

Social class bias in welfare sanctioning judgements: Experimental evidence from a nationally representative sample

Robert de Vries¹  | Aaron Reeves²  | Ben Geiger¹ 

¹School of Social Policy, Sociology, and Social Research, University of Kent, Canterbury, UK

²Department of Social Policy and Intervention, University of Oxford, Oxford, UK

Correspondence

Robert de Vries, School of Social Policy, Sociology, and Social Research, University of Kent, Canterbury, UK.

Email: r.devries@kent.ac.uk

Funding information

University of Kent

Abstract

In this study, we examine whether welfare deservingness judgements in the United Kingdom are affected by a bias against claimants from stigmatised social class backgrounds. In the United Kingdom, as in other countries, stereotypes of a perceived social ‘underclass’ are widespread. Political and media discourse frequently portrays members of this ‘underclass’ as lazy, feckless and not genuinely in need of support. Yet despite strong academic interest in perceived welfare deservingness, existing research has largely neglected the role of social class bias in deservingness judgements. To address this gap, we use a novel vignette experiment administered to a representative sample of British respondents to provide the first direct evidence of discrimination against welfare claimants with ‘underclass’ signifiers. We find that the British public are more likely to endorse a sanction against a claimant from an ‘underclass’ background than against an otherwise identical claimant from a less stigmatised class background. We also asked respondents to justify their decisions and, applying computational methods to analyse these free-text responses, we find that ‘underclass’ claimants are more likely to be blamed for violating the conditions of their benefit, while claimants from other class backgrounds are more often given the ‘benefit of the doubt.’ Our findings have important implications for our understanding of the relationship between social class

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. Social Policy & Administration published by John Wiley & Sons Ltd.

background and public deservingness perceptions, and potentially for the differential treatment of claimants by the benefits system.

KEYWORDS

benefits, bias, deservingness, discrimination, sanctions, social class, underclass, welfare

1 | INTRODUCTION

Contemporary welfare systems increasingly emphasise labour market activation and re-integration (Van Kersbergen & Hemerijck, 2012; Watts & Fitzpatrick, 2018). This ‘activation turn’ has produced a dramatic increase in the degree of conditionality attached to claiming unemployment benefits (Watts & Fitzpatrick, 2018). Inherent in such conditionality is an underlying logic of discipline and punishment (Deeming, 2015; MacLeavy, 2011). Unemployed claimants must adhere to strict ‘conditions of conduct’ in order to continue receiving welfare payments—for example, providing evidence of extensive job search activities, and participating in work placement programmes (Clasen & Clegg, 2007). Deviations from acceptable conduct are punished through partial or full withdrawal of payments—a benefit ‘sanction’ (Adler, 2016). In the United Kingdom for example (our focus in this study), this has been formalised through the ‘claimant commitment,’ an agreement signed by the claimant that outlines the steps they will take to find work (UK Department for Work and Pensions, 2020a). Failure to adhere to these commitments represents grounds for sanction.

The punitive logic of conditionality is closely linked to conceptions of welfare deservingness. ‘Good behaviour’ (adherence to benefit conditions) demonstrates that a claimant is deserving of continued welfare support; while ‘bad behaviour’ suggests that they are undeserving (Redman, 2020). Perceptions of deservingness are coloured by bias and discrimination; with group stereotypes informing harsher or more generous deservingness judgements (Larsen, 2008). For example, groups stereotyped as ‘lazy,’ and therefore more blameworthy are judged to be less deserving of support (Larsen, 2008). Such stereotypes partially explain lower levels of support for welfare payments to young people (Larsen, 2008; Van Oorschot, 2000), immigrants (Heuer & Zimmermann, 2020; Van Oorschot, 2006) and potentially people of colour (though evidence on this is mixed: Ford, 2016; Kootstra, 2016). In this study, we add to a small but growing literature investigating how such biases affect sanctioning judgements specifically (Buss, 2019; Houtman, 1997; Naumann et al., 2020; Schram et al., 2009; Terum et al., 2018). Are claimants from stigmatised groups punished more harshly than other claimants for the same ‘bad behaviour’?

Previous research on bias in sanctioning judgements has found that, all else being equal, younger claimants are judged more deserving of sanctions than older claimants (Buss, 2019; Houtman, 1997; Naumann et al., 2020), and potentially that claimants of colour are also judged more harshly (though again the evidence is mixed: Schram et al., 2009; Buss, 2019; Terum et al., 2018). We also examine potential age, gender and ethnic biases; however, our primary focus is on the role of social class stereotypes in sanctioning judgements. While the existing deservingness literature is extensive, there has been remarkably little theoretical or empirical work on the role of social class stereotypes. This is striking because such stereotypes map strongly onto well-explored deservingness criteria. In particular, members of a perceived ‘underclass’ are often stigmatised as feckless, workshy and not genuinely in need (Hayward & Yar, 2006; Jones, 2011). Claimants with ‘underclass’ signifiers may therefore be judged uniquely undeserving of state support, and potentially highly deserving of punishment in the form of sanctions.

We begin by articulating a theoretical basis for class-based discrimination in deservingness perceptions, and specifically in sanctioning judgements. We then test such discrimination in public sanctioning attitudes using an experimental vignette study. In a representative sample of the UK population, we find a significantly higher level of

support for sanctioning a hypothetical claimant with 'underclass' signifiers, than for sanctioning an otherwise identical claimant from a non-stigmatised class background. This evidence of class-based discrimination in sanctioning judgements has important implications for how we understand the drivers of sanctioning support, and of deservingness perceptions more generally (Monnat, 2010).

1.1 | Who deserves to be sanctioned?

The core propositions of existing welfare deservingness research are (i) that public perceptions of deservingness are based on relatively consistent criteria (such as whether a potential recipient is genuinely in need, and whether they are perceived to be at fault for their circumstances) (e.g., Van Oorschot, 2000); and (ii) given that cues to these criteria are often invisible or ambiguous, deservingness judgements are commonly based on group stereotypes (Aarøe & Petersen, 2014). For example, popular stereotypes of young people as lazy and ungrateful mean they are widely perceived as relatively undeserving of welfare support (Larsen, 2008).

There are a number of different ways in which 'deservingness' attitudes can be conceptualised and measured, including: access (whether a claimant or group deserves access to welfare support at all), generosity (how much support they should receive), and—our focus here—punitive conditionality (under what circumstances should claimants be punished for violating the conditions of their benefit). Existing deservingness research has not drawn a strong theoretical distinction between these different types of attitudes. The small literature focusing on sanctioning judgements in particular has tended to assume that such judgements are based on essentially the same logic and criteria as other deservingness attitudes (e.g., Buss, 2019; Naumann et al., 2020).

Considerable overlap between sanctioning judgements and other deservingness attitudes does appear likely, given that both involve deciding how much someone should be supported by the state. From this perspective, withholding 'new' benefits, is not qualitatively different from stopping or reducing payments to an existing claimant. However, sanctions are different from other welfare judgements in that they are explicitly punitive—a sanction is a punishment for violating a rule (and the associated principle). This element of sanctions may invoke a different mindset than more passive judgements about access and generosity (Houtman, 1997). Certain deservingness-related stereotypes may also be more or less salient when rendering punitive judgements (e.g., Schram et al., 2009 on racial bias in punitive social policy decisions).

If their punitive nature makes sanctioning attitudes to some extent qualitatively different from other deservingness attitudes, then the centrality of punitive logic to contemporary 'activation based' welfare policy makes research on such attitudes particularly important. However, as we have noted, relatively little research on deservingness has adopted this focus. Evidence from this small literature suggests that, first, younger claimants are deemed more deserving of sanctions than older claimants (Buss, 2019; Houtman, 1997; Naumann et al., 2020), as are claimants who appear less motivated to find work (Buss, 2019). These findings are consistent with robust evidence from the wider deservingness literature. Evidence for the effects of other claimant characteristics is more mixed, with some studies suggesting an ethnic bias in sanctioning judgements (Schram et al., 2009), and others not (Buss, 2019; Terum et al., 2018). Buss' (2019) work is notable in that he measures the effect of a variety of claimant characteristics on both welfare generosity and sanctioning judgements. He finds that non-White ethnicity has a negative effect on welfare generosity, but no effect on sanctioning judgements—supporting a potential detachment between these two types of deservingness judgements. To our knowledge, no existing work on sanctioning judgements has examined the effect of claimant social class.

1.2 | Social class stereotypes, deservingness and sanctioning judgements

As we note, very little work on welfare deservingness discusses the potential effect of social class. This may be due to the overlap between the concepts involved—particularly when considering working age, non-disabled people. The deservingness literature seeks to understand which groups of poor people are considered more or less deserving

of state support (Van Oorschot, 2000). If being poor or unemployed is considered to determine one's social class, then all who are potentially eligible for welfare are perforce part of the same social class. Hence the question of how social class affects deservingness perceptions is unlikely to arise. However, people who are currently in need of state support inhabit many different class positions, and come from a variety of class backgrounds (Geiger, 2018). This variety will often be reflected in visible, class-marked differences in lifestyle and presentation (including residential disadvantage, clothing, accent, smoking behaviour and other class signifiers).

The potential effect of social class on deservingness judgements depends on the relationship between social class stereotypes and deservingness criteria. In European deservingness research, Van Oorschot's (2000) CARIN criteria are widely adopted. However, the CARIN model is not universally accepted and others, such as the RICE model (Knotz et al., 2020) have suggested amended criteria. Criteria common to both the CARIN and RICE models are:

- *Reciprocity*: to what extent claimants are considered to be contributors to the system they are now 'withdrawing' from. This may be conceived in terms of past contributions only (RICE) or both past and potential future contributions (CARIN).
- *Identity*: to what extent claimants are perceived to be a socially separate outgroup.
- *Control*: how much control claimants are perceived to have over the circumstances leading to their need for support. Under the CARIN model, *Control* refers to both the initial reason for need (e.g., control over a job loss) and to reasons for ongoing need (e.g., control over failure to find work). Under the RICE model, *Control* refers only to the former, and the latter is theorised as a separate *Effort* criterion.

Under both models there is considerable overlap between class-based (and particularly 'underclass') stereotypes and deservingness criteria. Perceptions of laziness and graft are central to stereotypes of 'the underclass'; particularly in the United Kingdom (Baumberg et al., 2012; Hayward & Yar, 2006; Jones, 2011; Kallschmidt & Eaton, 2019). It is these stereotyped characteristics which are most strongly reinforced in political and public discourses around welfare, which valorise 'hard working people' while demonising 'workshy scroungers' (Baumberg et al., 2012; Okoroji et al., 2021). These characteristics relate very strongly to *Control/Effort* criteria—with an unwillingness to work placing the claimant at fault for needing state support.

The relationship between class-based stereotypes and established deservingness criteria leads to a clear expectation: that claimants from stigmatised 'underclass' backgrounds will be seen as more to blame for their circumstances, with this being a strong driver of their perceived (un)deservingness for welfare. There is limited existing evidence of a link between social class and deservingness—comprising only two studies of which we are aware. Although primarily focused on ethnicity and migration status, Kootstra (2016) found that claimants who were previously cleaners were judged to deserve lower levels of welfare support than claimants who were previously senior managers. By contrast, Reeskens and van der Meer (2019) found that claimants with a lower previous income were judged to deserve higher levels of benefit—though they were also more likely to be awarded no benefit at all. In both studies, class manipulations were weak—with claimant social class indicated only by a single class-linked characteristic (occupation or income). Crucially, neither study specifically manipulated 'underclass' signifiers or examined sanctioning judgements.

The implications for sanctioning judgements of the link between class-based stereotypes and deservingness criteria are complicated by the extent to which such judgements may depart from the logic and criteria of other, more widely researched, deservingness measures (such as access and generosity). However, we argue that any such departures are unlikely to weaken the expectation that 'underclass' claimants will be judged more harshly than claimants from less stigmatised class backgrounds. Sanctioning judgements are punitive, and punitive judgements, more so than other deservingness judgements, are likely to rest on perceptions of *fault* (Houtman, 1997). Is a potentially sanctionable offence likely to be the fault of the claimant, or might it be a product of genuine extenuating circumstances? As such, stereotypes of 'workshyness' appear unlikely to have a weaker effect on sanctioning judgements than they would have on other deservingness judgements. On this basis, we propose the following hypotheses:

Hypothesis 1. *Public sanctioning judgements are subject to class-based bias, such that claimants with 'underclass' signifiers are judged more deserving of sanctions than otherwise identical claimants from non-stigmatised class backgrounds.*

Hypothesis 2. *When reasoning about sanctioning judgements, people attribute a greater degree of blame/control to claimants with 'underclass' signifiers than to claimants from non-stigmatised class backgrounds.*

1.3 | The present study

We test the above hypotheses using a factorial vignette experiment in a nationally representative sample of respondents in the United Kingdom. In these vignettes, we manipulate the perceived social class of a hypothetical unemployed welfare claimant who has breached the conditions of their benefit. In particular, we manipulate two class-linked characteristics (parental occupation and current housing situation) to signify membership of either (i) the professional middle-class, (ii) the 'reputable' working class or (iii) the 'underclass.' Our study is therefore the first to directly examine the role of 'underclass' signifiers in sanctioning judgements (and indeed in deservingness judgements more generally). We also manipulate three other demographic factors commonly investigated in the deservingness literature—age, gender and ethnicity (see above)—in order to compare any observed social class effect with the effect of these characteristics.

This study also goes beyond previous quantitative research in this field by asking respondents to describe, using a free-text response, *why* they did or did not support a hypothetical sanction. To our knowledge, ours is the first study to use computational methods to examine large-scale qualitative data on such deservingness judgements.

2 | METHODS

2.1 | Sample

To test our hypotheses, we commissioned the polling company *YouGov* to conduct a survey using their opt-in panel of active users. To achieve an approximately nationally representative sample, panel members were invited to participate according to sociodemographic criteria, and the resulting sample was then weighted to be representative of the UK population in terms of age, gender, occupational class (see below for the class measure used), government office region, and level of education. Ethical approval was given by the lead author's institution; the final survey was conducted in January 2020 and included 1519 respondents.

2.2 | Experimental procedure

Participants first read a vignette describing a hypothetical claimant of unemployment benefits who had committed a potentially sanctionable infraction. We described the claimant as someone who had lost their job and had been claiming benefits for 3 months. They initially seemed highly motivated to find work, but their motivation appeared to have declined over time. We described the claimant as having recently missed the second of two mandatory meetings with their welfare caseworker.

In the United Kingdom, attending scheduled caseworker meetings is part of the commitment claimants must make in order to begin receiving benefit payments. Failure to attend such meetings without a 'good reason' (Department for Work and Pensions, 2020b) therefore represents grounds upon which a caseworker may refer a claimant for sanction. To date, missing a caseworker meeting is the most common reason for sanction referral (data retrieved from StatXplore,¹ June 2021). In our vignette, we describe our hypothetical claimant as offering one of

TABLE 1 Social class signifiers used in the vignette

	Parental occupation	Housing situation
Professional middle class	Both parents are architects	A suburban home
'Reputable' working class	Father is a plumber, mother is a 'dinner lady' ^a	A flat
'Underclass'	Both parents are long-term unemployed and claiming benefits	A council estate

^aA colloquial term for a school kitchen assistant.

three exculpatory reasons for missing their most recent meeting: either (i) a bad cold, (ii) failure to receive the appointment letter or (iii) their bus was very late. These reasons were randomised across vignettes to ensure that our results were generalisable beyond a single specific reason.

In this vignette, we aimed to describe a scenario which (i) matched what a welfare caseworker might realistically encounter and (ii) provided reasonable grounds to question the claimant's exculpatory justification (this being the second missed appointment, alongside an observation of apparently declining motivation).

To test the effect of the hypothetical claimant's class background on responses to this scenario, we manipulated two class signifiers: (i) the occupation of the claimant's parents and (ii) the claimant's current housing situation. Specific details of the manipulation are given in Table 1, above.

'Council estates' in the United Kingdom comprise housing that was originally owned by local authorities; often medium or high-rise blocks in urban areas. Properties in 'council estates' are now often privately rented or owned by residents (Hanley, 2017). However, they are still highly stigmatised in public discourse as being associated with a problematic underclass (Hanley, 2017). Multi-generational worklessness is also a common element of the discourse around this 'underclass' (Macmillan, 2011).

By employing two social class cues, we aimed to increase the salience of class in the vignette, as well as increasing the realism and vividness of the vignette as a whole (Petersen & Aarøe, 2013).

In addition to social class, we also manipulated three other demographic characteristics of the claimant:

- *Ethnicity (Black/White)*. The claimant was given a name more common among White British people (Hannah/Andrew) or more common among Black Britons of Nigerian descent (Hassana/Abeo). The claimant was also explicitly described as either Black British or White British. To further increase the salience of ethnicity, the claimant's parents were either described as having been born in London (for the White British claimant), or having migrated from Nigeria before the claimant was born (hence, all claimants were themselves born in Britain).
- *Age (19/35/55)*. The claimant was described as either being 19, 35 or 55.² In order to increase the salience of the claimant's age, we also inserted a reference to the claimant's 19th, 35th or 55th birthday having recently passed.
- *Gender (Male/Female)*. The claimant was given either a male (Abeo/Andrew) or female (Hannah/Hassana) name, and corresponding gender pronouns were used throughout the vignette.

All five of the above characteristics were varied randomly across surveys. Hence respondents saw one of 108 possible unique vignettes. A full example vignette is given in online supporting information (Web Appendix A).

2.3 | Measures

After reading the vignette, respondents were asked to what extent they agreed that the named claimant 'should be sanctioned (i.e., have their benefits stopped temporarily),' giving answers on a scale of 1 (Strongly agree) to 5 (Strongly disagree). Responses were dichotomised for the purposes of the analyses reported below (1 = agree/strongly agree, 0 = all other valid responses).

The purpose of the random experimental manipulation is to purge our estimate of the social class effect from omitted variable bias due to respondent characteristics (or due to other characteristics of the hypothetical claimant) (Mutz, 2011). However, it is possible for a random process to produce an incidental correlation between these factors and the experimental manipulation of interest. To ensure that our estimate was robust to this possibility, we collected data on the following socio-demographic characteristics of respondents (sourced primarily from YouGov's background data on panel members, collected prior to the survey):

- Age in categories: 18–24, 25–39, 40–64, 65+
- Gender (male/female)
- The *occupational status* of the highest income earner in the household, as measured by National Readership Survey (NRS) social grade (NRS, 2018). We contrasted: (i) grades A/B (higher or intermediate managerial, administrative and professional occupations), (ii) C1 (supervisory or clerical, junior managerial, administrative or professional occupations)/C2 (skilled manual occupations), and (iii) D (semi and unskilled manual workers)/E (state pensioners, casual and lowest grade workers and unemployed people).
- *Political affiliation*. As Attewell (2021) shows, perceptions of welfare deservingness correlate more strongly with the libertarian-authoritarian political axis—strongly reflected in the Leave/Remain divide in the United Kingdom—than with conventional distinctions between political parties. We therefore operationalised political affiliation according to the respondent's vote in the 2016 Brexit referendum. This variable has three categories: 'Remain,' 'Leave' and 'Did not vote.' A small number of respondents (1.1%) who did not remember how they voted were excluded.
- *Newspaper readership*. Right-leaning British tabloids often publish highly stigmatising coverage of welfare claimants (Baumberg et al., 2012). We therefore derived a dichotomous variable capturing whether respondents reported reading a right-leaning tabloid in the last 12 months (excluding omnivorous news consumers who also reported reading a left-leaning broadsheet in the last 12 months—please see Web Appendix A2 for the list of newspapers covered by these terms). Respondents who did not report reading any newspaper were categorised as not reading a right-leaning tabloid.
- *Claimant status*. We contrasted i) respondents who reported never having claimed, or having previously claimed only 'for a few months or less,' with ii) respondents who reported claiming currently or having previously claimed for 'longer than a few months.'³

For all analyses reported below, we restrict our sample to respondents with valid responses to all of the above measures ($N = 1402$).

To investigate H2, we incorporated an open-text question which asked respondents why they thought the claimant should or should not receive the sanction. We analysed this data using a Structural Topic Model (STM) (Roberts et al., 2014). STMs inductively classify open-text responses into topics through a machine-learning procedure which tries to identify similarities between texts based on the words they use. Once the topics have been defined, use of each topic can be predicted from covariates. We used the latter feature of STM to estimate whether claimant social class predicted the types of responses people used to justify their decisions. The STM, then, allows us to uncover not just whether there is bias in sanctioning decisions, but also whether there is evidence of bias in respondents' reasoning about those decisions. More details on the STM can be found in Web Appendix B.

3 | RESULTS

3.1 | Descriptive statistics

Overall, a small majority of respondents (62.8%, 95% CI: 59.8%–65.6%) agreed that the claimant described in the vignette should be sanctioned.

Table 2 shows the proportion of respondents endorsing a sanction, by respondent socio-demographic characteristics.

TABLE 2 Percentage (and 95% CI) of respondents who agreed/strongly agreed that the claimant should be sanctioned, by respondent characteristics ($N = 1402$)

Age	
Under 25	49.4 (37.5–61.3)
25–39	61.1 (54.8–67.1)
40–64	63.6 (59.3–67.6)
65+	68.5 (62.7–73.7)
Gender	
Male	61.6 (57.0–66.0)
Female	63.8 (60.0–67.5)
Social grade	
A/B	62.9 (57.8–67.7)
C1/C2	65.7 (61.6–69.6)
D/E	55.4 (48.1–62.5)
EU referendum vote	
Remain	53.0 (48.7–57.3)
Leave	72.5 (68.2–76.5)
Did not vote	62.1 (53.1–70.3)
Newspaper readership	
Reads right-wing tabloid	73.0 (66.9–78.3)
Does not read right-wing tabloid	60.4 (57.0–63.7)
Benefit status	
Never claimed out of work benefits, or only claimed for short period	65.6 (62.5–68.6)
Currently claiming out of work benefits, or previously claimed for long period	50.2 (42.1–58.3)

Older people were more likely to agree that the claimant should be sanctioned. Those who voted to Leave the European Union were also more likely than those who voted Remain to endorse a sanction, as were readers of right-leaning tabloids. Respondents who were currently claiming out of work benefits, or who had previously claimed for a long period, were substantially less likely to support a sanction. Respondents in social grades D and E (which includes unemployed people) were also less likely to support a sanction than those in higher grades.

3.2 | Evidence of class-based discrimination in sanctioning judgements (H1)

Table 3 shows the proportion of respondents who agreed that the claimant should be sanctioned at each level of the social class manipulation.⁴

Consistent with our hypothesis, respondents were 13–14 percentage points more likely to recommend a sanction against a claimant with underclass signifiers than against an (otherwise identical) claimant from a 'reputable' working-class (Wald $F = 15.81$, $p < 0.001$) or professional (Wald $F = 13.62$, $p < 0.001$) background.

3.2.1 | The relative magnitude of class-based discrimination

Table 4 gives the proportion of respondents who agreed that the claimant should be sanctioned at each level of the age, gender and ethnicity manipulations.

TABLE 3 Percentage (and 95% CI) of respondents who agreed/strongly agreed that the claimant should be sanctioned, by claimant social class ($N = 1402$)

Stigmatised working class	71.6 (66.6–76.2)
Reputable working class	57.5 (52.4–62.5)
Middle class	58.4 (53.2–63.4)

TABLE 4 Percentage (and 95% CI) of respondents who agreed/strongly agreed that the claimant should be sanctioned, by claimant age, gender and ethnicity ($N = 1402$)

Age	
19	69.5 (64.7–73.8)
35	62.7 (57.5–67.7)
55	56.3 (51.0–61.4)
Gender	
Male	62.9 (58.8–66.7)
Female	62.7 (58.4–66.9)
Ethnicity	
Black	63.4 (59.3–67.3)
White	62.2 (57.9–66.2)

We found no apparent effect of gender or ethnicity, but a substantial effect of age. Respondents were around seven percentage points less likely to agree to a sanction against a 35-year-old claimant than against an otherwise identical 19-year-old claimant; and were around six percentage points less likely again to endorse a sanction against a 55-year-old claimant. The difference in responses to a 19 versus a 55-year-old claimant is therefore equivalent in magnitude to the difference observed between ‘underclass’ claimants and claimants from non-stigmatised class backgrounds (Wald $F = 13.84$, $p < 0.001$).

3.2.2 | Robustness checks

We examined the effect of our randomisation procedure by comparing respondent and claimant characteristics across the three levels of the social class manipulation, finding no statistically significant differences. However, respondents who read the ‘underclass’ vignette were slightly more likely than respondents who read the ‘professional’ or ‘reputable working class’ vignettes to (i) have never claimed benefits themselves, (ii) have voted Remain and (iii) to be in social grade C1/C2 as opposed to D/E. We therefore re-ran our primary analysis as a linear probability model, including respondent benefit status, political affiliation and social grade as covariates. This had no substantive effect on our findings.

We also re-ran the analyses addressing H1 using OLS rather than logistic regressions, preserving the underlying 1–5 response scale for the dependent variable. This also had no substantive effect on our findings.

3.3 | Reasons for sanctioning judgements (H2)

Our second hypothesis was that, when reasoning about sanctioning judgements, people attribute more control/blame to claimants with ‘underclass’ signifiers. To test this, we used a Structural Topic Model (STM) to unpack

the kinds of justifications our respondents gave for their decisions to sanction or not to sanction. Across all who answered this free-text question ($N = 1170$), we found a 10-topic model to be the most parsimonious (see Web Appendix B). Figure 1 visualises the association between each of these topics and whether respondents endorsed sanctioning the claimant. It also provides the three words most strongly associated with each topic.⁵

Figure 1 shows a clear divide between those who would and who would not sanction the claimant in terms of the language they used to describe their decision. This can be seen most clearly by contrasting topics 4, 6, 7 and 10. Topics 7 and 10 were associated with a decision *not* to sanction. Responses falling into these categories tended to give claimants the ‘benefit of the doubt’—accepting the claimant’s given reason for missing the appointment as potentially genuine or suggesting an alternative mitigating justification, such as mental health. Example responses falling under these topics include ‘Her reasons may be genuine. After 2 missed appointments it would be unfair to stop her benefits completely’ and ‘She sounds depressed and needs medical help not [to be] put in a worse financial situation’ (many responses falling under Topic 7 suggested that the claimant may be suffering from depression) (see Web Appendix B4 for additional sample responses).

By contrast, Topics 4 and 6 were more prevalent among respondents who *would* sanction. Responses falling into these categories often described the claimant as making insufficient effort to attend the appointments. Example responses include ‘if she can’t be bothered to make the effort then her benefits should be taken away’ and ‘If an individual has the capability to get a job but chooses not to work, under normal circumstances, they should not be paid by the state....’

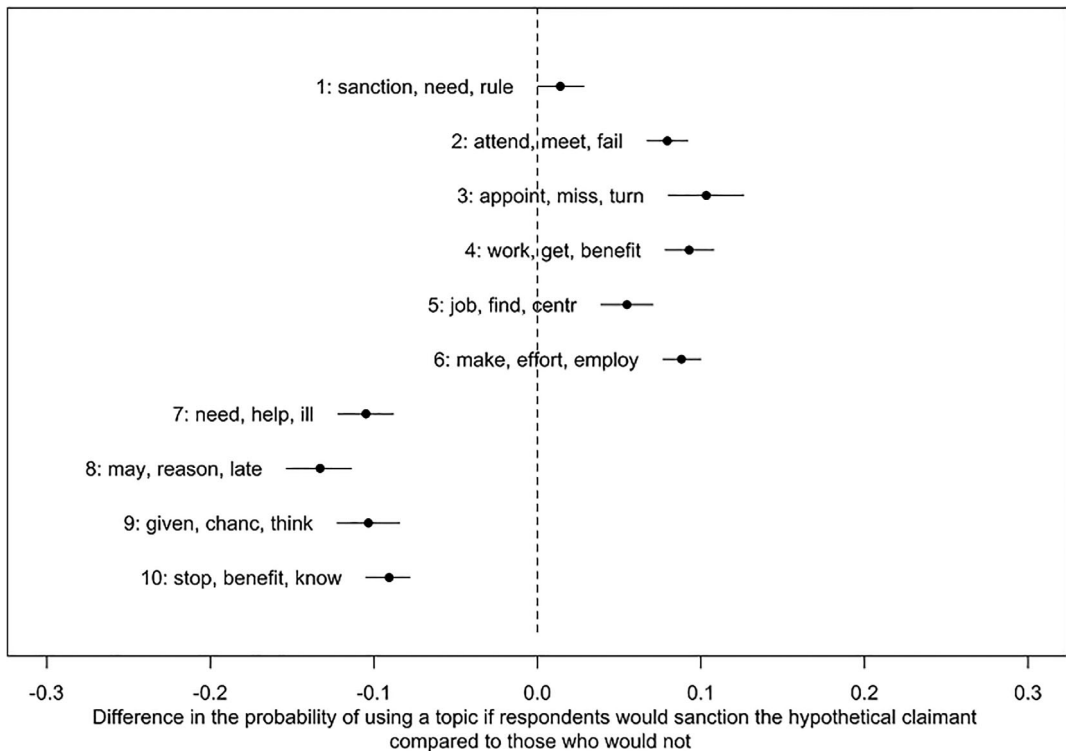


FIGURE 1 Association between STM topics and sanctioning decisions. The point estimates represent the difference in the probability of using each topic given that the survey respondent would sanction the hypothetical claimant. In other words, if respondents would sanction then the probability of deploying topic 4 is around 0.1 higher than if respondents would not sanction

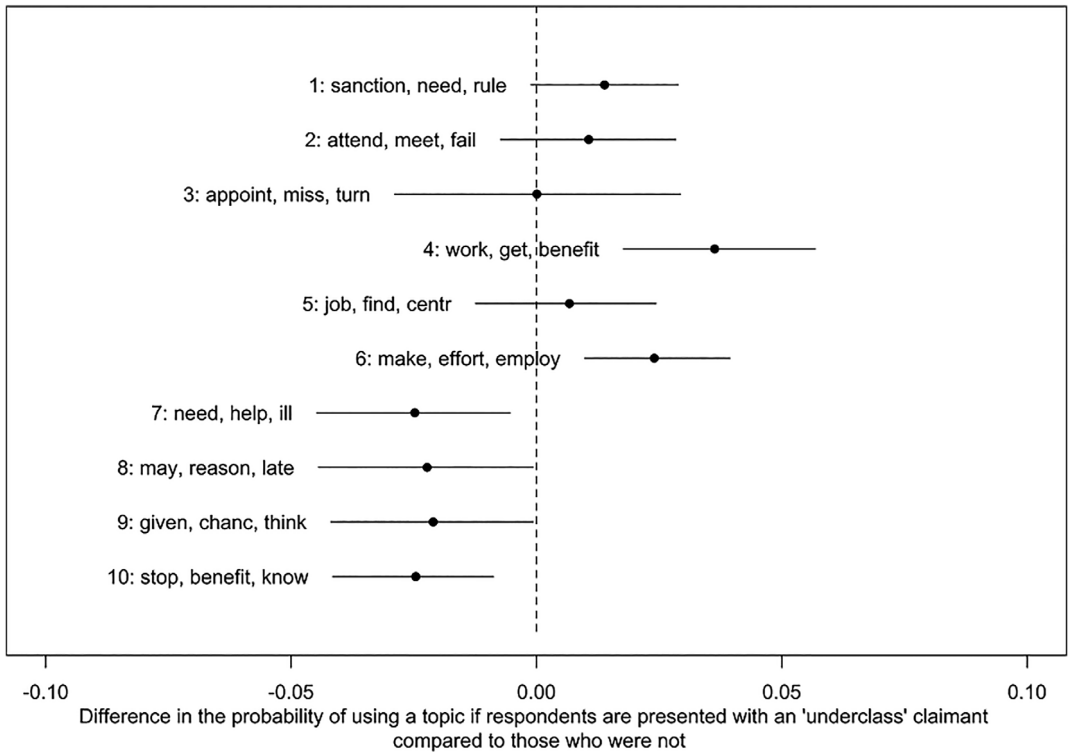


FIGURE 2 Association between STM topics and claimant social class ('underclass' vs. 'reputable working class' and professional middle class)

Next, we examined whether certain topics were more commonly used if respondents read about claimants with 'underclass' signifiers (see Figure 2). These results show that, after reading about a claimant with 'underclass' signifiers, respondents were *less* likely to use the language associated with Topic 7 or Topic 10 (the topics associated with a decision *not* to sanction, and with giving claimants 'the benefit of the doubt'). Respondents exposed to an 'underclass' claimant were also *more* likely to use language associated with Topics 4 and 6 (the topics associated with a decision to sanction, and with a suspicion that claimants were making excuses and failing to make an effort).

4 | DISCUSSION

In this paper we have examined whether public sanctioning judgements are affected by claimant social class background. We hypothesised that, due to pernicious class-based stereotypes (Hayward & Yar, 2006; Jones, 2011), claimants from stigmatised 'underclass' backgrounds are judged significantly more harshly than claimants inhabiting less denigrated class positions for the same behaviour.

Our results support this hypothesis. We provide the first direct experimental evidence of discrimination against claimants from stigmatised class backgrounds in public sanctioning judgements. Our respondents read a vignette describing a realistic scenario in which, after appearing to lose motivation to find work, a person claiming unemployment benefits had missed two compulsory caseworker appointments. When details about the claimant's housing situation and parental employment status appeared to situate them within the stigmatised 'underclass,' 72% of respondents agreed that they should be sanctioned. When an otherwise identical vignette described a claimant from a 'reputable' working class or professional background, this figure was 58%.

To contextualise the magnitude of the social class effect, we also experimentally manipulated the age, gender and ethnicity of the hypothetical claimant—demographic characteristics correlated with deservingness (Ford, 2016; Larsen, 2008) and sanctioning (Buss, 2019; Naumann et al., 2020; Schram et al., 2009) judgements. Consistent with previous work, we found that respondents were more likely to endorse sanctioning a younger claimant than an otherwise identical older one (Buss, 2019; Houtman, 1997; Naumann et al., 2020). Comparing the scale of the age effect—which is one of the most robust effects observed in the deservingness literature—to that of social class, it is notable that the two are approximately equivalent. We found no comparable effect of claimant gender or ethnicity, a finding that is in line with the mixed results of previous research on these factors (e.g., Schram et al., 2009; Terum et al., 2018).

Our methodology went beyond existing quantitative deservingness research by using qualitative data to explore the *mechanism* by which ‘underclass’ signifiers yield discriminatory judgements. Examining respondents’ qualitative explanations for their decision to sanction (or not), we found a strong emphasis on *effort/control* as a basis for judgement. Respondents often justified their sanctioning decisions on the basis of whether missing the interview was seen as *outside the claimant’s control* (due to, e.g., mental health difficulties), or *within their control* (usually due to a perceived lack of effort). Respondents judging a claimant with ‘underclass’ signifiers were substantially more likely to fall into the latter camp.

These opposing responses can be seen as a differential application of the ‘benefit of the doubt’ (a phrase many respondents used explicitly). Our vignettes described a scenario in which there were plausible grounds to suspect that the claimant’s infraction was due to a lack of motivation. In the context of a claimant from a ‘reputable’ working class or professional background, respondents were more likely to either (i) ignore these grounds—focusing on the possibility that the claimant’s stated reason for missing the appointment was genuine; or (ii) to attribute any lack of motivation to a potential mental health problem—thereby also placing the infraction outside the claimant’s control. By contrast, in the context of a claimant with ‘underclass’ signifiers, such ‘benefit of the doubt’ tended to be withheld. ‘Underclass’ claimants were instead more often suspected of missing their appointment due to intentional laziness (they ‘couldn’t be bothered’ to attend). The results from our qualitative responses therefore strongly suggest *why* respondents were more likely to support sanctioning a claimant with ‘underclass’ signifiers: they were attributing different deservingness relevant characteristics—specifically *blame/control*—to these claimants based on social class cues.

One possible alternative explanation for our results is that they arise not from class-based bias, but from some other process. Perhaps living on a council estate and having parents who were long-term unemployed may signify (in some sense ‘objectively’) low reciprocity at either the individual or family level. However, the fact that respondents almost never raised notions of reciprocity in their justifications makes this unlikely.

A notable facet of our results is that we found no distinction in responses between vignettes describing a claimant from a professional background and those describing a claimant from a ‘reputable’ working-class background. We specifically focused our hypotheses on discrimination against claimants with ‘underclass’ signifiers. As we have noted, it is this group who are subject to the most extreme stigmatisation (Jones, 2011). However, one might have expected class-based stereotypes to produce some differences between claimants from the two other class backgrounds. After all, even members of the ‘respectable’ British working class are not immune to negative stereotypes (Hesmondhalgh, 2017), particularly in the wake of Brexit (Ford & Goodwin, 2017). There are a number of possible explanations for this result; including conflicting positive and negative stereotypes (Fiske et al., 1999), perceptions of need (Kootstra, 2016) and a suppression of negative stereotypes of the ‘hard working’ working class in the context of unemployment benefits. Further research is necessary to untangle these potential mechanisms.

4.1 | Limitations and directions for future research

One potential limitation of the present study arises from the nature of our vignette. In order to (i) realistically mirror caseworker experience and (ii) generate a ‘vivid’ mental image in our respondents (Petersen & Aarøe, 2013), our vignette included a number of details which were extraneous to the central experimental manipulation (for example, details about the claimant’s recent employment history). The vignette also included a plausible reason to suspect the

claimant of a genuine infraction (missing an appointment *without a good reason*) by explicitly referencing concerns about the claimant's motivation.

Both of these factors may affect the generalisability of our results. Most importantly, the social class bias we observe may only arise in the presence of an explicit cue to question the claimant's motivation (or another deservingness related factor). This would be consistent with Schram et al.'s (2009) finding that racial bias in sanctioning judgements arose only in the presence of a stereotype-consistent discrediting marker (such as being a single mother with many children). Schram et al. (2009) argue, based on the Racial Classification Model, that when a racially minoritized policy target is seen to possess such a marker, this increases the salience of their link with group-level stereotypes, thereby increasing the likelihood of discrimination. In our case, concerns about motivation may have increased the salience of the link between our 'underclass' claimant and associated negative stereotypes about this group. Whether social class bias would be present, or as strong, in the absence of such a stereotype-linked marker therefore remains an open question. Nevertheless, our free-text responses give us a novel window into the nature of the double-standard applied to 'underclass' claimants with respect to this cue. Of particular note is the extent to which respondents who read about non-'underclass' claimants ignored the motivation cue, or interpreted it as evidence of an imputed exculpatory factor (mental health) which was not otherwise cued in the vignette.

Despite this limitation, our results represent strong evidence of a causal effect of class signifiers on sanctioning judgements. It remains for future research to determine the extent to which such class-based bias is affected by other contextual factors—including the presence or absence of stereotype-consistent discrediting markers. Such research could also be used to investigate how disambiguating information with respect to specific deservingness characteristics (for example, reciprocity or need) affects the extent of class-based bias. This would extend the work of, for example, Aarøe and Petersen (2014), who examine the effect on deservingness judgements of 'filling in the blanks' that are otherwise filled by group stereotypes. Our findings also suggest that it would be fruitful to investigate how class-based (and other) stereotypes might affect how the same behavioural cues are *interpreted*. Are behaviours which are taken to indicate a lack of motivation in negatively stereotyped claimants seen very differently when displayed by non-stigmatised claimants?

A further limitation of our study is our focus on a single country—the UK. While deservingness criteria are relatively consistent across countries (Van Oorschot, 2006), the nature of class-based stereotypes is likely to differ across national contexts (Aarøe & Petersen, 2014). Elements of the British media have sharply vilified benefit claimants, and have consistently linked benefit receipt to stereotypical portrayals of a stigmatised 'underclass' (Baumberg et al., 2012; Jones, 2011; Okoroji et al., 2021). This distinction between a 'reputable' working class and a disreputable 'underclass' may be less salient or even entirely absent in other countries.

5 | CONCLUSIONS

This study provides robust evidence of discrimination in public sanctioning judgements against claimants with signifiers tying them to a stigmatised social class group. These findings move the literature forward in three main ways. First, social class has been largely overlooked in previous welfare deservingness research. Our results suggest a strong, causal effect of social class background on public deservingness judgements.

Second, the existing empirical literature on deservingness attitudes focuses primarily on welfare access and generosity. Such attitudes do not strongly incorporate the punitive, coercive logic of contemporary activation-based welfare policy. Our research builds on the smaller, but growing empirical literature focusing specifically on punitive sanctioning judgements. By demonstrating a discriminatory effect of claimant characteristics on sanctioning judgements, our results raise substantial policy questions. For example, how strongly is public support for a coercive approach to welfare linked to perceptions of the class background of claimants? Would a decrease in the salience of the link between benefit receipt and the 'underclass' decrease support for punitive welfare policies (de Vries et al., 2021)?

Finally, our results also raise the possibility that claimants from stigmatised class backgrounds will be treated more harshly by the benefits system (e.g., *ceteris paribus* being more likely to experience a sanction). Our research focused on the general public rather than caseworkers, so we cannot draw this conclusion definitively. However, given research in other frontline contexts, it seems highly likely that public biases would be reproduced among welfare caseworkers. Given that claimants from denigrated class backgrounds are the least likely to have alternative forms of financial support, such discrimination could lead to high levels of avoidable harm.

ACKNOWLEDGEMENTS

We would like to thank the University of Kent Faculty of Social Sciences for financially supporting this research. We would also like to thank Jemma Conner at YouGov for her assistance in the design and deployment of the survey.

ENDNOTES

- ¹ StatXplore is the UK Department for Work and Pensions' (DWP) portal for accessing welfare statistics.
- ² For a 55-year-old claimant, having both parents still working is implausible. Therefore, in these vignettes, parental occupations were described in the past tense.
- ³ This data was collected from a novel survey question rather than from YouGov's background data
- ⁴ Because claimant characteristics were varied randomly across the vignettes, these estimates are not adjusted for other claimant characteristics (Mutz, 2011). Similarly, we do not adjust for the socio-demographic characteristics of respondents (Mutz, 2011) (though, see 'robustness checks').
- ⁵ See Web Appendix B2 for topic prevalence and an expanded list of key words, and Web Appendix B3 for the coefficients visualised in Figures 1 and 2.

DATA AVAILABILITY STATEMENT

The research data and code used to produce the results reported here are available through the Open Science Foundation at: <https://osf.io/vjx7m/>

ORCID

Robert de Vries  <https://orcid.org/0000-0002-6776-836X>

Aaron Reeves  <https://orcid.org/0000-0001-9114-965X>

Ben Geiger  <https://orcid.org/0000-0003-0341-3532>

REFERENCES

- Aarøe, L., & Petersen, M. B. (2014). Crowding out culture: Scandinavians and Americans agree on social welfare in the face of deservingness cues. *The Journal of Politics*, 76(3), 684–697.
- Adler, M. (2016). A new leviathan: Benefit sanctions in the twenty-first century. *Journal of Law and Society*, 43(2), 195–227.
- Attewell, D. (2021). Deservingness perceptions, welfare state support and vote choice in Western Europe. *West European Politics*, 44(3), 611–634.
- Baumberg, B., Bell, K., & Gaffney, D. (2012). Benefits stigma in Britain. Turn2us: <http://www.jrf.org.uk/publications/attitudes-economic-inequality>
- Buss, C. (2019). Public opinion towards targeted labour market policies: A vignette study on the perceived deservingness of the unemployed. *Journal of European Social Policy*, 29(2), 228–240.
- Clasen, J., & Clegg, D. (2007). Levels and levers of conditionality: Measuring change within welfare states. In J. Clasen & N. A. Siegel (Eds.), *Investigating welfare state change: The 'dependent variable problem' in comparative analysis*. Edward Elgar Publishing.
- de Vries, R., Baumberg Geiger, B., Scullion, L., Summers, K., Edmiston, D., Ingold, J., Robertshaw, D., Young, D., (2021). Solidarity in a crisis? Trends in attitudes to benefits during COVID-19. Welfare at a social distance project report <https://tinyurl.com/47yk2faj>.
- Deeming, C. (2015). Foundations of the workfare state—reflections on the political transformation of the welfare state in Britain. *Social Policy & Administration*, 49(7), 862–886.

- Department for Work and Pensions (2020a). Universal credit and your claimant commitment. <https://www.gov.uk/government/publications/universal-credit-and-your-claimant-commitment-quick-guide/universal-credit-and-your-claimant-commitment>.
- Department for Work and Pensions (2020b). Advice for decision making: Staff guide. <https://www.gov.uk/government/publications/advice-for-decision-making-staff-guide>.
- Fiske, S. T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (dis) respecting versus (dis) liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues, 55*(3), 473–489.
- Ford, R. (2016). Who should we help? An experimental test of discrimination in the British welfare state. *Political Studies, 64*(3), 630–650.
- Ford, R., & Goodwin, M. (2017). Britain after Brexit: A nation divided. *Journal of Democracy, 28*(1), 17–30.
- Geiger, B. B. (2018). Benefit ‘myths’? The accuracy and inaccuracy of public beliefs about the benefits system. *Social Policy & Administration, 52*(5), 998–1018.
- Hanley, L. (2017). *Estates: An intimate history*. Granta Books.
- Hayward, K., & Yar, M. (2006). The ‘chav’ phenomenon: Consumption, media and the construction of a new underclass. *Crime, Media, Culture, 2*(1), 9–28.
- Hesmondhalgh, D. (2017). The media’s failure to represent the working class: Explanations from media production and beyond 1. In J. Deery & A. Press (Eds.), *Media and class*. Routledge.
- Heuer, J. O., & Zimmermann, K. (2020). Unravelling deservingness: Which criteria do people use to judge the relative deservingness of welfare target groups? A vignette-based focus group study. *Journal of European Social Policy, 30*(4), 389–403.
- Houtman, D. (1997). Welfare state, unemployment, and social justice: Judgments on the rights and obligations of the unemployed. *Social Justice Research, 10*(3), 267–288.
- Jones, O. (2011). *Chavs: The demonization of the working class*. Verso Books.
- Kallschmidt, A. M., & Eaton, A. A. (2019). Are lower social class origins stigmatized at work? A qualitative study of social class concealment and disclosure among white men employees who experienced upward mobility. *Journal of Vocational Behavior, 113*, 115–128.
- Knotz, C., Gandenberger, M., Bonoli, G., & Fossati, F. (2020). RICE—an integrated model of welfare deservingness perceptions. NCCR on the Move Working Paper #26.
- Kootstra, A. (2016). Deserving and undeserving welfare claimants in Britain and The Netherlands: Examining the role of ethnicity and migration status using a vignette experiment. *European Sociological Review, 32*(3), 325–338.
- Larsen, C. A. (2008). The political logic of labour market reforms and popular images of target groups. *Journal of European Social Policy, 18*(1), 50–63.
- MacLeavy, J. (2011). A “new politics” of austerity, workfare and gender? The UK government’s welfare reform proposals. *Cambridge Journal of Regions, Economy and Society, 4*(3), 355–367.
- Macmillan, L. (2011). Measuring the intergenerational correlation of worklessness. Centre for Market and Public Organisation Working Paper 11/278.
- Monnat, S. M. (2010). The color of welfare sanctioning: Exploring the individual and contextual roles of race on TANF case closures and benefit reductions. *The Sociological Quarterly, 51*(4), 678–707.
- Mutz, D. C. (2011). *Population-based survey experiments*. Princeton University Press.
- Naumann, E., De Tavernier, W., Naegele, L., & Hess, M. (2020). Public support for sanctioning older unemployed—a survey experiment in 21 European countries. *European Societies, 22*(1), 77–100.
- Okoroji, C., Gleibs, I. H., & Jovchelovitch, S. (2021). Elite stigmatization of the unemployed: The association between framing and public attitudes. *British Journal of Psychology, 112*(1), 207–229.
- Van Oorschot, W. (2000). Who should get what, and why? On deservingness criteria and the conditionality of solidarity among the public. *Policy & Politics, 28*(1), 33–48.
- Petersen, M. B., & Aarøe, L. (2013). Politics in the mind’s eye: Imagination as a link between social and political cognition. *American Political Science Review, 107*(2), 275–293.
- Redman, J. (2020). The benefit sanction: A correctional device or a weapon of disgust? *Sociological Research Online, 25*(1), 84–100.
- Reeskens, T., & van der Meer, T. (2019). The inevitable deservingness gap: A study into the insurmountable immigrant penalty in perceived welfare deservingness. *Journal of European Social Policy, 29*(2), 166–181.
- Richards-Gray, L. (2020). Political discourse and gendered welfare reform: A case study of the UK coalition government. *Journal of Elections, Public Opinion and Parties, 1–19*. <https://doi.org/10.1080/17457289.2020.1760283>
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science, 58*(4), 1064–1082.
- Schram, S. F., Soss, J., Fording, R. C., & Houser, L. (2009). Deciding to discipline: Race, choice, and punishment at the frontlines of welfare reform. *American Sociological Review, 74*(3), 398–422.

- Terum, L. I., Torsvik, G., & Øverbye, E. (2018). Discrimination against ethnic minorities in activation programme? Evidence from a vignette experiment. *Journal of Social Policy*, 47(1), 39–56.
- Van Kersbergen, K., & Hemerijck, A. (2012). Two decades of change in Europe: The emergence of the social investment state. *Journal of Social Policy*, 41(3), 475–492.
- Van Oorschot, W. (2006). Making the difference in social Europe: Deservingness perceptions among citizens of European welfare states. *Journal of European Social Policy*, 16(1), 23–42.
- Watts, B., & Fitzpatrick, S. (2018). *Welfare conditionality*. Routledge.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: de Vries, R., Reeves, A., & Geiger, B. (2022). Social class bias in welfare sanctioning judgements: Experimental evidence from a nationally representative sample. *Social Policy & Administration*, 1–16. <https://doi.org/10.1111/spol.12812>