Board of Colombia, which had produced qualitative knowledge which could not be readily assimilated to the models produced in American universities. Hirschman himself was not unaware of the methodological and epistemic stakes. He argued against what he saw as excessive emphasis on parsimony in model construction, and for deliberate complication of economic theory (Hirschman 1984).

The analogy between the cartography of Africa and the analytical-deductive versus interpretative-inductive styles is, perhaps, obvious. But in some ways the analogy is deficient. It portrays qualitative knowledge as a "second-best" alternative, to be used as a stand-in when more precise methods are yet unavailable or intractable. And by focussing on technique, it ignores the more important question of theory. The moral of the parable can be improved by upon by the addition of a second one, also in keeping with the theme of European maps of Africa.

While European cartographical science was emptying the African interior of ghouls and monsters, as well as actually existing, if hitherto imprecisely indicated, geographical features, at least one fantastical entity remained: a vast mountain range stretching across the western part of the country at a latitude roughly coterminous with the geographical region known as the Sahel. The mountain range does not exist, but was commonly depicted on maps throughout the 19th century (Bassett and Porter 1991). These fantastical mountains were termed "The Mountains of Kong". (Note: most of the following account is based on Bassett and Porter [1991]).

In an account of his expedition through West Africa, Mungo Park, a British explorer, claimed to have seen, from a distance, a range of mountains situated in "a large and powerful kingdom called Kong" (cited in [Bassett and Porter 1991, 377]). Based on Park's account, a map was drawn up by James Rennell, a prominent contemporary geographer. This was the first map to feature the Mountains of Kong. But the rationale for the placement of the mountain was due to more than just the passing comment of an explorer who had seen the peaks from a very great distance. It had already been theorised by previous geographers that there should be some elevated land in the area where the mountains were supposedly situated, in order to serve as a drainage divide between the Niger river and the coastal water basin. Rennell's placement of the mountains in that particular location complemented his own theory of the course of the river Niger, and in defence of this theory he provided calculations of the volume and evaporation rate of the rivers waters. Hence, the location of the mountain was supported by hydrological theory, mathematical calculations, and reputable scientific methods and techniques – though not, of course, personal observation.

Despite the accounts of subsequent explorers of West Africa who did not come across mountains of the size or location that the Mountains of Kong, they continued to be placed on maps for the better part of a hundred years. The rise of colonialism and scientific racism also helped to contribute to the epistemological ignorance. European cartographers and administrators ignored first-hand testimony of native Africans that the mountains did not exist in the places they were claimed to. It was not until the surveying expedition of Louis-Gustave Binger, a French military officer, that the non-existence of the mountains (for Europeans) was definitively established nearly a century after Rennell had placed them on the map: "on the horizon, not even a ridge of hills!" (cited in [Bassett and Porter 1991, 396]). In a reversal of intellectual traditions, French empiricism had trumped British rationalism.

This second parable teaches us that the new scientific techniques of European cartography did not merely destroy imprecise, or vague knowledge. Rather, deductive scientific theory also contributed to the creation of very precise, but totally inaccurate, geographical features. And it was "qualitative" data that contributed to refuting the "theory" of the Mountains of Kong, through rich, first-hand accounts of the actual terrain. By opening the door to other kinds of knowledge, qualitative evidence can undermine the false certainties produced by the analytic-deductive method.

The purpose of this parable is not to disparage theory. Quite the opposite. Probably there is not as much theory as there should be in applied social sciences, and that which is there may benefit from reformulation in an analytic-deductive style. But I think qualitative evidence plays a crucial complementary role for social researchers who are confronted with the inherent uncertainty of the social world. Uncertainty, in its stricter sense, refers to the presence of factors which are unknown or unknowable to the subject. It is in this particular aspect that I think qualitative data is of vital importance. Binger's journey through the Sahel provided evidence for a proposition which most western scholars would not have considered: that the mountains existed only on in theory.

To conclude this digression on a personal note, the experiment reported in Chapter 4 represents my own personal Mountains of Kong. Even the abbreviated, sometimes gnomic responses provided by the respondents to the question about their fairness rationale evinced a creativity and richness which could never be fully anticipated by any theory. In future experiments, I will endeavour to include open-ended questions in order to aid the interpretation of the results. And, although I am not trained in qualitative methods such as interviewing or ethnography, I will consider including such methods in future research projects, either by learning them myself or by collaborating with scholars who are familiar with them. With the aid of these complementary tools – the broad, diffuse radiance of the inductive-interpretative style and the narrow, cutting beam of the deductive-analytic style – I would hope to discover if not whole mountains then at least some foothills of reasonable proportion.

References

Bassett, Thomas J., and Philip W. Porter. 1991. '"From the Best Authorities": The Mountains of Kong in the Cartography of West Africa*'. *The Journal of African History* 32(3):367–413. doi: 10.1017/S0021853700031522.

Bremer, Björn. 2018. 'The Missing Left? Economic Crisis and the Programmatic Response of Social Democratic Parties in Europe'. *Party Politics* 24(1):23–38. doi: 10.1177/1354068817740745.

Clancy, Patrick, and Sara O'Sullivan. 2020. 'Gender Parity in Higher Education Enrolments: Trends and Paradoxes'. *Irish Educational Studies* 39(3):337–54.

Durlauf, Steven N., and Ananth Seshadri. 2018. 'Understanding the Great Gatsby Curve'. *NBER Macroeconomics Annual* 32(1):333–93. doi: 10.1086/696058.

Eagly, Alice H., Christa Nater, David I. Miller, Michèle Kaufmann, and Sabine Sczesny. 2020. 'Gender Stereotypes Have Changed: A Cross-Temporal Meta-Analysis of U.S. Public Opinion Polls from 1946 to 2018'. *American Psychologist* 75(3):301–15. doi: 10.1037/amp0000494.

Elster, Jon. 1985. *Sour Grapes : Studies in the Subversion of Rationality*. Cambridge University Press.

Fernández, Raquel. 2013. 'Cultural Change As Learning: The Evolution of Female Labor Force Participation over a Century'. *American Economic Review* 103(1):472–500.

Goldin, Claudia. 2014. 'A Grand Gender Convergence: Its Last Chapter'. *American Economic Review* 104(4):1091–1119.

Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. 2006. 'The Homecoming of American College Women: The Reversal of the College Gender Gap'. *Journal of Economic Perspectives* 20(4):133–56.

Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Vol. 25. Harvard university press.

Hirschman, Albert O. 1984. ‘Against Parsimony: Three Easy Ways of Complicating Some Categories of Economic Discourse’. *Bulletin of the American Academy of Arts and Sciences* 37(8):11–28. doi: 10.2307/3823226.

Hobsbawm, E. J. (Eric J.). 1971. *Primitive Rebels : Studies in Archaic Forms of Social Movement in the 19th and 20th Centuries*. [Third edition]. Manchester University Press.

Hufe, Paul, Ravi Kanbur, and Andreas Peichl. 2022. ‘Measuring Unfair Inequality: Reconciling Equality of Opportunity and Freedom from Poverty’. *The Review of Economic Studies* rdab101. doi: 10.1093/restud/rdab101.

Krugman, Paul. n.d. ‘The Fall and Rise of Development Economics’. Retrieved 30 May 2022 (https://web.mit.edu/krugman/www/dishpan.html).

Nolan, Brian, Matteo G. Richiardi, and Luis Valenzuela. 2019. ‘The Drivers of Income Inequality in Rich Countries’. *Journal of Economic Surveys* 33(4):1285–1324. doi: 10.1111/joes.12328.

Rennwald, Line, and Jonas Pontusson. 2021. ‘Paper Stones Revisited: Class Voting, Unionization and the Electoral Decline of the Mainstream Left’. *Perspectives on Politics* 19(1):36–54. doi: 10.1017/S1537592720000067.

Ridgeway, Cecilia L. 2019. *Status*. New York: Russell Sage.

Ridgeway, Cecilia L., and Susan R. Fisk. 2012. ‘Class Rules, Status Dynamics, and “Gateway” Interactions’. Pp. 131–51 in *Facing social class: How societal rank influences interaction*. New York, NY, US: Russell Sage Foundation.

Schnabel, Claus. 2013. ‘Union Membership and Density: Some (Not so) Stylized Facts and Challenges’. *European Journal of Industrial Relations* 19(3):255–72. doi: 10.1177/0959680113493373.

Thévenon, Olivier. 2013. ‘Drivers of Female Labour Force Participation in the OECD"’. *OECD Social, Employment and Migration Working Papers* (145). doi: 10.1787/5k46cvrgnms6-en.

Vincent-Lancrin, Stéphan. 2008. ‘The Reversal of Gender Inequalities in Higher Education: An On-Going Trend’.

Waddington, Jeremy. 2015. ‘Trade Union Membership Retention in Europe: The Challenge of Difficult Times’. *European Journal of Industrial Relations* 21(3):205–21. doi: 10.1177/0959680114538708.