Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 25-1354

Understanding the decision (not) to become a teacher: evidence from survey experiments with undergraduates in the UK and US

Sam Sims
University College London

Clare Routledge
University College London

Teacher shortages are widespread, yet the reasons people choose (not) to enter the profession remain poorly understood. We conducted two survey experiments in which thousands of undergraduates chose between pairs of hypothetical jobs. This allowed us to evaluate the effects of differences in pay, working patterns and other job attributes on job choices, as well as explore how personality type and values underpin job preferences. Contrary to existing research, which is largely based on self-reports, we found that extrinsic rewards have the most influence on job choices, even among those who are considering teaching. Policymakers looking to address shortages should improve the extrinsic rewards of teaching and communicate these, alongside the many altruistic and meaningful aspects of teaching, to potential new recruits.

VERSION: December 2025

Suggested citation: Sims, Sam, and Clare Routledge. (2025). Understanding the decision (not) to become a teacher: evidence from survey experiments with undergraduates in the UK and US. (EdWorkingPaper: 25-1354). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/2qgv-x483

Understanding the decision (not) to become a teacher: evidence from survey experiments with undergraduates in the UK and US

Sam Sims[1]

Clare Routledge[1]

Teacher shortages are widespread, yet the reasons people choose (not) to enter the profession remain poorly understood. We conducted two survey experiments in which thousands of undergraduates chose between pairs of hypothetical jobs. This allowed us to evaluate the effects of differences in pay, working patterns and other job attributes on job choices, as well as explore how personality type and values underpin job preferences. Contrary to existing research, which is largely based on self-reports, we found that extrinsic rewards have the most influence on job choices, even among those who are considering teaching. Policymakers looking to address shortages should improve the extrinsic rewards of teaching and communicate these, alongside the many altruistic and meaningful aspects of teaching, to potential new recruits.

Keywords: teachers, occupational choice, recruitment, survey experiment, conjoint experiment

[1] Centre for Education Policy and Equalising Opportunities, UCL

This research was generously supported by the Gatsby Foundation.

Across rich nations, one in five teachers works in a school where the principal reports that a shortage of appropriately qualified teachers is hindering learning (OECD, 2019). This figure increases to almost one in four in the USA and one in three in England. Such shortages tend to be disproportionately concentrated in schools serving disadvantaged students (Bruno, 2025) and in particular subjects, with science and math teachers being in particularly short supply (Dee & Goldhaber, 2017; Worth & Faulkner-Ellis, 2021). In the face of such shortages, principals often resort to increasing class sizes or using temporary/supply teachers, both of which are known to harm student achievement (Benhenda, 2022; Schanzenbach, 2006).

Researchers have now provided considerable evidence on why people leave the profession and what school leaders and policymakers can do to prevent this. For example, quasi-experiments have shown the importance of pay (Biasi, 2025), panel data analysis has shown the importance of working environments (Boyd et al., 2011; Ladd, 2011; Kraft et al., 2016), and survey experiments have shown the importance of skilled and supportive colleagues (Johnston, 2021; Lentini et al., 2024; Lovison & Mo, 2024). By contrast, we know relatively little about why people decide (not to) enter the profession. In a recent review of the literature in this area, Bruno (2025) concludes that we "do not know much about the earliest stages of the teacher pipeline." This is despite the fact that interest in becoming a teacher among students in the USA is at the lowest rate since records began (Kraft & Lyon, 2024).

This lack of evidence on entry to teaching reflects several methodological challenges. In general, there is a lack of longitudinal data following undergraduate students into the labour market. The cohort studies that do exist (e.g., Bartanen et al., 2025) tend to record only limited information about the preferences of participants and often lack the coverage to be appropriate for evaluating the effects of targeted recruitment policies, such as training bursaries. The evidence we do have therefore tends to come overwhelmingly from simple self-reports of job preferences or motivations for teaching, usually collected via cross-sectional surveys with undergraduate students (Fray & Gore, 2018; See et al., 2022). However, it is not clear to what extent the findings from these surveys, e.g. the apparent importance of altruistic motives, reflect social desirability bias.

We addressed this gap in the literature by conducting two online survey experiments with thousands of photo-ID verified undergraduates in the US and UK. Job choice survey experiments have previously been employed extensively to study preferences for different types of work (for a review, see Sims et al., 2025) and several such studies have been conducted with in-service teachers to understand how they might be better retained or attracted to specific schools (Burge et al., 2021; Johnston, 2021; Lentini et al., 2024; Levatino et al., 2024; Lovison & Mo, 2024; Allen et al., 2025). However, we are not aware of any existing research that has used survey experiments to understand what might attract undergraduates *into* the teaching profession. This design allows us to gather experimental evidence on the effects of various reforms to the teaching profession and measure preferences in a way that mitigates social desirability bias (Horiuchi et al, 2022).

The research makes a number of novel contributions to the literature. First, we show that undergraduates place a high weight on extrinsic rewards (such as pay, hours and time off) and, contrary to the existing literature, this also holds true among those who report that they are considering entering teaching. Second, undergraduates are also altruistically motivated, in that they are willing to sacrifice £1,200 ($1,500) dollars in salary to choose a job with ‘significant’ (as opposed to ‘small’) social impact. Third, people with high levels of compassion, civic duty, and openness to new experience exhibit stronger preferences for the job attributes found in teaching, such as social impact and interaction with youth. The findings thus shed new light on who is likely to enter teaching and how policymakers can persuade more of them to do so.

Theory and Existing Evidence

General Models of Occupational Choice

In explaining occupational choice, economists have tended to emphasize the importance of extrinsic rewards. For example, the Roy model predicts that people will choose to enter the occupation in which they can make the largest return, given their specific skills (Roy, 1951). Returns here refer to the compensation (e.g., salary) received, minus the costs (e.g., leisure time forgone) of doing the job. Psychologists and sociologists, by contrast, have tended to emphasize the intrinsic rewards of work. On the positive side, they have studied the benefits of work as a source of meaning, or the felt significance of work for one's life (Deci et al., 2017; Deci & Ryan, 2000). They have also documented negative effects of work in terms of stress and potential burnout (Galanakis & Tsitsouri, 2022).

More recently, these two research traditions have been synthesized into a single inter-disciplinary model by Cassar & Meier (2018) - henceforth C&M. This states the overall value (or utility) of a job is a function of three things: income, meaning and costs. Income depends most directly on salary but may also depend on benefits in-kind, bonuses, and employer pension contributions. Meaning depends on whether work feels as though it is aligned with one's values or mission (autonomy) and whether it provides opportunities for developing new competencies and social connection (C&M, 2018). Costs depend most directly on the hours spent at work but may also be affected by whether employees have flexibility around when or where to do their work. This model (Equation 1, below) predicts that individuals will compare the available jobs or occupations on these dimensions and choose the one with the highest overall utility.

$$Utility = Income\ (salary) + Meaning\ (mission, autonomy, competence, connection) - Costs\ (hours, flexibility) \quad (1)$$

Empirical evidence supporting various components of this model comes from time-use surveys (Bryce, 2018), panel data studies (Benz & Frey, 2008; Bartling et al., 2013), lab experiments (Ariely et al., 2008), field experiments (Carpenter & Gong, 2020; Chandler & Kapelner, 2013; Grant, 2008, Gosnell et al., 2020) and survey experiments (Battaglio et al., 2022; Schouwer & Kesternich, 2022; Valet et al., 2021).

Models of the Choice to Enter Teaching

The leading framework for understanding what motivates people to enter the specific occupation of teaching is the Factors Influencing Teaching (FIT) Choice model (Richardson & Watt, 2006). This is built on Wigfield & Eccles' (2000) expectancy value theory, which predicts that choices are a function of expectancies for future success and how much individuals value the job. With respect to teaching, expectancies for success relate to individuals' perceptions of how good they would likely be as a teacher. Alongside this, Richardson and Watt (2006) considered multiple ways in which people might value teaching. The first is *social utility value*, which is similar to the mission aspect of the C&M framework. The second is *personal utility value* (which includes time for family, job security and job transferability) and the third is *task demand value*, which captures perception of the expertise and hours required. Personal utility value and task demand value have no direct analogue in the C&M framework but overlap partially with Costs. The fourth is *task reward value*, which captures the monetary and status benefits of a job. This overlaps partially with the income term in the C&M framework.

The FIT-Choice Model was developed to underpin the FIT-Choice questionnaire, which has since been used in surveys with undergraduates and early-career teachers in at least 12 different countries around the world (Navarro-Asencio et al., 2021). A 2018 review of research on why people choose to teach found 18 papers on the topic using FIT-Choice and concluded that "interest in teaching is affected by a range of individual and societal factors, especially intrinsic and altruistic motivation" (Fray & Gore, p.158). A 2022 review on the same topic found that "A large majority of studies of teachers' motivation utilized the FIT-Choice Likert-scale self-report questionnaire" (p.15) and concluded that "across countries in Europe, USA, Australia and New Zealand, the highest rated motivations for choosing teaching among pre-service teachers… were the intrinsic reasons… and the altruistic value of teaching" (See et al., p.17).

Individual Differences and Occupational Choice

The C&M and Fit-Choice models largely attempt to explain occupational choice based on characteristics of the job, combined with general accounts of human motivation. However, there are other complementary frameworks which seek to explain job preferences based on differences in motivators across people (individual differences). Foremost among these is the concept of Public Service Motivation (PSM), originally defined as "an individual's predisposition to respond to motives grounded primarily or uniquely in public institutions and organizations" (Perry & Wise, 1990, p.368). PSM comprises four sub-constructs: attraction to public service, civic duty, compassion, and self-sacrifice (Kim et al., 2012). PSM varies considerably across people but is reasonably stable over time within individuals (Vogel & Kroll, 2016), consistent with the view that it is a trait-like individual difference. Empirical evidence shows that PSM does predict broad occupational choice (Ritz et al., 2016).

Although the reasons for the stability of PSM remain somewhat unclear, a leading theory is that it is rooted in personality type, which is itself highly stable over time (Florczak et al., 2023). Personality refers to the "the enduring configuration of characteristics and behavior that comprises an individual's unique adjustment to life" (APA, n.d.). Personality is also a multi-dimensional construct, generally thought to comprise openness, conscientiousness, extraversion, agreeableness and neuroticism (O'Connor, 2002). Empirical evidence shows that personality type shows a relationship with choice of academic major (Vedel et al., 2015) and occupational choice (Hurtado Rua et al., 2019). Openness has been found to be related to entry to teaching in previous research (Sims, 2018; Törnroos et al., 2019).

Although C&M do not discuss specific reasons (e.g., PSM) that preferences might differ across individuals, they do allow for individual differences in their model by including a set of weights represented by the Greek letter Θ ('theta'), as shown in Equation 2:

$$Utility = Income\ (\Theta, salary) + Meaning\ (\Theta, mission, autonomy, competence, connection) \\ -Costs\ (\Theta, hours, flexibility) \tag{2}$$

This version of the model predicts that all individuals' utility is affected by income, meaning and costs, but to differing extents depending on their individual traits. We adopt the C&M model as the main theoretical framework for our study because it can flexibly incorporate individual differences in this way.

The Current Study

Drawing on the above theory, this study addresses two research questions. The first (RQ1) is: how would reforms to teaching affect the probability of people choosing teaching and non-teaching jobs? To answer this, we experimentally tested the effects of varying a wide range of different job attributes – consistent with the C&M and FIT-Choice models – on undergraduates' choice of (hypothetical) jobs. Crucially, we calibrated the values of these job attributes to reflect those of teaching and other graduate jobs, which allowed us to quantify the distinctive benefits (such as social impact) and downsides (absence of flexibility) of teaching.

Our second research question (RQ2) is: how do stable individual differences (captured by the ϴ term in the C&M model) affect preferences for teaching and non-teaching jobs? To answer this, we measured a range of individual differences about our respondents (PSM, openness, gender) and then compared their preferences for different types of jobs. This allowed us to characterize the 'types' of people most likely to enter teaching.

Methods

Design

We pre-registered details of the two studies reported in this paper prior to beginning data collection (https://osf.io/qn357; https://osf.io/a7cs9). To understand people's choice about whether to enter teaching, we used full factorial, paired profile, forced choice survey experiments (Hainmueller & Hopkins, 2015; Bansak et al., 2021b). This involved providing respondents with a pair of hypothetical jobs and asking them to indicate which of the two was most attractive (a choice task). Each job was represented as a bundle of attributes (e.g., salary, working hours) and each attribute could take on a finite set of values (e.g., salary of £30,000 or £35,000). We presented the choice task in tabular form, with jobs in the columns and attributes in the rows. Separately randomizing the value of each attribute in each job allowed us to isolate the causal effect of each job attribute on participants' choices (Hainmueller et al., 2015). Importantly, several papers have shown that the results from such job choice experiments predict real-world job choice behavior (Maestas et al., 2023; Viano et al., 2021; Wiswall & Zafar, 2018).

Survey experiments have several advantages for understanding job choices and hence the decision to enter teaching. First, they reduce social desirability bias relative to traditional surveys by asking respondents to select between two jobs, both of which contain socially desirable features (Horiuchi et al., 2022). Second, survey experiments allow for the collection of large samples, allowing extensive exploration of heterogeneity in preferences across respondents. Third, survey experiments allow preferences to be converted into a common, cash-equivalent-value metric, which is informative for policymakers and school leaders weighing up the costs and benefits of different reforms aimed at improving the attractiveness of teaching.

Sample

We chose to collect data via the Prolific survey platform because it contains a large number of undergraduates and because empirical research consistently shows that responses are of high quality (Albert et al., 2023; Douglas et al., 2023; Krefeld-Schwalb et al., 2024; Peer et al., 2017; Peer et al., 2022). Unlike other platforms, participants have to verify their email address, phone number and photo ID before they are allowed to respond to any surveys (Peer et al., 2023) which makes it much harder for bots to enter the respondent pool (Westwood, 2025).[i] Once participants join, Prolific monitors their IP addresses, internet service provider, device and browser information, and virtual private network usage – and ban users that appear to be acting suspiciously. Users also pre-register a wide range of personal characteristics (including whether an undergraduate and geographic location) meaning eligibility for our survey was fixed prior to the survey being listed on the site. Across both the UK and US surveys we implemented a number of pre-registered additional checks to further check the quality of our responses (see Appendix A for details).

We conducted two survey experiments. The first focused on undergraduates located in the UK. We targeted a minimum detectable effect size of 0.03, which is just over half the average effect size (0.05) found in survey experiments (Schuessler & Freitag, 2020). Allowing each of our attributes to take on up to four values, and with each participant responding to ten job choice tasks, achieving this MDES required 871 unique respondents.[ii] We pre-registered our intention to recruit this number via Prolific and ultimately achieved a sample of 871 eligible respondents. In addition to collecting the choice task data, we measured a range of participant individual differences (e.g., openness). We paid UK participants £3 to complete the survey (approximately £12 per hour).

To check the generalizability of our main findings, we also conducted a replication study in an analogous US sample. We did not measure individual differences in this survey, which made it shorter and less costly. We therefore pre-registered our intention to recruit 1,452 respondents or stop the survey on 31st July 2025, whichever came sooner. Ultimately, we recruited 1,242 eligible US respondents by the cut-off date. We paid US participants £1.50 to complete the survey (approximately £10 per hour). Table 1 shows descriptive statistics for the two samples.

Table 1

Participant Demographics by Sample

	UK Sample	US Sample
Age, median	23	22
Gender		
Female (%)	60.2%	59.2%
Male (%)	39.5%	37.8%
Non-binary/Other (%)	< 1%	3.1%
Year of Study		
1st Year (%)	—	17.3%
2nd Year (%)	17%	23.9%
3rd Year (%)	56.2%	27.4%
4th Year (%)	26.8%	27.8%
5th+ Year (%)	—	3.4%
Subject Area		
Social Sciences/Humanities/Law (%)	32.8%	26.7%
Biological/Life Sciences (%)	31.9%	26.2%
Mathematics (%)	20.1%	14%
Physics/Engineering (%)	1.7%	13%
Creative Arts/Media (%)	8.3%	10.7%
Education/Teaching (%)	4.1%	5.1%
Other Physical Sciences (%)	1%	4.3%
Sample Size (N= unique respondents)	**871**	**1242**

Attributes and attribute values

We included ten job attributes in our UK survey experiment. The first four attributes fall within the ‘cost’ category of the C&M framework in that they relate to what people have to give up to do their work. The first of these is the typical number of hours worked per week. This attribute could take on one of four values in our choice tasks, one of which (40) corresponds to five 8-hour days (which is typical in England), and another (52) corresponds to the average number reported by teachers in England (Adams et al., 2023). The second attribute in this category is the number of weeks of paid leave per year. This could take on three values, one of which corresponds to the legal minimum in England (6) and another of which (13) corresponds to the number that teachers typically receive in England. The third attribute is days worked from home per week, which could take on four values, one of which corresponds to the typical value in office-based graduate jobs (2), and another corresponds to the typical number for teachers (0). The fourth attribute is flexibility over hours worked, which could take on three values, one of which (fixed working hours) corresponds to the typical case in teaching. These attributes have been shown to influence job preferences in a range of other job choice studies (Maestas et al., 2023; Schouwer & Kesternich, 2022; Valet et al., 2021; Wiswall & Zafar, 2018; Woźniak-Jęchorek et al., 2022).

The second set of attributes fall within the 'meaning' category of the C&M framework, which has been shown to influence job preferences in previous research (Maestas et al., 2023; Ripoll et al., 2023; Woźniak-Jęchorek et al., 2022). The first of these is the extent to which the job uses knowledge from respondents' undergraduate degree. This could take on three values, one of which (daily) corresponds to the situation for a typical secondary school (students aged 11-18) teacher in England. The second attribute in this category is the frequency with which the job involves working with young people. This could take on three values, one of which (daily) corresponds to the typical situation for teachers. The third attribute is the level of social or community impact: none, small, moderate or significant. We believe that most people would consider teaching to have significant social impact.

The third and final set of attributes relate to the income category in the C&M framework. The first of these is annual starting salary (before tax). This could take on four values, two of which were chosen to correspond to typical graduate starting salaries in England (£28,500[iii]) and typical teacher starting salaries in England (£31,650[iv]). The second attribute in this category is monthly employer pension contributions (over and above salary). This could take on three values, two of which were chosen to correspond approximately with the amount offered in defined contribution pension schemes in England (10%) and the proportion of schools' remuneration bill that goes into typical defined-benefit teacher pensions schemes in England[v]. The final attribute is a one-off cash bonus paid after two years in the job. This could take on three values, which correspond to those offered in existing 'retention bonus' schemes for teachers in England (Sims & Benhenda, 2022). Similar attributes have been shown to influence job choice in previous research (Jost & Möser, 2023; Ripoll et al., 2023; Valet et al., 2021; Wiswall & Zafar, 2018; Woźniak-Jęchorek et al., 2022). All attributes and values for our UK sample are summarized in Table 2 below. The ten attributes correspond to the first ten hypotheses in our UK pre-registration. An example of a choice task can be seen in Figure 1.

Figure 1
Example of a choice task in the UK undergraduate survey experiment.

(1/10) Choose your preferred option below:

	Option 1	Option 2
One-off bonus paid after two years	£7,500	£3,000
Typical working hours per week	48 hours per week	48 hours per week
Paid leave per year	13 weeks per year	17 weeks per year
Days worked from home per week	5 days per week from home	2 days per week from home
Flexibility over working hours	Fixed working hours	Complete flexibility over working hours, within demands of the role
Uses knowledge from your undergraduate degree	Never	Daily
Frequency of working with young people	Never	Daily
Level of social or community impact you can make	Moderate impact	No impact
Employer payment added into your pension pot each month	Equivalent to 14% of salary	Equivalent to 5% of salary
Starting salary per year (before tax)	£28,500	£40,000
	○	○

.

Table 2
Attributes and possible attribute values for the UK survey experiment.

C&M	FIT	Attribute	Value 1	Value 2	Value 3	Value 4
Costs	Task demand	Typical working hours per week	40 hours	48 hours	52 hours	60 hours
	Personal utility	Paid leave per year	6 weeks	13 weeks	17 weeks	-
		Days worked from home per week	0 days per week from home	2 days per week from home	4 days per week from home	5 days per week from home
		Flexibility over working hours	Fixed working hours	Ability to shift start/end times forward or back by one hour	Complete flexibility over working hours, within demands of the role	-
Meaning	Task demand	Uses knowledge from undergraduate degree	Never	Weekly	Daily	-
	Social utility	Frequency of working with young people	Never	Weekly	Daily	-
		Level of social or community impact you can make	No impact	Small impact	Moderate impact	Significant impact
Income	Task reward	Starting salary per year (before tax)	£28,500 per year	£31,650 per year	£40,000 per year	£49,000 per year
		Monthly employer contribution to pension (additional to salary)	5% of salary	10% of salary	14% of salary	-
		One-off bonus paid after two years	£0	£3,000	£5,000	£7,500

Note: C&M refers to the terms from the Cassar & Meier (2018) framework. FIT refers to the categories from the FIT-Choice framework. Each hypothetical job presented to respondents in the choice tasks comprised the ten attributes listed above. For each hypothetical job, the value for each of these attributes was randomly selected from the values shown in the columns.

For our US sample, we selected analogous values that are aligned with the typical situation for new teachers and recent graduates in the US (see Appendix B). Based on consultation with US-based teachers, we adapted the values for some of the attributes to match them to the US context. For example, the amount of paid leave has been reduced to reflect the typical values in US schools (11 days) and a plausible value for US graduate employers (14 days). We also added one additional value for the employer pension contribution to reflect plausible values for the US (3%). Since there are no national teacher pay scales for the US, the salary and bonus values were converted from pound to dollar values using 2025 exchange rates.

Analytical approach

To answer RQ1, we need to quantify how varying the characteristics of the jobs in our choice tasks affect participants' job choices. Estimating average treatment effects (ATEs) would require a model including all possible interactions between all attributes, which would require a very large sample. Instead, we estimated average marginal component effects (AMCEs), which capture the difference in the probability of choosing a job for a one unit increase in the focal attribute, averaged over the joint distribution of all other attributes. By subsuming any interactions into the estimand, the AMCE can be estimated using a much smaller sample, using the following additive linear probability model (Equation 3):

$$Y_{ij} = \beta_0 + \sum_{l=1}^{10} \beta_l A_l + \varepsilon_{ij} \tag{3}$$

Where:

- Y_{ij} is a binary variable capturing whether job profile j was preferred by individual i
- A_l is vector of l job attributes
- β_l is the estimate of the AMCE for each of the attributes. In practice, there is one coefficient for each of the values of each attribute, excluding the reference categories.

Since one of our attributes is salary, we can also estimate salary equivalent values (SEV) for each of our attributes by dividing the coefficient on the focal attribute value (e.g., 48 hours per week) by the coefficient on the second lowest salary value (£31,500, relative to a reference category of £28,500). Since both coefficients are estimated with some degree of imprecision, we estimate the standard error on the SEVs using bootstrap methods (Hole, 2007a; Hole, 2007b).

To answer RQ2, we need to quantify and compare the strength of preferences for different job attributes across our sample. To do this, we estimated marginal means for each attribute value. Marginal means capture the proportion of times a job profile containing the focal attribute value was chosen, among all instances of job profiles containing the focal attribute value. For example, among all job profiles with a salary value of £40,000, how often was the job profile with a salary of £40,000 chosen. Since the value of all job attributes were separately randomized, this provides a descriptive measure of the intensity of preferences for that attribute value. To explain how stable, trait-like individual differences explain these preferences, we then compared the intensity of these preferences across groups.

Results: Research Question 1

UK undergraduates

Figure 2 shows the AMCEs estimated from our UK sample. All ten attributes and their possible values are arrayed on the vertical axis. The horizontal axis shows the change in probability of choosing a job when changing the value of an attribute from the reference category. For example, moving from a salary of £28,500 to a salary of £31,500 increases the probability of choosing a job by 0.09, or 9 percentage points (pp). The attributes have been sorted so that the attribute values with the largest AMCEs (e.g., salary) appear at the top.

The top cluster of coefficients in Figure 2 (dark blue) relate to pay. Increasing the typical graduate starting salary in England (£28,500) to the typical teacher starting salaries in England (£31,650) - which represents an 11% uplift - increases the probability of choosing a job by 0.09

or 9pp. This suggests that (assuming undergraduates are informed) the relatively high starting pay for teachers in England is helping attract undergraduates into teaching. A 40% increase from the typical graduate starting salary (up to £40k) increases the probability of choosing a job by 0.29 or 29pp and a 72% increase from the typical graduate starting salary (to £49k) increases it by 0.37 or 37pp.

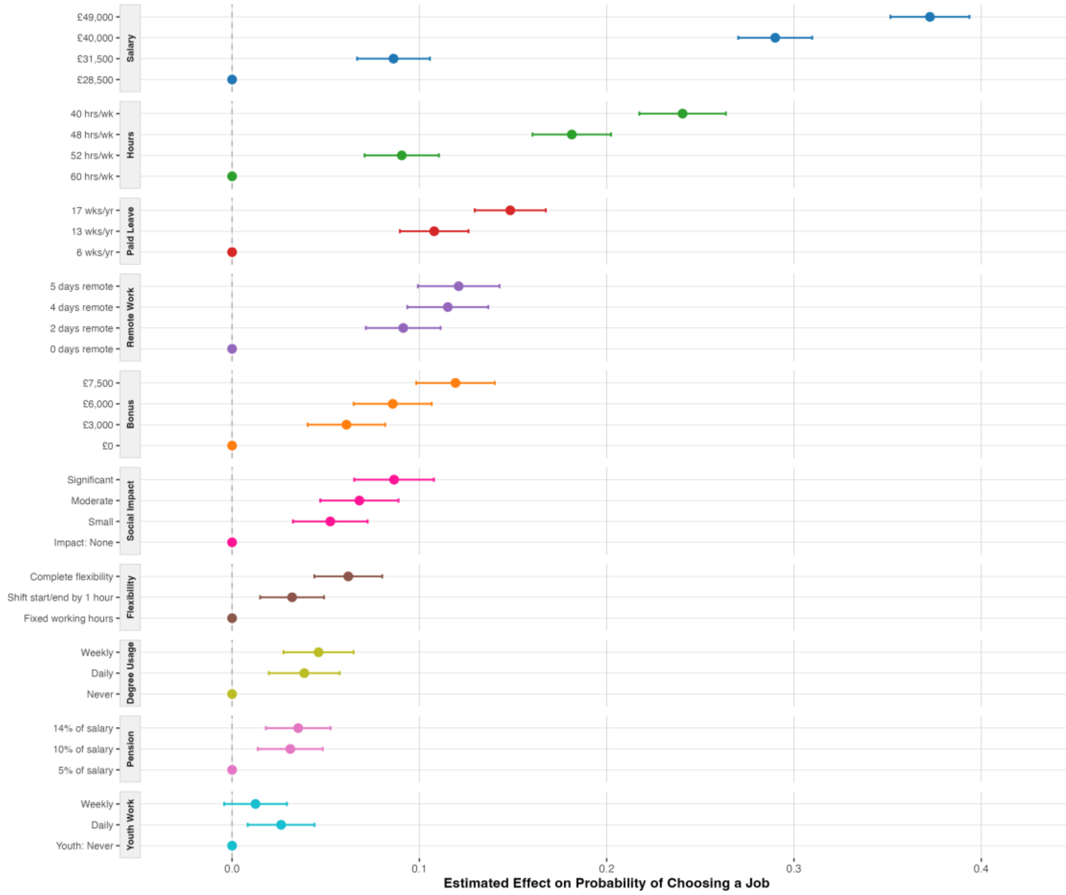
The second cluster of coefficients (green) relate to hours worked. Increasing hours from the typical full-time working week (40 hours) to the typical teacher working week in England (52 hours) reduces the probability of choosing a job by 0.15 or 15pp. The term-time hours worked by UK teachers over and above standard full-time hours have a salary equivalent value of approximately -£3.2k (full SEV results in Appendix C). The third cluster of coefficients (red) relate to paid leave. Increasing the amount of paid leave per year from the typical amount for a full-time job in England (six weeks) to the typical paid leave for a teacher in England (13 weeks) increases the probability of choosing a job by 0.11 or 11pp. This suggests that (assuming undergraduates are informed) the relatively high paid leave for teachers in England is helping attract undergraduates into teaching. Indeed, this extra paid leave has a salary equivalent value of approximately £3.7k. Some schools in England are currently experimenting with giving teachers one day off per fortnight, without any decrease in pay (Cumiskey, 2024). This raises teachers' paid leave from 13 to 17 weeks per year, which increases the probability of choosing a job by a further 0.04 (or 4pp).

The fourth cluster of coefficients (purple) relate to the number of days per week worked remotely. Decreasing the number of days worked from home from the typical amount for a graduate job (two days per week) to the value typical in teaching (zero days per week) leads to a 0.09 (or 9pp) reduction in the probability of choosing a job. This is similar to the effect of an approximately -£3.2k change in starting pay (dark blue coefficients). The fifth cluster of coefficients (orange) relate to a cash bonus paid after two years in the job. Increasing the bonus from £0 to the value seen in various existing targeted teacher retention policies in England (£3k-7.5k) increases the probability of choosing a job by 0.06-0.12 (or 6-12pp). This suggests that, if government can inform undergraduates about the existence of such retention bonuses, it would help attract them into teaching. Cash bonuses paid after two years on the job have a salary equivalent value less than the face value of the bonus. The difference between the SEV and the face value grows as the value of the bonus increases, suggesting diminishing returns.

The sixth cluster of coefficients (pink) relate to the social impact of the job. Moving from a job with ‘small’ social impact to a job with ‘significant’ social impact is associated with a 0.03 (or 3pp) increase in probability of choosing the job. Our salary equivalent value estimates suggest that teachers would be willing to forgo approximately £1.2k per year to obtain a job with significant (as opposed to small) social impact. The remaining four clusters show weaker (0.06 or less change in probability) relationships with the probability of choosing a job. These are: flexibility over when hours are worked, use of knowledge from undergraduate degree, employer pensions contribution over and above salary, and frequency of working with young people. The small effects for pensions are particularly striking. Having employer pension contributions equivalent to 14% of salary compared to 5% of salary has a salary equivalent value of £1.2k. Given that this 11 percentage point increase in employer contributions would have a cash value far in excess of £1.2k, this shows that a sizable discount rate is placed on retirement income by our sample.

Figure 2

How does varying the characteristics of jobs affect their attractiveness to UK undergraduates?



Note. Coefficients are average marginal component effects, estimated using Equation 3. Horizontal lines show 95% confidence intervals with standard errors clustered at respondent level. N=871 unique respondents, each responding to 10 paired profile choice tasks.

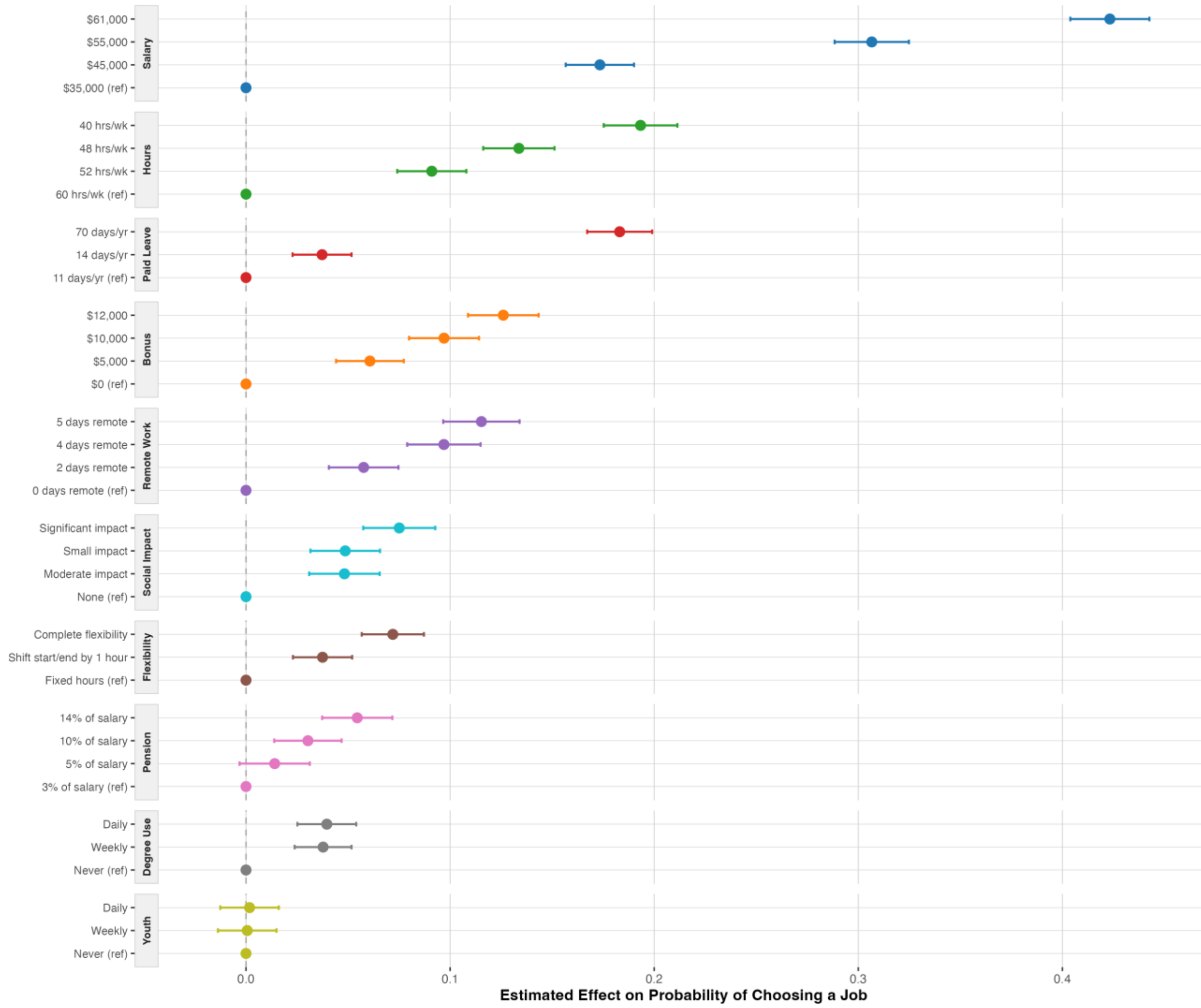
Replication and extension with US undergraduates

Figure 3 shows the AMCEs estimated from our US sample. The layout is equivalent to Figure 2, contains the same attributes, and uses analogous attribute values. The ten attributes in the figure correspond to the first ten hypotheses in our US pre-registration. The results are similar to those in Figure 2. The ranking of the attributes (on the vertical axis) is very similar. Indeed, only two pairs of attributes swap places in the ranking. The absolute values of the coefficients also tend to be similar. For example, increasing the number of days working from home from 0 per week to 5 per week leads to a 0.12 (or 12pp) increase in the probability of choosing a job in both the UK and the US. More generally, the Pearson correlation coefficient between the attribute importances in the two figures is 0.87. The only effect that shows a noticeable difference between Figure 2 and Figure 3 is moving from the lowest salary value to the second lowest salary value, which has a larger effect (+17pp) in the US than in the UK (+9pp). The other two salary coefficients are, however, comparable. The SEV for the US sample can be found in Appendix D.

The results in Figure 2 and Figure 3 are informative about how the average undergraduates' job choices are affected by changes in different job characteristics. However, some of these respondents would likely never consider a career in teaching, while others may be actively considering it. Policymakers looking to entice more undergraduates into teaching are likely to be particularly interested in the latter, more marginal group. An additional motivation for conducting our second US survey was therefore to directly inquire about whether respondents were considering teaching and check whether the results differ based on this.

Figure 3

How does varying the characteristics of jobs affect their attractiveness to US undergraduates?



Note. Coefficients are average marginal component effects. Horizontal lines show 95% confidence intervals with standard errors clustered at respondent level. N=1242 unique respondents, each responding to 10 paired profile choice tasks.

Figure 4 shows the results broken down across the two groups, with error bars representing +/- 1SD. We define an individual as 'open to teaching' (blue coefficients) if they agreed or strongly agreed with the statement 'I would consider becoming a K-12 teacher' or reported that their 'current plan is to become a K-12 teacher'. By and large, the coefficients are of similar magnitude across the two groups, with the possible exception of salary and pensions.

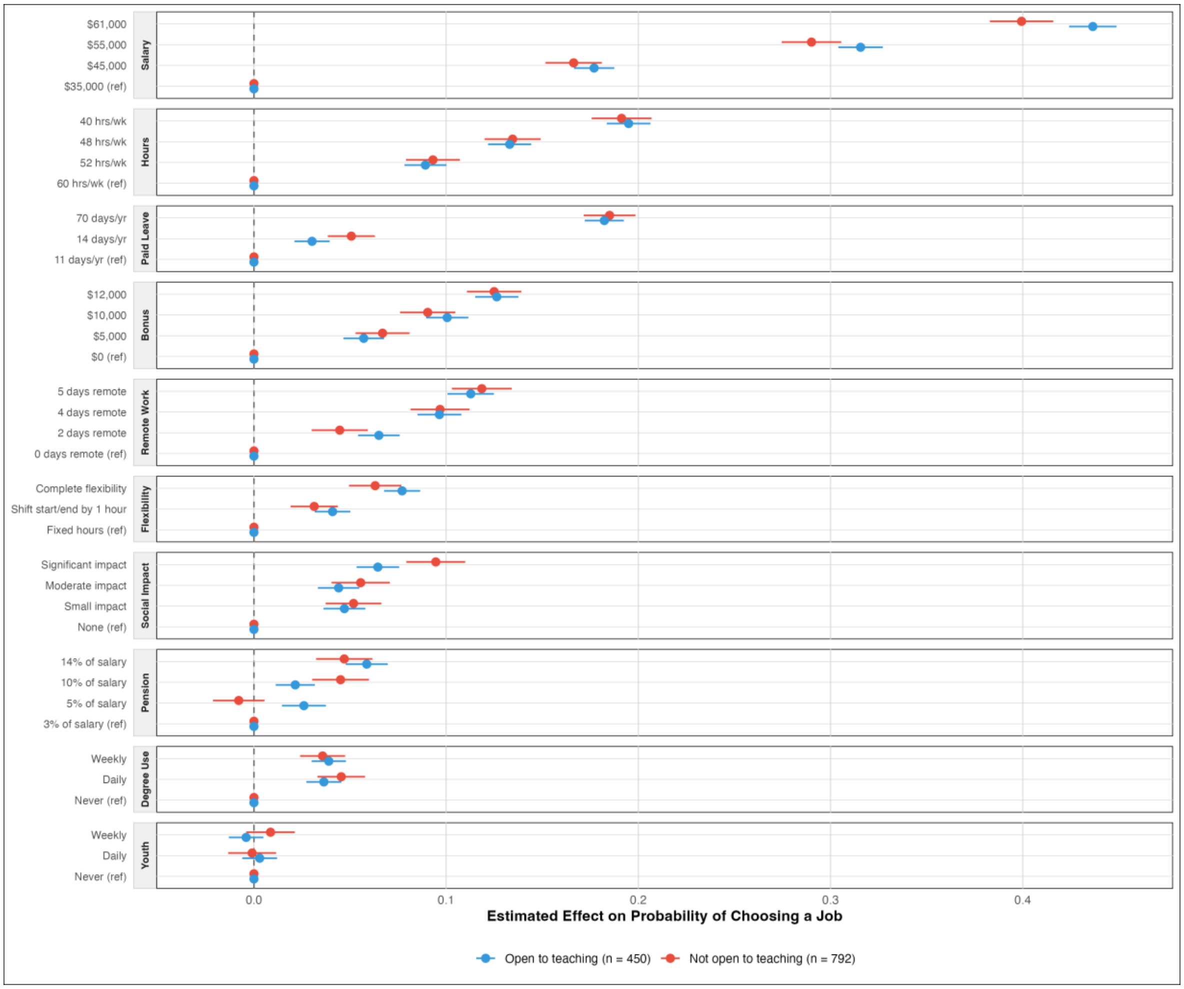
The effects of increasing salary to \$55,000 (*p*=0.18) and \$61,000 (*p*=0.07) are slightly larger for the open to teaching group than for the not open to teaching group, albeit not statistically significant. The effect of increasing pension contributions from 3% to 5% isslightly larger for the open to teaching group. However, once again, this difference is not statistically significant (*p*=0.05). Broadly speaking then, we do not find compelling evidence of heterogenous effects depending on whether individuals are open to teaching.

Robustness checks

The internal validity of our results depends on a number of assumptions (Hainmueller et al., 2013). First, decisions made by respondents in earlier choice tasks should not affect their decision in later choice tasks (stability and no carryover effects). We tested this by comparing the coefficients estimated during the first five and last five choice tasks. Second, the decision made by respondents should not be affected by whether a job appears on the left profile or the right profile (no profile order effects). We tested this by comparing the coefficients estimated using just the data from left-hand side profiles and just the data using right-hand side profiles. Third, there should be complete randomization of all attribute values. We present the results of these robustness checks for the UK (Appendix E) and US (Appendix F) samples. None of these robustness checks provide reasons to doubt the validity of our analysis.

Figure 4

How does varying the characteristics of jobs affect their attractiveness to US undergraduates by whether they are 'open to teaching'?



Note: US university undergraduates (n = 1,242). Values show estimated effects of changing job attributes on probability of choosing a job, relative to reference levels (shown in parentheses). Error bars show ±1 standard deviation with standard errors clustered by respondent. 'Open to teaching' includes respondents who would consider becoming a K-12 teacher (agree/strongly agree) OR plan to teach (n = 450). 'Not open to teaching' includes those who would not consider teaching (disagree/strongly disagree) AND do not plan to teach (n = 792).

Results: Research Question 2

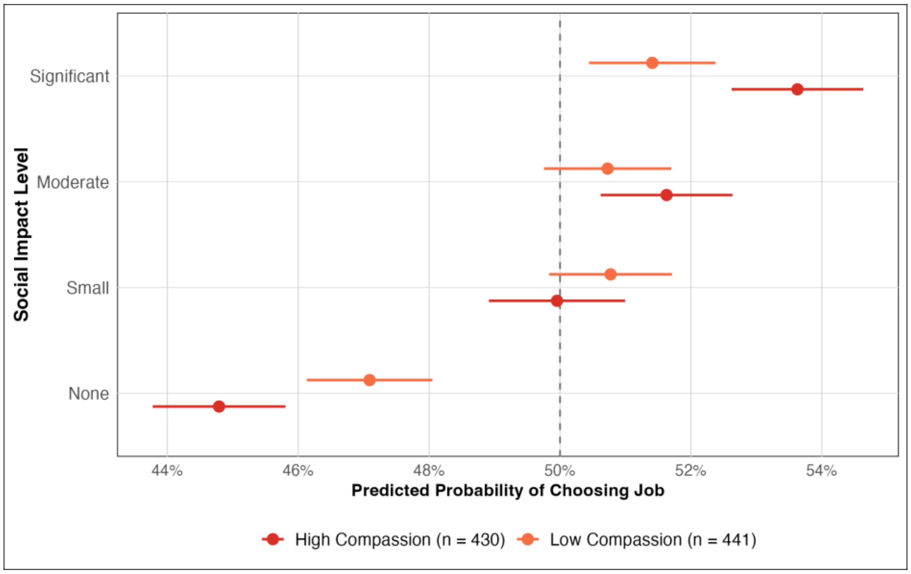
We pre-registered five hypotheses (H11-H15) about how preferences might differ across respondents in our UK data. First, we hypothesized (H11) that respondents with higher levels of 'public service motivation - compassion' will prefer jobs with greater societal and community impact more strongly than respondents with lower levels of 'public service motivation –

compassion'. Compassion is defined as love of others, especially fellow citizens, which suggests that those who are high in compassion will show a preference for jobs that benefit others (Perry et al., 1996). Figure 5 shows the marginal means for different levels of social impact, split by whether respondents are above or below the median for compassion. If respondents are indifferent about something, then they will choose job profiles that contain it approximately 50% of the time; whereas if they have preference for it, they will choose it more often. Supporting H11, there was a significant interaction between social impact and compassion ($p = 0.032$), with high compassion individuals showing stronger preferences for jobs with significant societal impact. For completeness, we report this analysis for all attributes in Appendix G.

Second, we hypothesized (H12) that respondents with higher levels of 'public service motivation – civic duty' will prefer jobs with greater societal and community impact more than respondents with lower levels of 'public service motivation – civic duty'. Civic duty is defined as a public service ethic, which suggests that those who are high in civic duty will show a preference for jobs that benefit others (Perry et al., 1996). Figure 6 shows the marginal means for different levels of social impact split by subgroups based on being above and below median values on the civic duty measure. In line with our hypothesis, those who are high in civic duty clearly exhibit a stronger preference for jobs with 'significant' social impact ($p = 0.002$).For completeness, we report this analysis for all attributes in Appendix H.

Figure 5

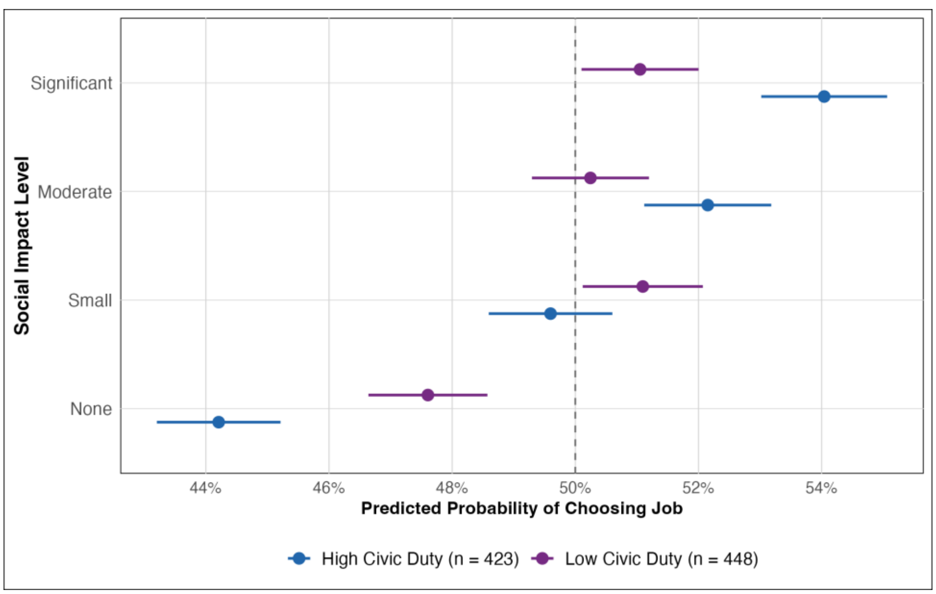
Differences in intensity of preference for social impact by compassion level



Note. UK survey experiment (n = 871). Coefficients are marginal means. Vertical line at 50% indicates indifference. Error bars show ±1 standard error. Standard errors clustered by respondent. 'High Compassion' includes respondents above the median PSM-compassion factor score (n = 430).

Figure 6

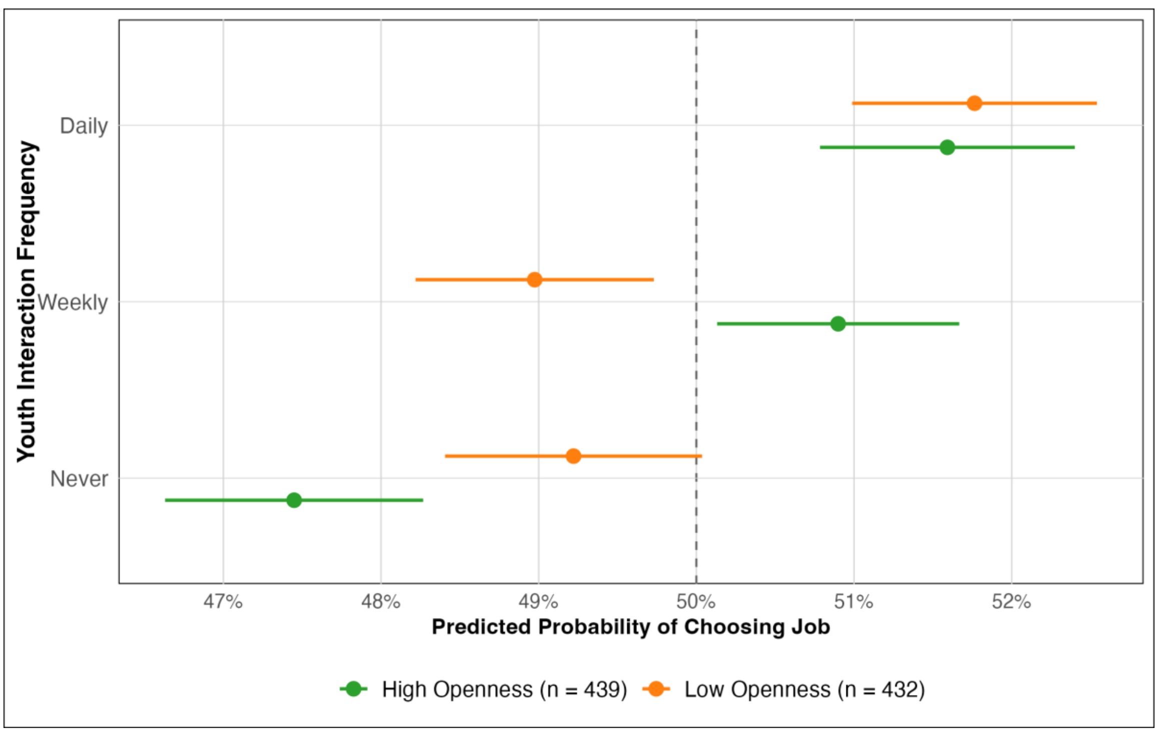
Differences in intensity of preference for social impact by civic duty level



Note. UK survey experiments (n = 871). Coefficients are marginal means. Vertical line at 50% indicates indifference. Error bars show ±1 standard error. Standard errors clustered by respondent. 'High Civic Duty' includes respondents above the median PSM-civic duty factor score (n = 423).

Third, we hypothesized (H13) that respondents with higher levels of the personality trait 'openness to experience' will prefer jobs with greater youth interaction. Openness to experience is characterized in part by curiosity and a preference for variety and learning (Gosling et al., 2003). We reasoned that working with youth would provide varied experiences and opportunities for learning and personal growth. Figure 7 shows the marginal means for different levels of social youth interaction split by subgroups based on being above and below median values on the openness measure. Partly in line with our hypothesis, those that are high in openness exhibit a stronger preference for jobs with 'weekly' youth interaction (p=0.046). However, there is no clear difference in preferences for 'daily interaction'. For completeness, we report this analysis for all attributes in Appendix I.

Figure 7
Differences in intensity of preference for youth interaction by openness level



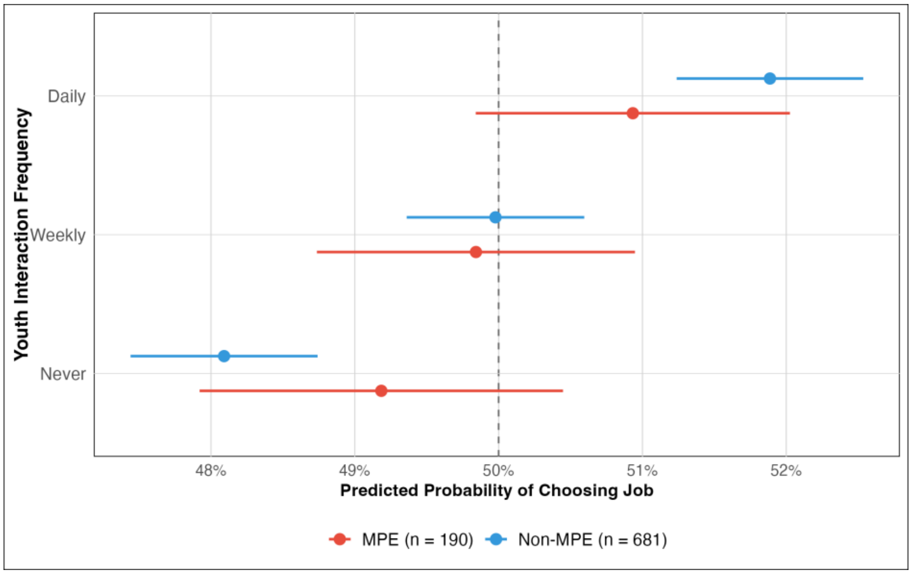
Note. UK survey experiment (n = 871). Coefficients are marginal means. Vertical line at 50% indicates indifference. Error bars show ±1 standard error with standard errors clustered by respondent. 'High Openness' includes respondents above the median openness to experience score from 10-item IPIP Big Five scale (n = 439). 'Low Openness' includes those below the median (n = 432).

Fourth, we hypothesized (H14) that respondents with an undergraduate degree (or major) in math, physics or engineering (MPE) will prefer jobs with less youth interaction. Figure 8 shows the marginal means for different levels of youth interaction split by whether they are doing an MPE undergraduate degree or not. However, there is no clear difference in preferences for interaction with youth across the two groups (p=0.26). For completeness, we report this analysis for all attributes in Appendix J. In line with our pre-registration, we recruited a small boost sample of MPE undergraduates through online recruitment efforts focused on universities in England. The motivation for this was to increase the precision of our MPE subgroup analyses, but results remained unchanged when tested with an expanded MPE sample (n=315) including additional participants recruited specifically for MPE representation (p = 0.37) (Appendix K).

Fifth, we hypothesized (H15) that male respondents will prefer jobs with higher salaries more than female respondents. Figure 9 shows the marginal means for different levels of salary split by whether they are male or female. Contrary to our hypothesis, and somewhat surprisingly given the amount of prior evidence on this point, we do not observe clear differences in intensity of preference for different salary levels between males and females (p=0.055). For completeness, we report this analysis for all attributes in Appendix L.

Figure 8

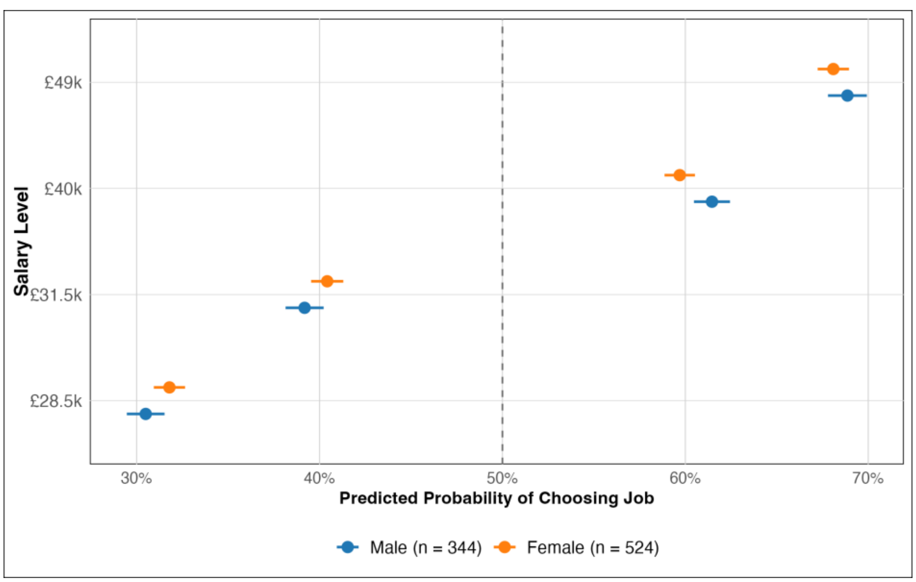
Differences in intensity of preference for youth interaction by academic background



Note. UK survey experiment (n = 871). Coefficients are marginal means. Vertical line at 50% indicates indifference. Error bars show ±1 standard error. Standard errors clustered by respondent. MPE = Mathematics, Physics or Engineering students (n = 190). Non-MPE = All other subjects (n = 681).

Figure 9

Differences in intensity of preference for salary by gender



Note. UK survey experiment (n = 871). Coefficients show marginal means. Vertical line at 50% indicates indifference. Error bars show ±1 standard error. Standard errors clustered by respondent. Male (n = 344). Female (n = 524).

Discussion

We set out to provide new evidence on why people do (not) enter teaching to better inform school leaders and policymakers trying to address shortages. Consistent with the existing literature (Fray & Gore, 2018; See et al., 2022) we found that undergraduates in the UK and US exhibit altruistic (what C&M would call mission-driven) motives in their preferences for jobs, in that they were willing to select jobs with lower salaries (-£3k or -\$4.6k) if the job had significant social impact. However, in stark contrast with the existing literature on entry to teaching, we found that extrinsic motives (what C&M would call pay and costs) were consistently more important determinants of undergraduates' job preferences than altruistic or intrinsic motives. This finding held across studies conducted in the UK and the US. Crucially, we calibrated the value of the attributes in our choice tasks to reflect those of teaching and non-teaching graduate jobs, meaning that they are reflective of the trade-offs faced by undergraduates facing the choice to teach. We did not find large interpersonal differences in the weight (ϴ in the C&M model) that different types of individuals place on different job attributes.

Why do our findings on the importance of extrinsic rewards contrast with those from existing literature? One possibility is that as undergraduates progress into pre-service teacher education/training courses, those with strong extrinsic motives select out of the teaching pipeline leaving those with more altruistic motives. Those who do remain in the pipeline may also be further socialized into the values of the profession during their training (Chao et al., 1994). However, when we restricted our analysis to those considering or planning to teach, we found that extrinsic motives remained the most important. Our findings from RQ2 also suggest that heterogeneity in preferences across undergraduates is in any case limited. The subset of existing research conducted with students, as opposed to trainee teachers, also tends to find that altruistic or intrinsic motives are the most important for entry to teaching (Gore et al., 2016; See et al., 2022). Taken together, this evidence suggests that selection effects are therefore unlikely to explain the discrepancy in results.

A second potential explanation for our contrasting results relates to research methods. As previously discussed, the existing literature on entry to teaching is dominated by self-report research, much of it conducted with the FIT-Choice scale. There are long-standing concerns among social scientists about whether participants in such research tend to exaggerate more socially desirable answers, such as being altruistically motivated (Nederhof, 1985). Survey experiments of the sort used in the present research are thought to mitigate social desirability bias in that they force participants to choose between two jobs, both of which include socially desirable characteristics. Horiuchi et al (2022) provide evidence from two studies showing that survey experiments reduced social desirability bias on sensitive attributes by around two thirds. Since our paper is, to our knowledge, the first in the literature on entry to teaching to use a survey experiment, this could explain why our findings contrast with the rest of the literature.

Of course, our research has its own limitations. Foremost among these is that participants were choosing between hypothetical jobs, rather than real jobs. It has been shown in empirical research that respondents often express higher values for attributes when the choices they face do not include real choices or monetary costs (Harrison, 2024). However, it is somewhat reassuring that results from job choice survey experiments have been shown to predict real world job choices (Maestas et al., 2023; Viano et al., 2021; Wiswall & Zafar, 2018). A second limitation is that participants in our experiment had access to full information about the options presented to them in the choice tasks. In practice, people may be more or less informed about the benefits and costs involved in teaching. Third, we studied jobs as a set of ten attributes. While we carefully reviewed the literature to select the ten attributes that seem most likely to affect undergraduates' choices, jobs clearly differ in more ways than we were able to capture in our simplified experimental set up. Studying jobs as bundles of attributes also has concomitant upsides in that it allows researchers to experimentally isolate the effect of each attribute.

Implications

Notwithstanding these limitations, this research has a number of practical implications. First, policymakers should not assume that the types of people who could potentially be persuaded to enter teaching have a fundamentally different motivational profile to the rest of the workforce. What makes a job attractive to those who are considering teaching shows considerable overlap with what makes a job attractive to undergraduates in general. It is notable, for example, that recruitment campaigns aimed at persuading more undergraduates to enter the teaching profession often emphasize the altruistic or meaning-related aspects of the job.[vi] However, our results suggest that recruitment campaigns for teaching should look more similar to those used in other occupations, since the motivational profile of potential teachers is more similar to those of undergraduates more generally than previously realized. At the very least, such campaigns should take a more balanced approach, emphasizing the extrinsic rewards alongside the many altruistic and meaningful aspects of teaching work. For example, although the nature of teachers' paid leave differs across school systems, school holidays mean that teachers usually qualify for considerably more paid leave than other professions. Our research suggests this is highly attractive to undergraduates.

Second, policymakers should focus on improving the extrinsic rewards of teaching. This is because the extrinsic rewards (for example, paid leave) are more malleable and because, as we showed in RQ1, undergraduates are more sensitive to changes in extrinsic rewards. For example, undergraduates are very sensitive to differences in working hours. However, working hours are high in teaching in many countries (Jerrim & Sims, 2019). Likewise, in line with the quasi-experimental literature on teacher pay (Biasi, 2025), our findings suggest that undergraduates' job choices are highly sensitive to known differences in starting pay and even somewhat sensitive to bonuses paid after two years in the profession. Where possible, policymakers looking to address shortages should increase pay and then clearly communicate this to potential new recruits. Even where there is limited budgetary scope to increase teacher pay, our research suggests potentially cost-neutral routes to improving recruitment. Although the details differ across jurisdictions, teachers are often paid generous public sector pensions (Dolton et al., 2019). However, our results show that a dollar of additional pensions payment on retirement has a much smaller effect on undergraduates' job choices than an additional dollar of salary now. Rebalancing teachers' compensation earlier in their career (away from pensions into salary) would therefore improve recruitment.

Conclusion

Using a survey experiment, which mitigates social desirability bias, we found that undergraduates' job choices are largely affected by extrinsic rewards. Contrary to existing research, which largely uses self-report methods, we found that this also holds true among those who are considering or planning to teach. The research therefore serves as a corrective to the commonly expressed view (Fray & Gore, 2018; See et al., 2022) that those who enter teaching are primarily motivated by altruistic concerns. Policymakers looking to address shortages should improve the extrinsic rewards of teaching and emphasize these, alongside the many altruistic and meaning-related aspects of the job, when trying to persuade undergraduates to enter the profession.

References

Adams, L., Coburn-Crane, S., Sanders-Earley, A., Keeble, R., Harris, H., Taylor, J., & Taylor, B. (2023). Working lives of teachers and leaders–wave 2. *Department for Education*.

Albert, D. A., & Smilek, D. (2023). Comparing attentional disengagement between Prolific and MTurk samples. *Scientific Reports*, *13*(1), 20574.

Allen, B. et al. (2025). Evaluating teacher recruitment strategies using fully randomised paired conjoint experiments. Education Endowment Foundation.

APA (nd.) https://dictionary.apa.org/personality

Ariely, D., Kamenica, E., & Prelec, D. (2008). Man's search for meaning: The case of Legos. *Journal of Economic Behavior & Organization*, *67*(3-4), 671-677.

Bansak, K., Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2018). The number of choice tasks and survey satisficing in conjoint experiments. *Political Analysis*, *26*(1), 112–119.

Bansak, K., Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2021a). Beyond the breaking point? Survey satisficing in conjoint experiments. *Political Science Research and Methods*, *9*(1), 53–71.

Bansak, K., Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2021b). Conjoint Survey Experiments. In J. Druckman & D. P. Green (Eds.), *Advances in Experimental Political Science* (1st ed., pp. 19–41). Cambridge University Press.

Bartanen, B., Avitabile, A., & Kwok, A. (2025). Who wants to be a teacher in America?. EdWorkingPaper 25-1275.

Bartling, B., Fehr, E., & Schmidt, K. M. (2013). Discretion, productivity, and work satisfaction. *Journal of Institutional and Theoretical Economics*, 4-22.

Battaglio, R. P., Belle, N., & Cantarelli, P. (2022). Self-determination theory goes public: experimental evidence on the causal relationship between psychological needs and job satisfaction. *Public Management Review*, *24*(9), 1411-1428.

Benhenda, A. (2022). Absence, substitutability and productivity: Evidence from teachers. *Labour Economics*, *76*, 102167.

Benz, M., & Frey, B. S. (2008). Being independent is a great thing: Subjective evaluations of self-employment and hierarchy. *Economica*, *75*(298), 362-383.

Biasi, B. (2025). Teacher compensation. *Live handbook of education policy research. Association for Education Finance and Policy.* *https://livehandbook.org/k-12-education/workforce-teachers/teacher-compensation/*

Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S., & Wyckoff, J. (2011). The influence of school administrators on teacher retention decisions. *American Educational Research Journal, 48*(2), 303– 333.

Bruno, P. (2025). Teacher supply and turnover. *Live handbook of education policy research. Association for Education Finance and Policy. https://livehandbook. org/k-12-education/workforce-teachers/teacher-supply-and-turnover*.

Bryce, A. (2018). *Finding meaning through work: eudaimonic well-being and job type in the US and UK* (No. 2018004). Sheffield Economic Research Paper Series.

Burge, P., Lu, H., & Phillips, W. D. (2021). Understanding Teaching Retention: Using a discrete choice experiment to measure teacher retention in England. Rand.

Carpenter, J., & Gong, E. (2016). Motivating agents: How much does the mission matter?. *Journal of Labor Economics*, *34*(1), 211-236.

Cassar, L., & Meier, S. (2018). Nonmonetary Incentives and the Implications of Work as a Source of Meaning. *Journal of Economic Perspectives*, *32*(3), 215–238.

Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization*, *90*, 123-133.

Chao, G. T., O'Leary-Kelly, A. M., Wolf, S., Klein, H. J., & Gardner, P. D. (1994). Organizational socialization: Its content and consequences. *Journal of Applied Psychology*, *79*(5), 730.

Cumiskey, L. (2024) Cambridge demands delay to new teacher training framework, *Schools Week*, 31 January, https://schoolsweek.co.uk/cambridge-demands-delay-to-new-teacher-training-framework/?s=03.

Deci, E. L., Olafsen, A. H., & Ryan, R. M. (2017). Self-Determination Theory in Work Organizations: The State of a Science. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*(1), 19–43. https://doi.org/10.1146/annurev-orgpsych-032516-113108

Deci, E. L., & Ryan, R. M. (2000). The 'What' and 'Why' of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, *11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Dee, T. S., & Goldhaber, D. (2017). Understanding and addressing teacher shortages in the United States. *The Hamilton Project*, *5*, 1-28.

Dolton, P., Samek, L., She, P., & National Institute of Social and Economic Research. (2019). Total reward and pensions in the UK by occupation in the public and private sectors. NIESR.

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *Plos one*, *18*(3), e0279720.

Florczak, C., Rasmussen, S. H. R., Jensen, U. T., Stritch, J. M., Christensen, K., Nørgaard, A. S., & Klemmensen, R. (2023). Exploring the foundational origins of public service motivation through the lens of behavioral genetics. *Public Administration*, *101*(3), 993-1013.

Focquaert, F., Steven, M. S., Wolford, G. L., Colden, A., & Gazzaniga, M. S. (2007). Empathizing and systemizing cognitive traits in the sciences and humanities. *Personality and Individual Differences*, *43*(3), 619-625.

Fray, L., & Gore, J. (2018). Why people choose teaching: A scoping review of empirical studies, 2007–2016. *Teaching and Teacher Education*, *75*, 153-163.

Galanakis, M. D., & Tsitouri, E. (2022). Positive psychology in the working environment. Job demands-resources theory, work engagement and burnout: A systematic literature review. *Frontiers in Psychology*, *13*, 1022102.

Gore, J., Barron, R. J., Holmes, K., & Smith, M. (2016). Who says we are not attracting the best and brightest? Teacher selection and the aspirations of Australian school students. *The Australian Educational Researcher*, *43*(5), 527-549.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528.

Gosnell, G. K., List, J. A., & Metcalfe, R. D. (2020). The impact of management practices on employee productivity: A field experiment with airline captains. *Journal of Political Economy*, *128*(4), 1195-1233.

Grant, A. M. (2008). The significance of task significance: Job performance effects, relational mechanisms, and boundary conditions. *Journal of Applied Psychology*, *93*(1), 108.

Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences*, *112*(8), 2395–2400.

Hainmueller, J., & Hopkins, D. J. (2015). The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes toward Immigrants. *American Journal of Political Science*, *59*(3), 529–548.

Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments. *Political Analysis*, *22*(1), 1–30.

Harrison, G. W. (2024). Real choices and hypothetical choices. In *Handbook of Choice Modelling* (pp. 246-275). Edward Elgar Publishing.

Hole, A. R. (2007a). A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Economics*, *16*(8), 827-840.

Hole, A. R. (2007b). wtp: Stata module to estimate confidence intervals for willingness to pay measures. Statistical Software Components S456808, Department of Economics, Boston College.

Horiuchi, Y., Markovich, Z., & Yamamoto, T. (2022). Does Conjoint Analysis Mitigate Social Desirability Bias? *Political Analysis*, *30*(4), 535–549.

Hurtado Rúa, S. M., Stead, G. B., & Poklar, A. E. (2019). Five-factor personality traits and RIASEC interest types: A multivariate meta-analysis. *Journal of Career Assessment*, *27*(3), 527-543.

Jerrim, J., & Sims, S. (2019). The teaching and learning international survey (TALIS). UK Government Department for Education.

Johnston, A. C. (2021). *Preferences, selection, and the structure of teacher pay*. SSRN 3532779.

Jost, M., & Möser, S. (2023). Salary, flexibility or career opportunity? A choice experiment on gender specific job preferences. *Frontiers in Sociology*, *8*, 1154324.

Kim, S., Vandenabeele, W., Wright, B. E., Andersen, L. B., Cerase, F. P., Christensen, R. K., ... & De Vivo, P. (2013). Investigating the structure and meaning of public service motivation across populations: Developing an international instrument and addressing issues of measurement invariance. *Journal of Public Administration Research and Theory*, *23*(1), 79-102.

Kraft, M. A., Marinell, W. H. and Shen-Wei Yee, D. (2016) School organizational contexts, teacher turnover, and student achievement: evidence from panel data. *American Educational Research Journal, 53* (5), 1411–1449.

Kraft, M. A., & Lyon, M. A. (2024). The rise and fall of the teaching profession: Prestige, interest, preparation, and satisfaction over the last half century. *American Educational Research Journal*, *61*(6), 1192-1236.

Krefeld-Schwalb, A., Sugerman, E. R., & Johnson, E. J. (2024). Exposing omitted moderators: Explaining why effect sizes differ in the social sciences. *Proceedings of the National Academy of Sciences*, *121*(12), e2306281121.

Ladd, H. F. (2011). Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement?. *Educational Evaluation and Policy Analysis*, *33*(2), 235-261.

Leeper, T. J., Hobolt, S. B., & Tilley, J. (2020). Measuring subgroup preferences in conjoint experiments. *Political Analysis*, *28*(2), 207-221.

Lentini, V., Gimenez, G., & Valbuena, J. (2024). Teachers' preferences for incentives to work in disadvantaged districts: A discrete choice experiment in Costa Rica. *Economic Analysis and Policy, 82*, 831-845.

Levatino, A., Ferrer-Esteban, G., & Verger, A. (2024). Unveiling teachers' work preferences: A conjoint experiment on the implications of school governance reform across three countries. *Teaching and Teacher Education, 146*, 104631.

Lovison, V. S., & Hyunjung Mo, C. (2024). Investing in the teacher workforce: Experimental evidence on teachers' preferences. *American Educational Research Journal, 61*(1), 108-144.

Maestas, N., Mullen, K. J., Powell, D., Von Wachter, T., & Wenger, J. B. (2023). The value of working conditions in the United States and implications for the structure of wages. *American Economic Review*, *113*(7), 2007-2047.

Navarro-Asencio, E., Lopez-Martin, E., Asensio-Munoz, I., Exposito-Casas, E., Carpintero-Molina, E., & De Miguel, C. R. (2021). Meta-analysis of reliability generalisation of the FIT-Choice questionnaire (Factors Influencing Teaching Choice). *Revista de Educación*, *393*, 223-251.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, *15*(3), 263-280.

O'Connor, B. P. (2002). A quantitative review of the comprehensiveness of the five-factor model in relation to popular personality inventories. *Assessment*, *9*(2), 188-203.

OECD (2019). TALIS 2018 Results (Volume 1). Annex C Table I.3.63.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153-163.

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643-1662.

Peer, E., Rothschild, D., & Gordon, A. (2023). Platform over procedure: online platforms that pre-vet participants yield higher data quality without sacrificing diversity. *PsyArXiv*, 1-10.

Perry, J. L., & Wise, L. R. (1990). The motivational bases of public service. *Public Administration Review*, 367-373.

Perry, J. L. (1996). Measuring Public Service Motivation: An Assessment of Construct Reliability and Validity. *Journal of Public Administration Research and Theory: J-PART*, *6*(1), 5–22.

Richardson, P. W., & Watt, H. M. (2006). Who chooses teaching and why? Profiling characteristics and motivations across three Australian universities. *Asia-Pacific Journal of Teacher Education*, *34*(1), 27-56.

Ripoll, G., Ballart, X., Hernández, E., & Vandenabeele, W. (2023). 'It's a match!': a discrete choice experiment on job attractiveness for public service jobs. *Public Management Review*, 1-35.

Ritz, A., Brewer, G. A., & Neumann, O. (2016). Public service motivation: A systematic literature review and outlook. *Public administration review*, *76*(3), 414-426.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, *3*(2), 135–146. https://doi.org/10.1093/oxfordjournals.oep.a041827

De Schouwer, T., & Kesternich, I. (2022). *It's All About the Meaning: Work Meaning, Job Flexibility, and the Gender Wage Gap*. KU Leuven, Department of Economics.

Schanzenbach, D. W. (2006). What have researchers learned from Project STAR?. *Brookings Papers on Education Policy, 9*, 205-228.

Schuessler, J., & Freitag, M. (2020). Power analysis for conjoint experiments. *SocArXiv. doi*, *10*.

See, B. H., Munthe, E., Ross, S. A., Hitt, L., & El Soufi, N. (2022). Who becomes a teacher and why?. *Review of Education*, *10*(3), e3377.

Sims, S., & Benhenda, A. (2022). *The effect of financial incentives on the retention of shortage-subject teachers: evidence from England* (No. 22-04). UCL Centre for Education Policy and Equalising Opportunities.

Sims, S. (2018). *Essays on the recruitment and retention of teachers* (Doctoral dissertation, UCL (University College London)).

Sims, S., Lowes-Belk, H., & Routledge, C. (2025). Why do people choose to enter and exit the teaching profession? An interdisciplinary quantitative synthesis. *British Educational Research Journal*.

Törnroos, M., Jokela, M., & Hakulinen, C. (2019). The relationship between personality and job satisfaction across occupations. *Personality and Individual Differences*, *145*, 82-88.

Valet, P., Sauer, C., & Tolsma, J. (2021). Preferences for work arrangements: A discrete choice experiment. *PloS one*, *16*(7), e0254483.

Vedel, A., Thomsen, D. K., & Larsen, L. (2015). Personality, academic majors and performance: Revealing complex patterns. *Personality and Individual Differences, 85*, 69-76.

Viano, S., Pham, L. D., Henry, G. T., Kho, A., & Zimmer, R. (2021). What Teachers Want: School Factors Predicting Teachers' Decisions to Work in Low-Performing Schools. *American Educational Research Journal*, *58*(1), 201–233.

Vogel, D., & Kroll, A. (2016). The stability and change of PSM-related values across time: Testing theoretical expectations against panel data. *International Public Management Journal*, *19*(1), 53-77.

Westwood, S. (2025). The potential existential threat of large language models to online survey research. *PNAS*. https://doi.org/10.1073/pnas.2518075122

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68-81.

Wiswall, M., & Zafar, B. (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, *133*(1), 457-507.

Worth, J., & Faulkner-Ellis, H. (2021). Teacher Labour Market in England: Annual Report 2021. *National Foundation for Educational Research*.

Woźniak-Jęchorek, B., d'Urso, A. S., & Thurston, C. N. (2022). *Remote work preferences of American employees: evidence from conjoint survey experiment*. https://www.amandadurso.com/s/Remote-Work-Preferences-of-American-Employees.pdf

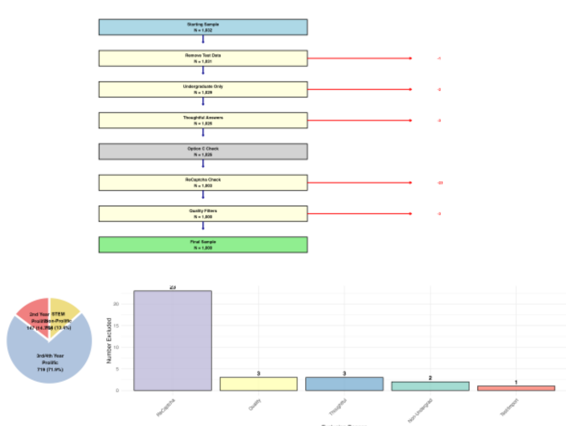
Appendices

Appendix A – Data quality checks

Across both the UK and US surveys, we implemented a number of pre-registered additional checks to help ensure the quality of our data. First, we double-checked that respondents were still engaged in undergraduate study at the time of responding and removed those that were not. Second, we included a Captcha exercise and excluded 23 UK and 19 US responses with a score of <0.5. Third, we checked for suspiciously fast responses (<90 seconds for UK and <60 seconds for US) and excluded responses that were faster. Fourth, we checked for attentiveness by asking participants whether they commit to providing thoughtful answers in the survey and also included a question that asks respondents to select a particular response (e.g., "Please select option C") to a given item. Respondents that failed either of these checks were excluded.

Figure A1

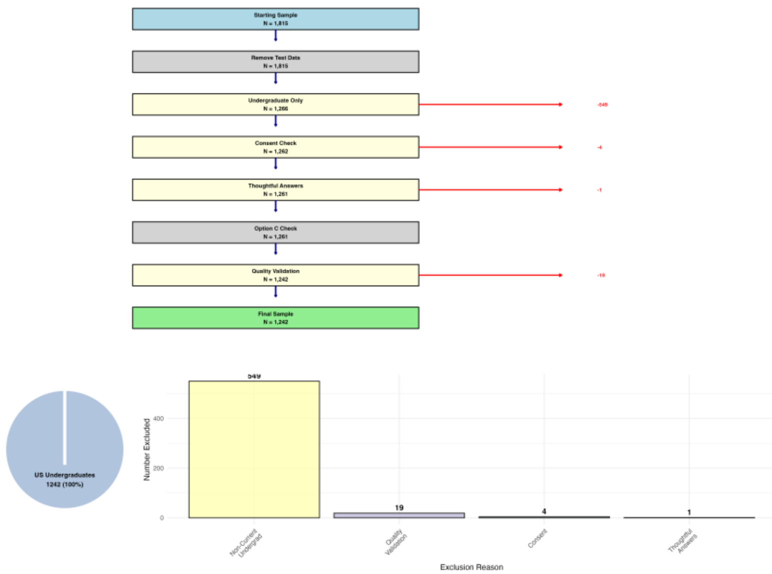
UK Participant Exclusion Process.



Note. This flowchart shows the complete exclusion process for the UK undergraduate sample (N = 1,032 initial). Pre-registered exclusion criteria were applied in the order participants encountered them during the survey, resulting in a final sample of 1,000 participants (96.9% retention rate).

Figure A2

US Participant Exclusion Process.



Note. This flowchart shows the complete exclusion process for the US undergraduate sample (N = 1,815 initial). Pre-registered exclusion criteria were applied in the order participants encountered them during the survey, resulting in a final sample of 1,242 participants (68.4% retention rate).

Appendix B – US attribute values

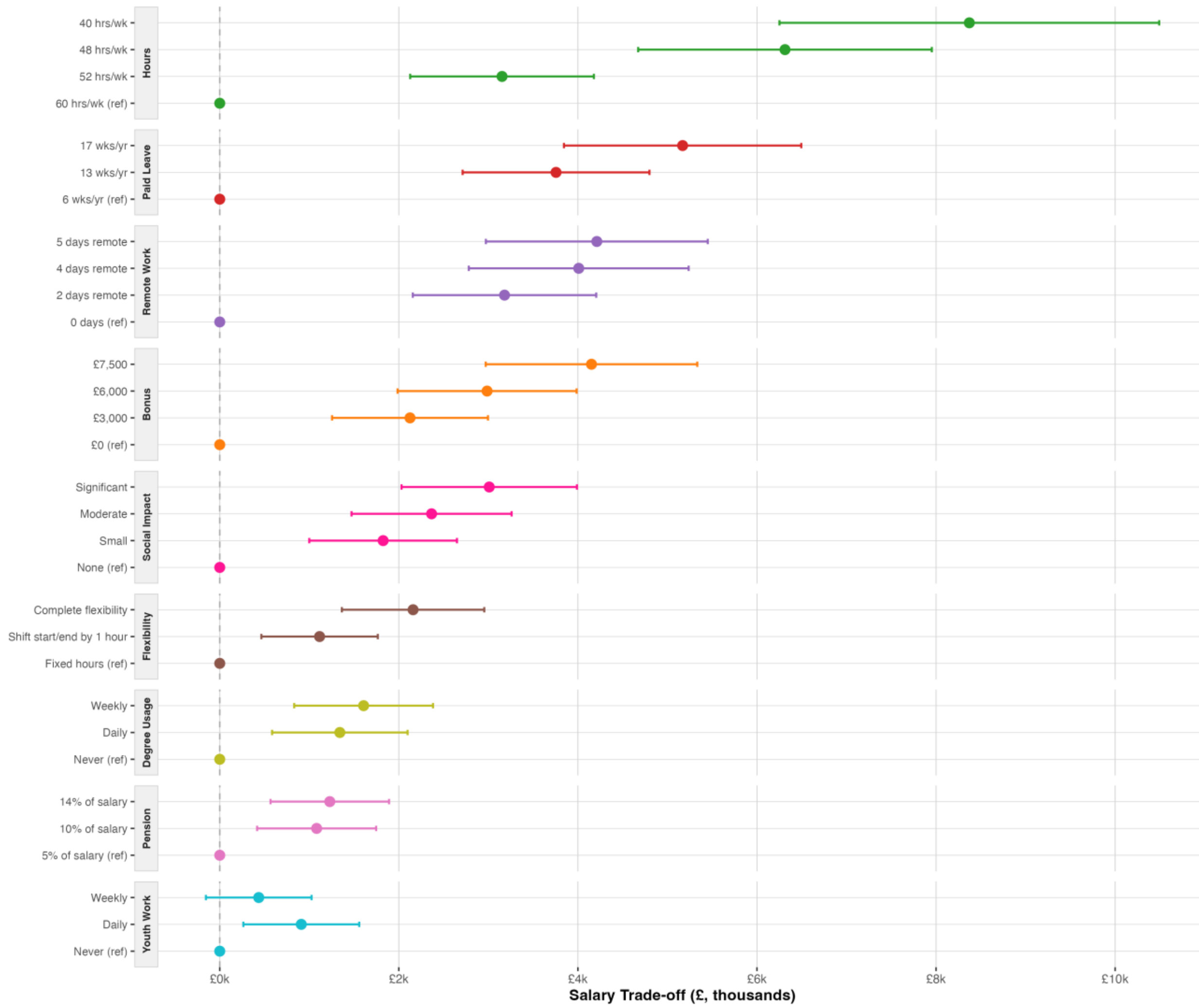
Attributes and attribute values for the US survey experiment.

C&M	FIT	Attribute	Value 1	Value 2	Value 3	Value 4
Costs	Task demand	Typical weekly working hours	40 hours	48 hours	52 hours	60 hours
	Personal utility	Paid Time Off (PTO) per year	11 days	14 days	70 days	-
		Number of remote workdays per week	0 days per week remote	2 days per week remote	4 days per week remote	5 days per week remote
		Flexibility over working hours	Fixed working hours	Ability to shift start/end times forward or back by one hour	Complete flexibility over working hours, within demands of the role	-
Meaning	Task demand	Uses knowledge from undergraduate degree	Never	Weekly	Daily	-
	Social utility	Frequency of working with young people	Never	Weekly	Daily	-
		Level of social or community impact you can make	No impact	Small impact	Moderate impact	Significant impact
Income	Task reward	Starting salary per year (before tax)	$35k	$45k	$55k	$61k
		Employer contribution to your retirement plan each month	3% of salary	5% of salary	10% of salary	14% of salary
		One-time retention bonus after two years	$0k	$5k	$10k	$12k

Note: Each hypothetical job presented to respondents in the choice tasks comprised the ten attributes listed above. For each hypothetical job, the value for each of these attributes was randomly selected from the options shown in the rows.

Appendix C – UK Salary equivalent values (SEV)

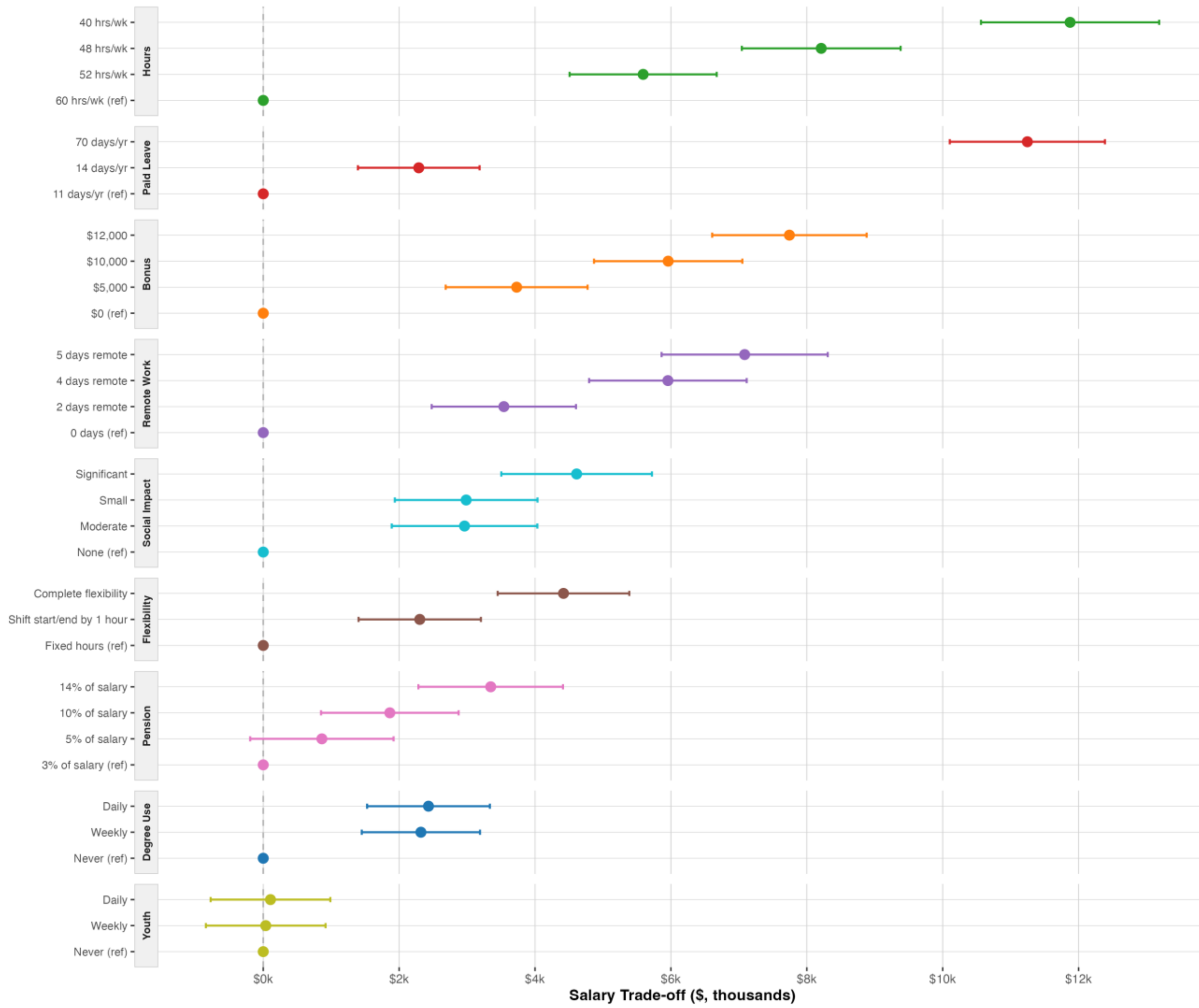
SEV for the UK survey experiment.



Note. Values show salary-equivalent trade-offs calculated using average marginal component effects. Horizontal lines show 95% confidence intervals with standard errors clustered at respondent level, calculated using the delta method. SEV represents how much salary graduates would accept as reduction to obtain each job attribute improvement relative to the baseline category. N=871 unique respondents, each responding to 10 paired profile choice tasks.

Appendix D – US Salary equivalent values (SEV)

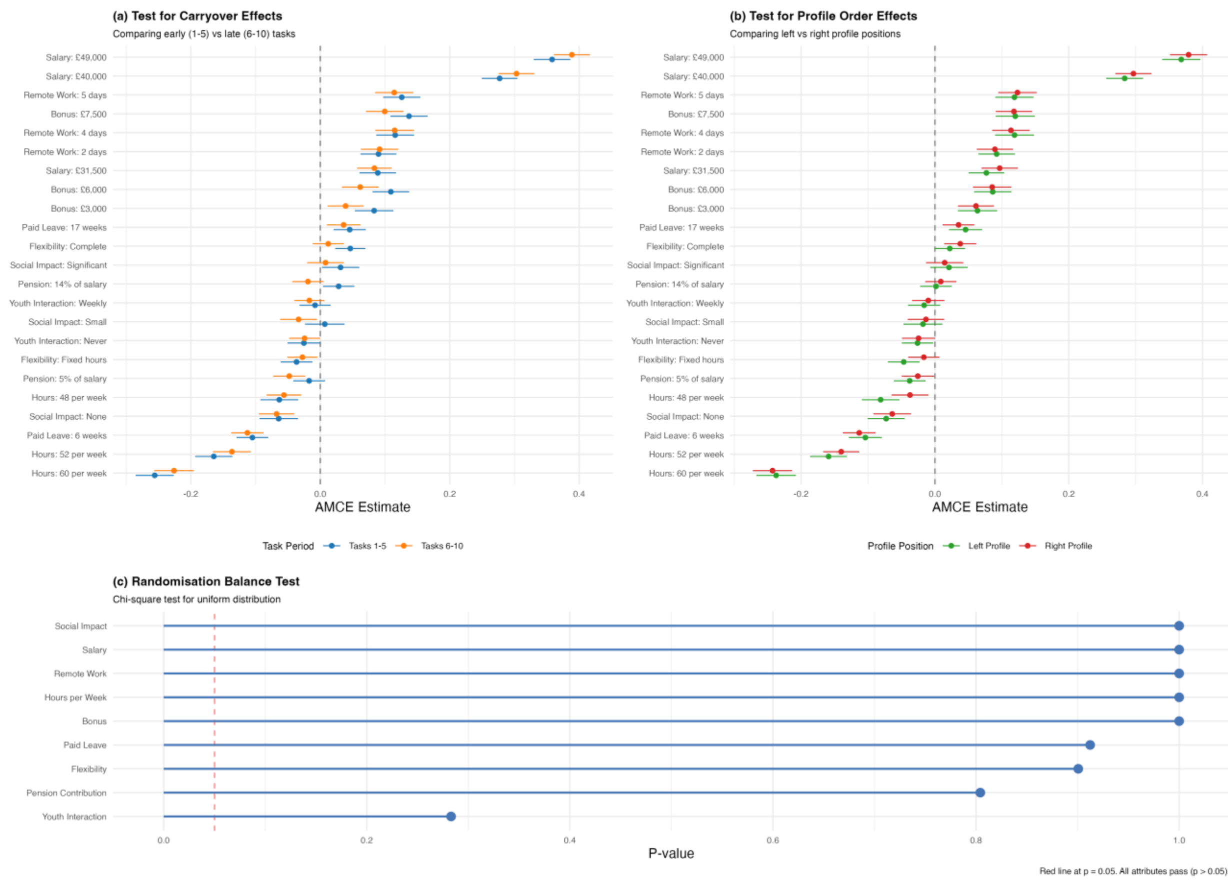
SEV for the US survey experiment.



Note. Values show salary-equivalent trade-offs calculated using average marginal component effects. Horizontal lines show 95% confidence intervals with standard errors clustered at respondent level, calculated using the delta method. SEV represents how much salary graduates would accept as reduction to obtain each job attribute improvement relative to the baseline category. N=1,242 unique respondents, each responding to 10 paired profile choice tasks.

Appendix E – UK Robustness

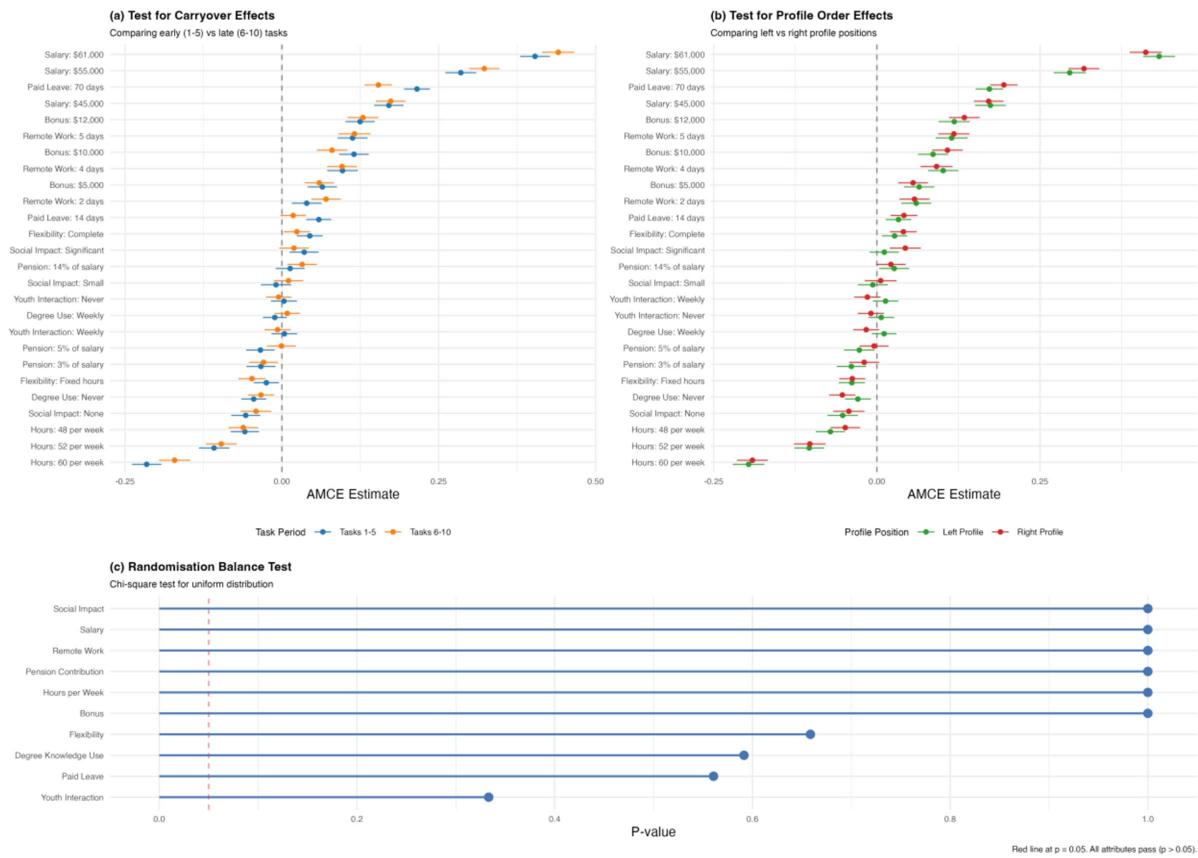
Robustness checks for the UK survey experiment.



Note. Prolific participants only (n = 871). Error bars represent 95% confidence intervals. Panel (a) examines carryover effects by comparing AMCE estimates from early tasks (1-5) versus late tasks (6-10). Panel (b) tests for profile order effects by comparing estimates when jobs appear in left versus right positions. Panel (c) validates randomization balance through chi-square tests examining whether each attribute level appears with expected frequency across all choice tasks.

Appendix F – US Robustness

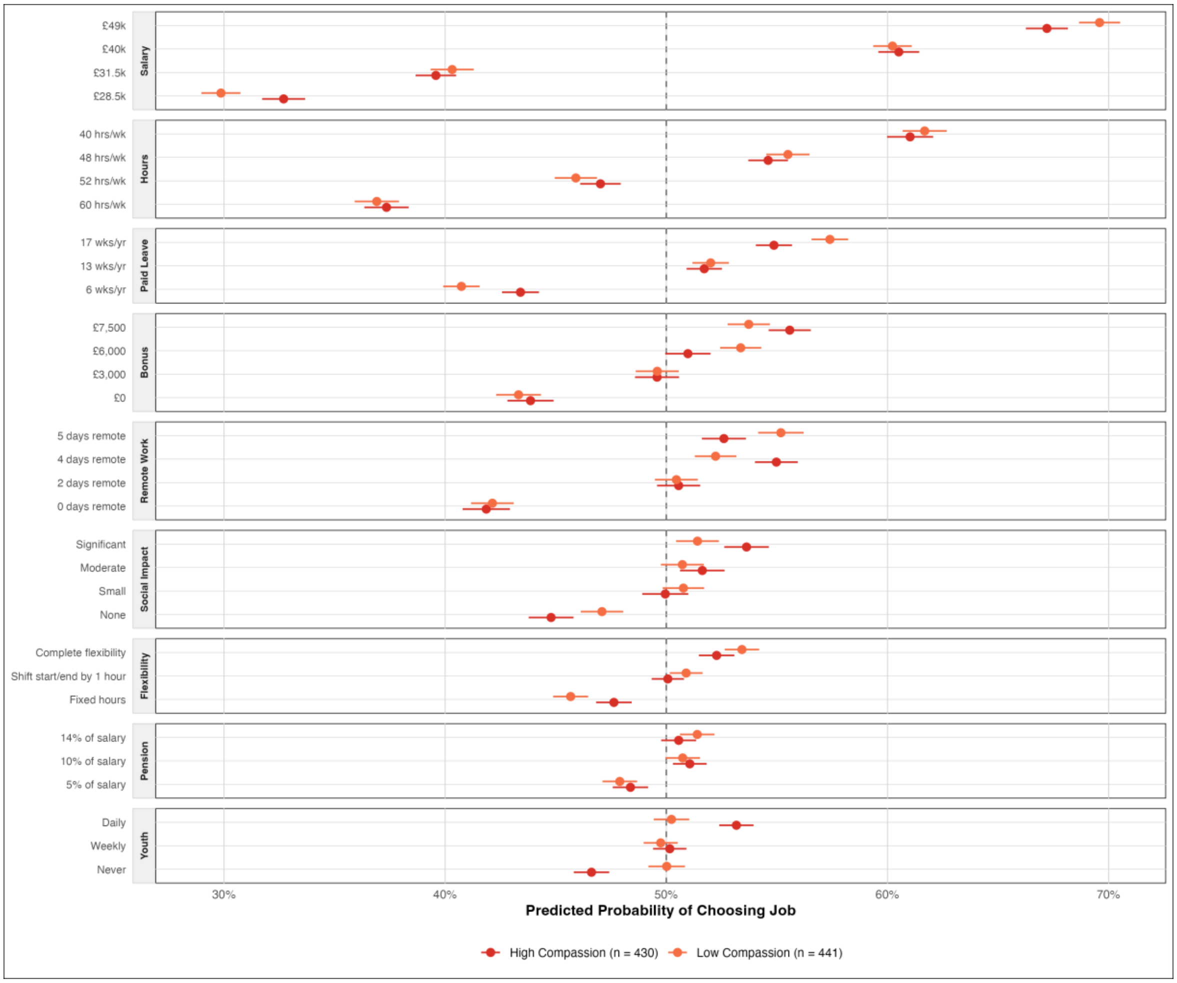
Robustness checks for the US survey experiment.



Note. (n = 1,242). Error bars represent 95% confidence intervals. Panel (a) examines carryover effects by comparing AMCE estimates from early tasks (1-5) versus late tasks (6-10). Panel (b) tests for profile order effects by comparing estimates when jobs appear in left versus right positions. Panel (c) validates randomization balance through chi-square tests examining whether each attribute level appears with expected frequency across all choice tasks.

Appendix G – Compassion

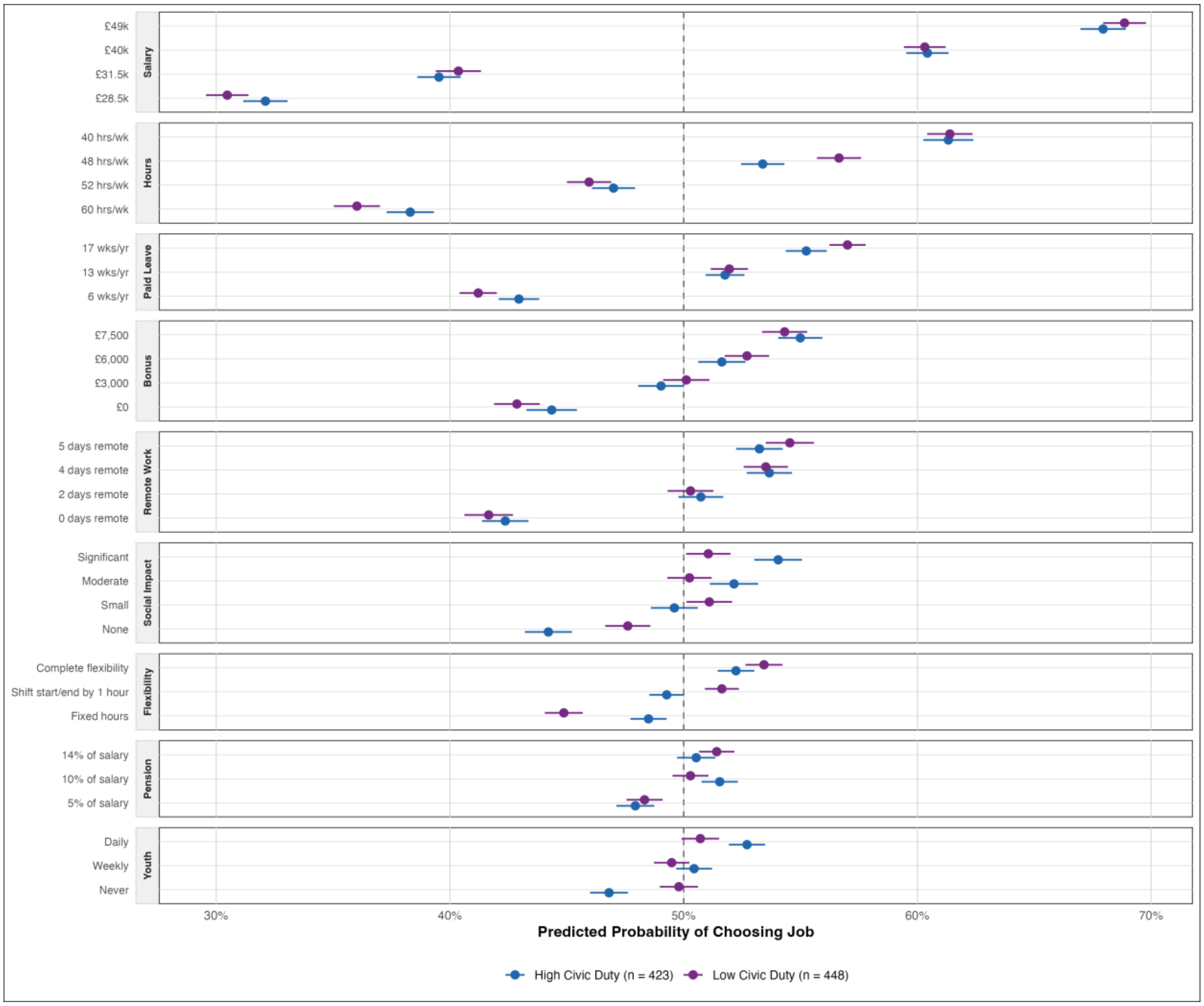
Comparing intensity of preferences across levels of compassion for all attributes (UK sample).



Note. UK survey experiment (n = 871). Coefficients are marginal means. Vertical line at 50% indicates indifference. Error bars show ±1 standard error with standard errors clustered by respondent. 'High Compassion' includes respondents above the median PSM-compassion factor score from confirmatory factor analysis (n = 430). 'Low Compassion' includes those below the median (n = 441).

Appendix H – Civic duty

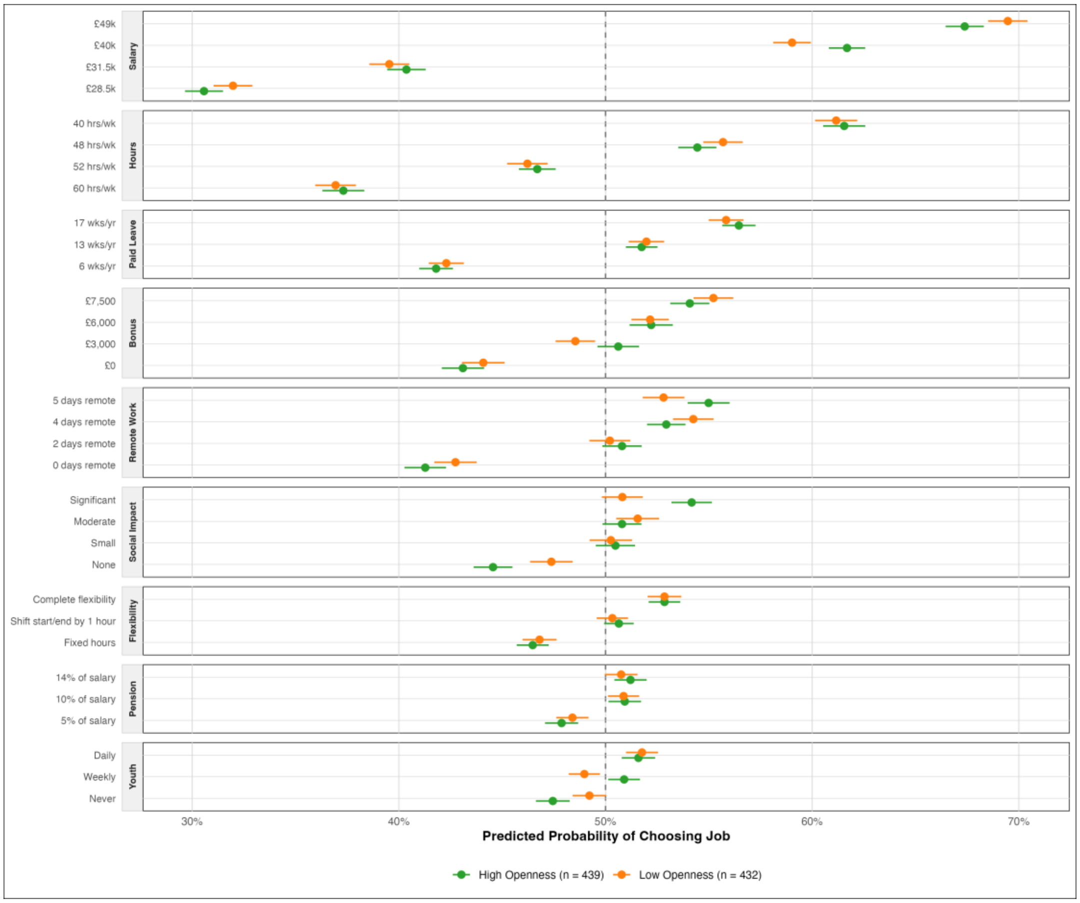
Comparing intensity of preferences across levels of compassion for all attributes (UK sample).



Note. UK survey experiment (n = 871). Coefficients are marginal means. Vertical line at 50% indicates indifference. Error bars show ±1 standard error with standard errors clustered by respondent. 'High Civic Duty' includes respondents above the median PSM-civic duty factor score from confirmatory factor analysis (n = 423). 'Low Civic Duty' includes those below the median (n = 448).

Appendix I – Openness

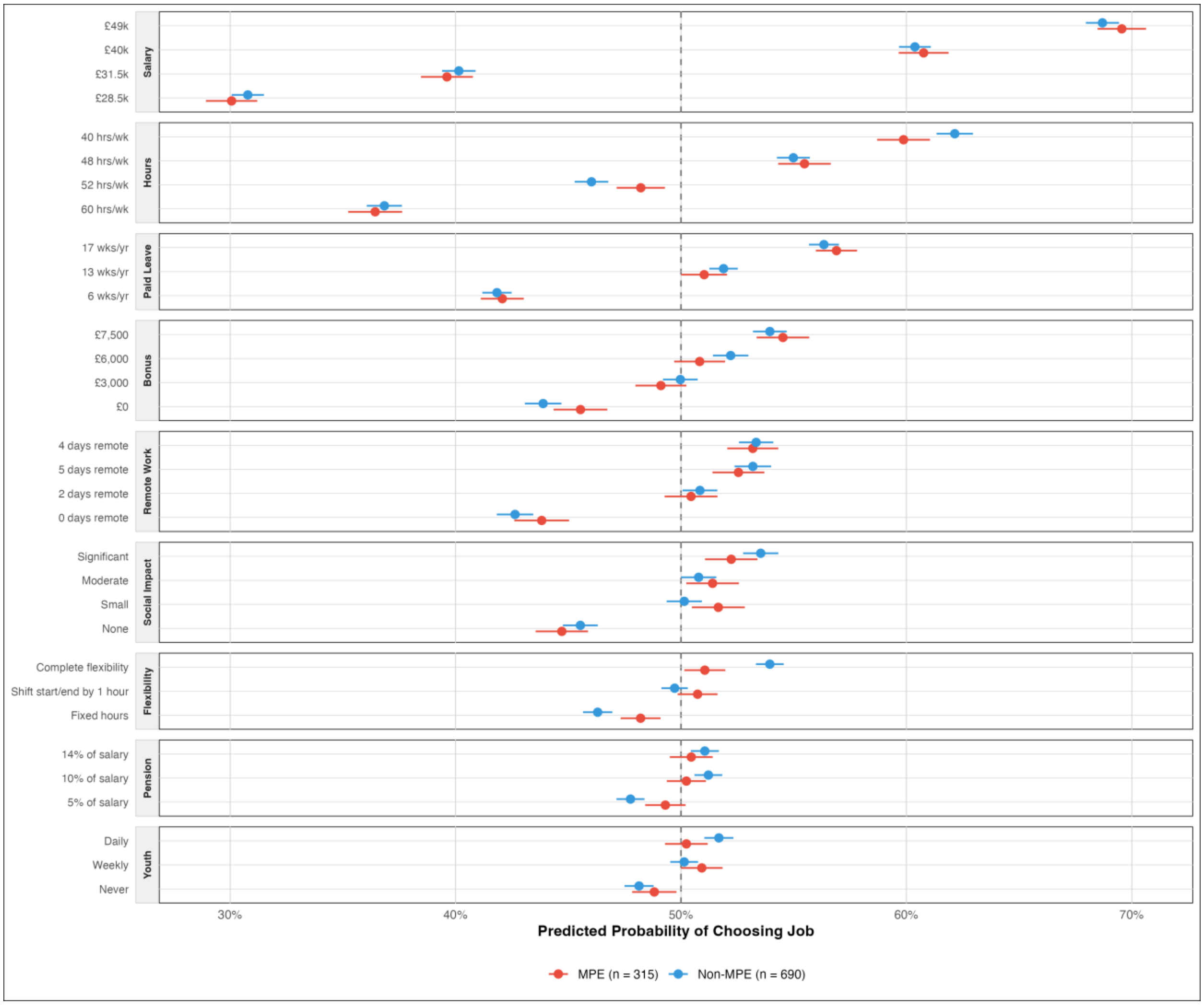
Comparing intensity of preferences across levels of openness for all attributes (UK sample).



Note. UK survey experiment (n = 871). Coefficients are marginal means. Vertical line at 50% indicates indifference. Error bars show ±1 standard error with standard errors clustered by respondent. 'High Openness' includes respondents above the median openness to experience score from 10-item IPIP Big Five scale (n = 439). 'Low Openness' includes those below the median (n = 432).

Appendix J – Math/physics/engineering (MPE) undergrads

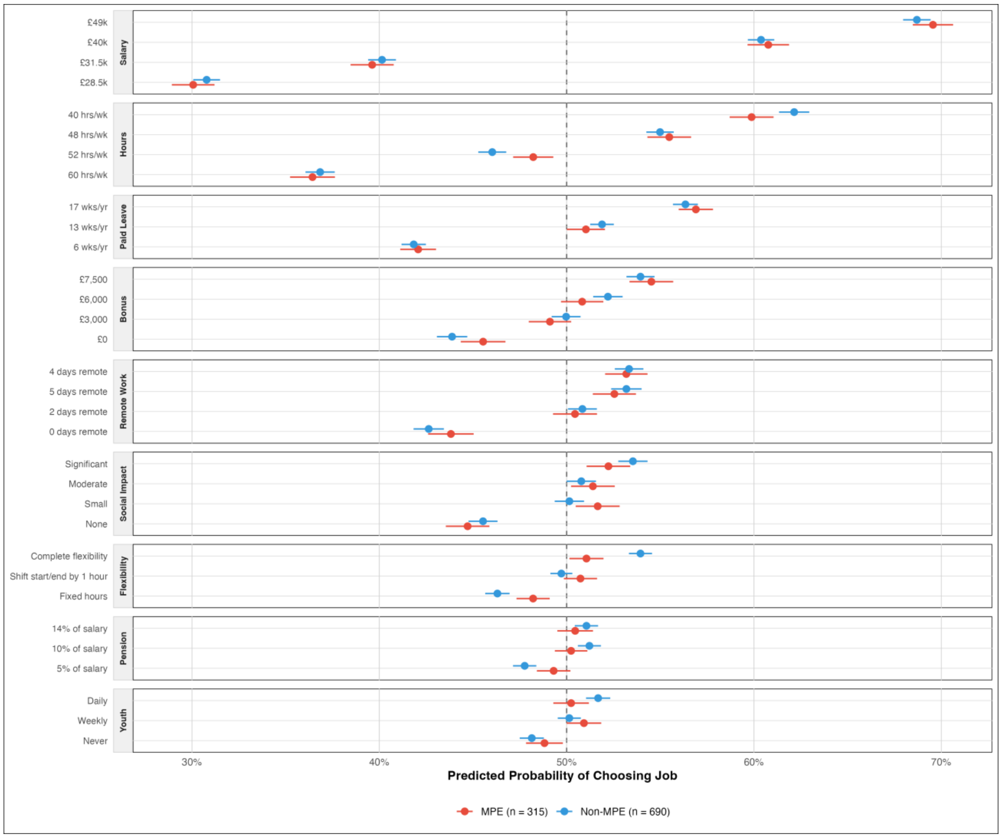
Comparing intensity of preferences between MPE and not for all attributes (UK sample).



Note. UK survey experiment (n = 871). Coefficients are marginal means. Vertical line at 50% indicates indifference. Error bars show ±1 standard error with standard errors clustered by respondent. 'High Openness' includes respondents above the median openness to experience score from 10-item IPIP Big Five scale (n = 439). 'Low Openness' includes those below the median (n = 432).

Appendix K – MPE not MPE boost sample

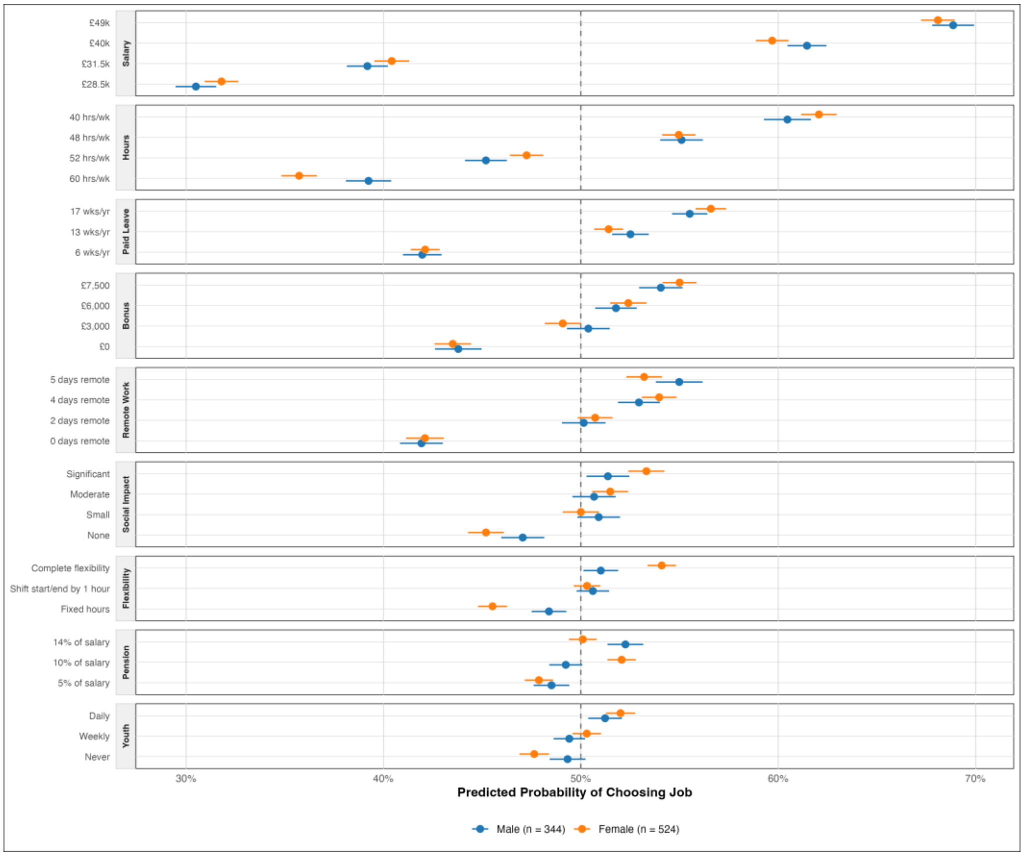
Comparing intensity of preferences between MPE and not for all attributes (UK sample, including STEM boost sample).



Note. UK survey experiment including MPE boost sample (n = 1,005). Coefficients are marginal means. Vertical line at 50% indicates indifference. Error bars show ±1 standard error with standard errors clustered by respondent. MPE = Mathematics, Physics or Engineering students (n = 315). Non-MPE = All other subjects (n = 690).

Appendix L – Gender

Comparing intensity of preferences across levels of compassion for all attributes (UK sample).



Note. UK survey experiment (n = 868). Coefficients are marginal means. Vertical line at 50% indicates indifference. Error bars show ±1 standard error with standard errors clustered by respondent. Male (n = 344). Female (n = 524).

---

[i] https://www.prolific.com/resources/why-participants-get-banned
[ii] https://markusfreitag.shinyapps.io/cjpowr/
[iii] https://www.hesa.ac.uk/news/13-06-2024/sb268-higher-education-graduate-outcomes-statistics/salary
[iv] https://www.tes.com/magazine/analysis/general/teacher-pay-scales-how-much-are-teachers-paid-england
[v] This is 28% of remuneration, which is twice 14%, which is the value actually included in our choice tasks.
[vi] https://www.youtube.com/watch?v=ge268Igl9LQ