Handling multiple demands in academia: Does gender play a role?

Nigel Ronald Francis Burnell

Healthcare Management and Policy Research PhD

2018

Supervisors: Irina Cojuharenco, Jo Blanden, Zahra Murad

© Nigel Burnell
Declaration of Originality

This thesis and the work to which it refers are the results of my own efforts. Any ideas, data, images or text resulting from the work of others (whether published or unpublished) are fully identified as such within the work and attributed to the originator in the text, bibliography or in footnotes. This thesis has not been submitted in whole or in part for any other academic degree or professional qualification. I agree that the University has the right to submit my work to the plagiarism detection service TurnitinUK for originality checks. Whether or not drafts have been so-assessed, the University reserves the right to require an electronic version of the document (as submitted) for assessment as above.
Summary

This thesis is a collection of three related papers examining how academics handle multiple demands in the UK higher education sector, as outlined in chapter one. It suggests why some academics progress faster through their careers than others and why gender imbalances may persist in academia. It uses a behavioural approach to these questions by examining the role of individual preferences, differing responsiveness to student expectations and engagement and biases in evaluations of teaching.

Chapter two reports the effect of individual research, teaching and time preferences on time allocated to research. I find that devoting more time to research is associated with higher levels of seniority. I find that preferences for research and teaching are in conflict and greater preferences for research predict more time allocated to research. Further, there are significant gender differences in research preference that provide a lens to examine a male career advantage in academia.

Chapter three reports a survey experiment testing the motivation of academics devoting more time to extra-curricular teaching effort, when it is in conflict with research time. I test whether student expectations of support and student engagement motivates extra-curricular teaching effort, as well as examine gender differences in the responsiveness to expectations and engagement. I find that expectations and engagement affect extra-curricular teaching effort positively for both male and female academics.

Chapter four tests for gender bias in teaching evaluations, following from findings that female academics may devote less time to research and more time to teaching (chapter two). I find evidence of gender bias and show that gender bias may be eliminated by the academic’s seniority but not by a high warmth teaching style. I find no evidence of less biased evaluations by those who anticipate gender bias.

Chapter five summarizes the findings of this thesis and concludes.
Acknowledgements (up to 1 page A4)

- Firstly I would like to thank Erica, my wife, and my children, Tom and Lia, for supporting me through the completion of this PhD thesis.

- Secondly I would like to thank all of my supervision team over my time at the University of Surrey. Irina Cojuharenco, Alex Gerbasi, Jo Blanden and Graham Cookson have all provided major supervisory input and I am grateful to all of them. Zahra Murad has been a supervisor for the duration of my PhD and provided encouragement and guidance from the outset.

- I would also like to thank Jane Hendy for giving me the opportunity to study at the University of Surrey.

- Finally, I would like to thank the Leverhulme Trust [https://www.leverhulme.ac.uk/](https://www.leverhulme.ac.uk/) (Grant RL2012-681) for providing funding for this research.
Table of Contents

Chapter 1 ................................................................................................................. 1
Introduction .............................................................................................................. 2
Chapter 2 ................................................................................................................ 9
“Understanding the research time allocations of academics and its association with career success.” 10
Abstract ................................................................................................................... 11
Introduction .............................................................................................................. 11
Theory Development ............................................................................................. 15
Study Design .......................................................................................................... 24
Results ..................................................................................................................... 28
Robustness Analysis ............................................................................................ 31
General Discussion ............................................................................................... 32
Tables and Figures ............................................................................................... 41
Appendix A ............................................................................................................ 62
Chapter 3 ............................................................................................................... 72
“Why give more? The role of student expectations and engagement in extra-curricular teaching
effort.” ................................................................................................................... 73
Abstract ................................................................................................................... 74
Introduction .............................................................................................................. 74
Theory Development ............................................................................................. 76
Study Design .......................................................................................................... 82
Results ..................................................................................................................... 90
General Discussion ............................................................................................... 95
Tables and Figures ............................................................................................... 100
Appendix A ............................................................................................................ 121
Chapter 4 ............................................................................................................... 124
“Now you see it now you don’t: The effect of teaching style and seniority on gender bias in teaching
evaluations.” .......................................................................................................... 125
Abstract ................................................................................................................... 126
Introduction .............................................................................................................. 127
Theory Development ............................................................................................. 129
Study 1 .................................................................................................................... 135
Results ..................................................................................................................... 139
Discussion ............................................................................................................... 142
Study 2 .................................................................................................................... 142
Results ..................................................................................................................... 144
List of Tables

Chapter 2
Table 1  Deductions from the survey sample in the study .......................... 41
Table 2  Confirmatory Factor Analysis for Preference Scales ......................... 42
Table 3  Means (and standard deviations in brackets) for study variables .......... 43
Table 4  Correlation Matrix showing Pearson's r for Research and Teaching variables.... 44
Table 5  Probit regression for Professor dummy variable ............................... 45
Table 6  Regression for research hours per week on preferences and gender ........ 48
Table 7  Standardized Mediation Effects between gender and research hours per week..... 51
Table 8  Standardized Mediation Effects between gender and research hours per week using pattern analysis ................................................................. 52
Table 9  Probit regression for Professor dummy variable (including administration hours) 53
Table 10 Regression for research hours per week (including administration hours) .... 56
Table 11 Regression of Research Hours per week; gender as a moderator ........... 59
Chapter 3
Table 1  Experimental Validity Checks .......................................................... 100
Table 2  Summary of academics’ likelihood of response and time donations to students... 101
Table 3  Effect sizes (Cohen’s D) for academics’ likelihood of response and time donations ................................................................. 102
Table 4  Tobit Regressions of responses to Students by Academics .................. 103
Table 5  OLS Regressions of responses to Students by Academics .................. 107
Table 6  Tobit Regressions of response for Expectations-Engagement Conditions ...... 111
Table 7  Tobit Regressions of response to students by Academics with Gender……….. 115
       Interaction for guilt………………………………………………………………………………………………..

Chapter 4
Table 1  Means (and standard deviations in brackets) for Study 1 variables……………… 152
Table 2  Regressions for Hiring Recommendation, Warmth and Competence in Study 1... 153
Table 3  Standardized Mediation Effects for Warmth and Competence on the Hiring…… 154
       Recommendation in Study 1…………………………………………………………………………………………
Table 4  Unstandardized Mediation Effects for Warmth and Competence on the Hiring 155
       Recommendation in Study 1. Statistics for r=5,000 Bootstrapped regressions…
Table 5  Means (and standard deviations in brackets) for Study 2 variables……………… 156
Table 6  Regressions for Warmth, Competence and Hiring Recommendation in Study 2... 157
Table 7  Comparison of Male Academic Coefficients in Study 1 and Study 2……………. 158
Table 8  Regression Analysis of the Role of Bias Awareness on Gender Bias in Study 2... 159

List of Figures

Chapter 3
Figure 1  Likelihood of responding (%) and time devoted (in minutes) by expectation… 119
         levels………………………………………………………………………………………………………………
Figure 2  Likelihood and Minutes Devoted by expectation levels by gender……………… 120

Chapter 4
Figure 1  Study 1: Simple Slopes Analysis for the Hiring Recommendation, Warmth and 160
         Competence…………………………………………………………………………………………………………
Figure 2  Study 2: Simple Slopes Analysis for the Hiring Recommendation, Warmth and 161
         Competence…………………………………………………………………………………………………………
Chapter 1
Introduction

This thesis is a collection of three related papers examining how academics handle multiple demands in the UK higher education sector. The handling of multiple demands in academia is important to individual careers, student outcomes and institutional productivity. The UK higher education sector generates up to £73 billion of output, accounts for 2.8% of GDP and for 2.7% of all UK employment, the equivalent of 757,268 full-time jobs (Kelly, McNicoll, & White, 2014). A strategic review of the sector highlights the need to develop and retain talented academics in an increasingly demanding environment (Deloitte LLP, 2015). Academics face the challenges of rising student expectations and a focus on value for money as well as rising quality standards for both research and teaching (Department for Business Innovation and Skills, 2016).

The papers, which follow at chapters two, three and four, are each self-contained but address common central questions of why some academics progress faster through their careers than others and why male academics may enjoy a career advantage. A central conceptual theme of these papers is that academics have some degree of discretion in their time allocation decisions between research, teaching and administration/service with consequences for career outcomes. Academic choice and discretion, associated with academic autonomy in time allocated to research, professionalism and trust, were characteristics of the system of liberal governmentality prevailing in UK higher education until the 1980s. This has subsequently given way to a system of neoliberal governmentality with a marketization of public sector higher education in the UK (Canaan & Shumar, 2008; Olssen & Peters, 2005; Olssen, 2016). The effect has been to establish principal-agent line management chains leading to control through workload accountability, facilitated by Information and Communication Technology (ICT), and reduced academic autonomy (Canaan & Shumar, 2008; J. Kenny & Fluck, 2014; J. Kenny, 2017; Vardi, 2009). With
widening student participation and universities largely funded on a per student basis, students’ choice of university, as consumers and economic beneficiaries, is central to this model. Student choice will depend on both teaching quality, for employment prospects, and research-driven prestige such that combining both high quality teaching and research is desirable for universities (Canaan & Shumar, 2008; Olssen & Peters, 2005).

However, for academics, acting as agents in a principal-agent relationship, teaching, research and administration/service may be viewed as substitutes in their use of time (Gautier & Wauthy, 2007; Hattie & Marsh, 1996; Marsh & Hattie, 2002). Arguably, some residual academic autonomy has persisted and re-evolved in the 21st century to permit the substitution of activities (Fumasoli, Gornitzka, & Maassen, 2014; Nokkala & Bacevic, 2014; Zgaga, 2012). Indeed, the incentives to devote time to research may have increased due to the requirements of the Research Excellence Framework, an increasing institutional focus on gaining research funding, and the existence of an international labour market where research publications are more portable than teaching (Canaan & Shumar, 2008; Gautier & Wauthy, 2007; Murphy & Sage, 2014). More recently, the Teaching Excellence Framework has also incentivised teaching (Department for Business Innovation and Skills, 2016). The incentives facing academics, and their intrinsic motivations for each role, may influence choices between research and teaching subject to institutional constraints. Indeed academics have been shown to use the discretion that remains in their roles to produce very different occupational profiles with differences in the allocation of their working time (Allgood & Walstad, 2013; Boyd & Smith, 2016; El Ouardighi, Kogan, & Vranceanu, 2013; Gautier & Wauthy, 2007; Kossi, Lesueur, & Sabatier, 2015; Wilkin & Tavernier, 2002). This remaining element of academic choice and discretion forms the basis for chapter two, focusing on individual preferences, and chapter three, focusing on differing responsiveness to student expectations and engagement.
A focus on differential career outcomes by gender is in the context of a continuing low female representation at professor level, standing at 23% (Higher Education Statistics Agency, 2017). In contrast, postgraduate student numbers, from which academics will emerge, are weighted 56.2% to 43.8% in favour of women. Differential career outcomes have been attributed to systemic cultural sexism in a male-centered academic system, affecting hiring and evaluation, with gendered role expectations and increased gender inequalities at the intersectionality of gender and motherhood (Heijstra, Bjarnason, & Rafnsdóttir, 2015; O'Hagan, 2018; Savigny, 2014; Savigny, 2017). Certainly female academics face different societal expectations for care giving which may lead to gendered differences in time usage and power over time (Beddoes, 2015; Heijstra et al., 2015; Rafnsdóttir & Heijstra, 2013). Further, gender differences in volunteering for non-promotable administration/service tasks, as part of their time allocation, and gender differences in recognizing contributions have been studied extensively (Babcock, Recalde, Vesterlund, & Weingart, 2017; Sarsons, 2017; Vesterlund, Babcock, & Weingart, 2014). In this context, analysing gender differences in handling the multiple demands of academia is of great interest. Conceptually, whilst controlling for domestic responsibilities (number of children, young children, job mobility), I assume that both male and female academics have some power over their allocation of time and in this context the descriptive statistics in chapter two show little difference in total hours worked per week by gender. Chapters two and three focus on gender differences in time allocations and chapter four on gender differences in outcomes as a result of gender biases in teaching evaluations.

In the three papers that follow, a specific conceptual position has been taken with regards to gender. A feminist distinction has been made between gender, as an imposed social construct, and sex, as biologically determined (Harrison, 2006; Mikkola, 2008). This has led to an extensive literature on gender differences (Croson & Gneezy, 2009; Niederle, 2016). In
chapters two and three, I study gender differences but using sex as an operationalisation of
gender, given that “biological sex is an appropriate operationalization when studying gender
…. at the social structural level. In this case, research examines around issues such as others'
perceptions, attributions, attitudes, or behaviours toward a target person (e.g., stereotyping,
sexism, and tokenism). In such instances, sex is a more salient and overt cue than gender
role; therefore, perceivers' judgments and actions will more likely be influenced by the sex of
the target person (i.e., their status or their attributed gender assignment as a man or a woman)
than by the target person's gender identity or gender-role traits or attitudes.” (Korabik, 1999,
p 20). Chapter four features an experiment that manipulates perceived gender, as
operationalised by sex.

Whilst each paper tackles a separate theme, there are commonalities in the methods
and research strategies employed, with behavioural experiments used to address specific
research questions. Chapter two examines the daily challenge faced by academics in
balancing the conflicting demands of career-enhancing research with teaching and
administration/service duties. I find devoting more time to research is associated with high
levels of academic seniority. I use individual research, teaching and time preferences to
examine differences in allocating time to research. I find that preferences for research and
teaching are in conflict in allocating time to career-enhancing research. Further, I find gender
differences in research preference that may contribute to a male career advantage in
academia.

Chapter three provides another resolution of the conflict between devoting time to
either research or teaching. I examine why some academics devote more time to supporting a
student, when extra-curricular teaching effort is in conflict with a career-enhancing research
grant application. I test whether, where academics’ have discretion over teaching effort, high
levels of student expectations of support and student engagement lead to increased extra-
curricular teaching effort. I also examine gender differences in the responsiveness to expectations and engagement.

I find that high student expectations of support and student engagement give rise to extra-curricular teaching effort. The findings are mixed with regard to the mechanism for these effects. On the one hand, I control for individual differences in guilt proneness (T. Cohen, Kim, Jordan, & Panter, 2014), and show a significant correlation of guilt proneness with extra-curricular teaching effort. This provides support for guilt aversion possibly being a mechanism for in-the-moment effects of student expectations and engagement on extra-curricular teaching effort. On the other hand, I find that female academics are more prone to guilt than male academics. Yet, there are no gender differences in how female versus male academics respond to student expectations of support and student engagement. I discuss these seemingly conflicting findings and suggest promising avenues for future research.

Chapter four examines gender biases in teaching evaluations, following from findings that female academics may devote less time to research and more time to teaching in chapter two. I test whether female academics are assessed more poorly than their male peers for equal performance, by manipulating the gender of the candidate being assessed, leading to unfair decisions during academics’ careers.

In two controlled experiments, I examine whether gender bias is eliminated by an academic’s high warmth teaching style and by seniority. I find that gender bias lowers recommendations to hire female academics delivering identical content as male academics, with the effect mediated by evaluations of the academic’s warmth and/or competence. In Study 1, I test competing hypotheses regarding the effect of teaching style on gender bias. I find that a high warmth teaching style increases women’s perceived warmth, but decreases their perceived competence, so gender bias in hiring recommendations remains. In Study 2, I find that gender bias disappears for senior academics. Finally, I find no evidence of less
biased evaluations by those who anticipate gender bias. I discuss my findings in the higher education context and make recommendations to mitigate gender bias in teaching evaluations.

As the basis for chapters two and three of this thesis, over 1,400 UK academics contributed to a unique dataset of preferences and experimental responses. This represents a point of difference to many behavioural studies using student samples in that responses were gathered from the same subject group who face multitasking decisions in every day real-world settings. For chapter four, the experiments were conducted using Prolific.com, an academic website aimed at recruiting and screening candidates for experimental manipulations (Peer, Brandimarte, Samat, & Acquisti, 2017). The use of behavioural experiments in an applied setting enables one to gain a more fine grained understanding of how academics handle multiple demands and to manipulate independent variables associated with time allocation decisions.

I contribute to the existing literature in each of chapters two to four. Academic’s preferences for research and teaching have received limited attention and my work represents a significant step in exploring the link between preferences and career outcomes (Callaghan & Coldwell, 2014; Matthews, Lodge, & Bosanquet, 2014). I add to this literature by finding the mediating role of a preference for research in devoting time to research activities. I also add to the literature on the relationship between the roles of research and teaching, by finding a conflict between research and teaching preferences (Hattie & Marsh, 1996; Marsh & Hattie, 2002) and to the study of gender differences in role preferences (Gino, Wilmuth, & Brooks, 2015).

In finding that student expectations and engagement lead to extra-curricular teaching effort, my work contributes to the applications of the social exchange theory (Blau, 1964; Emerson, 1976). Guilt aversion has been characterised as an individual altering their
behaviour to avoid guilt from failing to line up to the expectations of others, in order to maintain and strengthen relationships (Baumeister, Stillwell, & Heatherton, 1994). In testing for guilt as a possible mechanism for the in-moment decision to devote extra-curricular teaching effort, I contribute to the literature on guilt aversion in an applied setting (Battigalli & Dufwenberg, 2007; Charness & Dufwenberg, 2006).

Existing field studies of gender bias in teaching evaluations, the subject of chapter four, have yet to consider the role of evaluations of warmth and competence in this setting (MacNell, Driscoll, & Hunt, 2015; Ottoboni, Boring, & Stark, 2016). Very few studies have considered the role of academic seniority in teaching evaluations (Mengel, Sauermann, & Zölitz, 2017) or the debiasing of evaluations (Boring & Arnaud, 2017). I extend these studies, and derive new insights, in a controlled experimental setting.

To summarize, the focus of this thesis is in extending behavioural approaches into the applied, and economically relevant, setting of the UK higher education sector to study how academics handle multiple demands and the differential effects on their career outcomes. chapter five concludes this thesis by summarizing the results of chapters two to four, outlining the policy implications of these results, subject to their external (ecological) validity, and pointing out promising directions for further research.
Chapter 2
“Understanding the research time allocations of academics and its association with career success.”
Abstract

Academics face a daily challenge in balancing the conflicting demands of career-enhancing research with teaching and service duties. Using a unique dataset gathered from UK academics across multiple departments, I find, as anticipated, that those who devote more time to research hold higher academic titles. I use individual research, teaching and time preferences to examine differences in allocating time to research and I find that preferences for research and teaching are in conflict. I also find that there are significant gender differences in preferences for research and teaching. Gender differences in research preference are directly associated with a greater allocation of time to research by male academics, which may contribute to a male career advantage in academia. I discuss my findings in terms of their potential to shed light on the unexplained gender difference in career outcomes in academia.

Introduction

In this paper, I examine how academics handle the multiple demands of their work, and allocate time to research as against teaching and administration/service. Academic discretion in the allocation of time has been eroded by increasing workload accountability, facilitated by Information and Communications Technology (ICT) (J. Kenny & Fluck, 2014; J. Kenny, 2017; Vardi, 2009). This has been a feature of the neoliberal model of marketised public sector higher education in the UK (Canaan & Shumar, 2008; Olssen & Peters, 2005; Olssen, 2016). Student choice, and the flow of funding from students, is central to this model. This has made both high quality teaching, to enhance student employment prospects, and research-driven prestige, desirable for universities (Canaan & Shumar, 2008; Olssen & Peters, 2005).
However, subject to workload accountability constraints, academics may view teaching, research and administration/service as substitutes with both the perceived incentives for each role and the intrinsic motivations of academics resulting in very different role profiles in practice (Allgood & Walstad, 2013; Gautier & Wauthy, 2007; Kossi et al., 2015; Wilkin & Tavernier, 2002). Certainly beyond minimum requirements for teaching, how much more time and effort spent to improve the quality and quantity of teaching activities may be discretionary (Boyd & Smith, 2016; El Ouardighi et al., 2013). This is reflected in my descriptive statistics, with gender differences in time allocated to each role despite working similar hours. Compared to female academics, male academics allocate more time to research (19.30 hours versus 16.68 hours per week) and less time to teaching (18.05 hours versus 19.28 hours). Assessing differences in the relative amounts of time academics actually, rather than contractually, spend on research, teaching and administration/service is therefore of interest, with 60% of UK full-time academics engaged in concurrent research and teaching roles (Higher Education Statistics Agency, 2018).

Whilst many socio-organisational factors can influence research productivity (Beerkens, 2013; Fox & Mohapatra, 2007; Kern, 2011), my interest in the decision to allocate time to research arises from the positive relationship between time use and research productivity (Chen, Gupta, & Hoshower, 2006; Hattie & Marsh, 1996; Stack, 2004; Tower, Plummer, & Ridgewell, 2007). The further link between research productivity and career progression in academia (Cadez, Dimovski, & Zaman Groff, 2017; Parker, 2008), means that individual time allocation decisions may affect career outcomes.

I argue that individual decisions to allocate time to research will be influenced by preferences for the activities of research and teaching, together with time preferences (Callaghan & Coldwell, 2014; Frederick, Loewenstein, & O'Donoghue, 2002; Matthews et al., 2014). A preference for research may translate directly into time spent on research. A
preference for teaching, as the other central function of higher education, may translate into
teaching time allocations which can conflict with research time allocations (Hattie & Marsh,
1996). If academics are primarily motivated by extrinsic career rewards, then time
preferences, a measure of relative preference for larger but delayed rewards over smaller
immediate rewards, may also matter to time allocation decisions, given the different timings
of the benefits of research (later) and teaching (sooner) (Frederick et al., 2002).

Notably, research, teaching and time preferences may all differ by gender and
contribute to gender differences in career decisions and outcomes. This assumes that both
male and female academics have power over the allocation of time in a male-centered
academic system which I discuss below (Heijstra et al., 2015; O’Hagan, 2018; Savigny, 2014;
Savigny, 2017). There is potential for differences in preferences to help increase the
understanding of phenomena such as the glass ceiling in UK higher education. The
proportion of female academics at professor level was only 23% in 2014/15 despite initiatives
to promote gender equality (Higher Education Statistics Agency, 2017). I argue that gendered
role and social preferences (Azmat & Petrongolo, 2014; Konrad, Ritchie Jr, Lieb, &
Corrigall, 2000), socially constructed gender differences in delay gratification (Bjorklund &
Kipp, 1996) and a greater female weight to trade-offs arising from progression (Gino et al.,
2015) may result in gendered research, teaching and time preferences that individually and
together favour greater male allocations of time to research.

As anticipated, I find that academics devoting more time to research hold higher
academic titles. A direct relationship between hours allocated to research and a senior
academic title is found across a range of UK academic departments with different gender
ratios, academic cultures and traditions. I am not able to establish causality in this
relationship as my study is based on cross-sectional data. Cross-sectional data may also be
subject to sampling bias, omitted variable bias and the potential for two way causality
(Hoover, 2006) but this association provides a context for the study of individual preferences in allocating time to research and may be a starting point for a further longitudinal study (Andrews, 2018; Sedgwick, 2014).

I examine the role of research, teaching and time preferences in the decision to allocate time to research. In light of the importance of research time to career outcomes, I find that preferences for research and teaching have conflicting effects on devoting time to research. In testing for gendered research, teaching and time preferences, I find that a male preference for research does indeed mediate a male advantage in the amount of time allocated to research. Although I do not find strong evidence for the mediating effects of teaching and time preferences, I do find a female preference for teaching and higher female rates of time preference. In combination, these three individual preferences mediate the effect of gender on time allocated to research. So, I recoded my data to identify those conditions most aligned with my theoretical prediction of greater time allocated to research using a technique devised to provide a holistic view of the alignment of characteristics associated with a positive career outcome (Foti & Hauenstein, 2007). This verified that gender differences in the allocation of time to research appear to be explained by a greater likelihood of encountering individuals with an aligned pattern of preferences, with my theoretical prediction of preferences associated with greater time allocated to research, among male academics, and a greater likelihood of encountering individuals with a misaligned pattern of preferences among female academics.

Finding a pattern of high research preference, low teaching preference and low time preference associated with more research time contributes to the study of academic preferences (Callaghan & Coldwell, 2014; Matthews et al., 2014). In finding gendered differences in research preferences underlying greater male research time, I contribute directly to work on gender differences in role preferences (Azmat & Petrongolo, 2014;
Konrad et al., 2000) and to the developing conversation about gender differences in preferences as a whole (Croson & Gneezy, 2009). My analysis also adds to the understanding of gendered differences in time preferences (Bjorklund & Kipp, 1996; Dittrich & Leipold, 2014; Silverman, 2003).

Theory Development

The Importance of Time Allocated to Research

A degree of academic autonomy has arguably persisted and re-evolved dynamically despite the marketization of higher education (Fumasoli et al., 2014; Nokkala & Bacevic, 2014; Zgaga, 2012). Further, academics may regard teaching, research and administration/service as substitutes in their use of time (Gautier & Wauthy, 2007; Hattie & Marsh, 1996; Marsh & Hattie, 2002). Whilst the marketization of higher education has restricted academic autonomy and discretion, it has arguably increased the attractiveness of devoting time to research through the requirements of the Research Excellence Framework, an increasing institutional focus on gaining research funding, and the creation of an international labour market where research publications are more portable than teaching (Canaan & Shumar, 2008; Gautier & Wauthy, 2007; Murphy & Sage, 2014). Conversely, the recent introduction of the Teaching Excellence Framework has also incentivised teaching (Department for Business Innovation and Skills, 2016). With the discretion remaining to them, academics may seek to change their allocation of time to research (Boyd & Smith, 2016; Kossi et al., 2015).

Many determinants of academic research productivity have been identified including individual passion, research funding, institutional and collaborative networks and other socio-organisational factors (Beerkens, 2013; Fox & Mohapatra, 2007; Kern, 2011). Within these frameworks of determinants, time allocated to research has been identified as having a strong,
positive effect on research productivity (Chen et al., 2006; Hattie & Marsh, 1996; Stack, 2004; Tower et al., 2007). So, all other factors being equal, an academic’s discretionary time allocation decisions, in the shape of time allocated to research, can have a direct impact on their personal research productivity.

Personal research productivity and impact are often cited as the primary drivers of academic progression (Morley, 2014; West, 2016). Certainly, the Research Excellence Framework has focused attention on both institutional and personal research productivity (Murphy & Sage, 2014). In an increasingly global higher education labour market, research outputs are recognised as a key indicator of personal quality and may be more marketable than teaching efforts (Gautier & Wauthy, 2007; Mägi & Beerkens, 2016). Moreover, recently, research productivity has been linked to better teaching (Healey, Jordan, Pell, & Short, 2010; Shin, 2011; Zhang & Shin, 2015). All in all, research productivity thus becomes increasingly linked with positive career outcomes. So, time devoted to research may be expected to relate positively to career outcomes in academia.

**Hypothesis 1: (Academic Progression)** those academics who devote more time to research are more likely to achieve higher academic titles.

There is empirical evidence that research performance may be more important than teaching to academic progression to senior academic levels (Cadez et al., 2017; Parker, 2008).

**The Role of Preferences in Time Allocated to Research**

I have argued that individual preferences, together with incentives, may create differences in the allocation of working time and I will focus on three types of individual preferences that may matter for career outcomes. Research and teaching are considered to be
the primary functions of higher education (Gautier & Wauthy, 2007; Hattie & Marsh, 1996). Additionally, time preferences may influence a range of personal outcomes (Frederick et al., 2002; Urminsky & Zauberman, 2016). I will focus on these three preferences ahead of administration/service which may be non-promotable and plays a secondary role within higher education, generating knowledge specific to an institution compared to the more portable knowledge arising from research and teaching (Department for Business Innovation and Skills, 2016; Morley, 2014; West, 2016). I will verify whether time devoted to administration/service affects my findings in a robustness analysis of my results.

A preference for research may translate directly into research time allocations (Matthews et al., 2014). That would mean that some academics enjoy doing research more than others, as a matter of taste, and thus, choose to spend more time doing what they enjoy more. A preference for teaching, as the other central function of higher education, may translate into teaching time allocations which can conflict with research time allocations (Hattie & Marsh, 1996). Finally, time preferences, a measure of relative preference for larger but delayed rewards over smaller immediate rewards may also matter to research time allocations if academics are strongly motivated by extrinsic career rewards, given the different timings of the benefits of research (later) and teaching (sooner) (Frederick et al., 2002). I will develop the arguments for how each of these three preferences will affect time allocated to research in turn.

As explained above, a preference for research will, subject to incentives and institutional constraints, result in greater time being allocated to research. Positive incentives, in the form of the research excellence framework, personal recognition and marketability, and academic cultural norms should aid the translation of the research preference into action (Fishbein, 1979; Fishbein & Ajzen, 2011). Further, studies have found a clear correlation between allocating time to research and research productivity (Chen et al., 2006; Hattie &
Marsh, 1996; Stack, 2004; Tower et al., 2007). So, allocating time to research should develop self-efficacy in research skills, reinforcing a preference for research and the desire to turn that preference into action (Bandura, 1986). Thus I would expect a stronger preference for research to result in greater time allocated to research.

**Hypothesis 2: (Research Preference)** those academics who have a stronger preference for research will allocate more hours to research.

A stronger preference for teaching will, subject to incentives, similarly result in greater time being allocated to teaching. Developing self-efficacy in teaching is also likely to reinforce a preference for teaching (Tschannen-Moran, Hoy, & Hoy, 1998). Given that teaching and research represent the two central functions of higher education they represent alternatives in allocating working time (Hattie & Marsh, 1996). The scarcity model posits that limited time, energy and commitment will lead to a conflict between research and teaching and allocating time to teaching will certainly reduce time allocated to research, all else remaining equal (Moore, 1963). Hence I would expect a preference for teaching to negatively impact time allocated to research.

**Hypothesis 3: (Teaching Preference)** those academics who have a stronger preference for teaching will allocate fewer hours to research.

If academics are motivated by extrinsic career rewards, then time preferences may impact time allocated to research given the trade-off between the costs and benefits of research and teaching activities occurring at different times. The costs, in allocating time to research and teaching, occur each academic year. The economic benefits of teaching, in the
form of module evaluations and their impact on performance appraisals, will arise within the appraisal period. However, the benefits of research may be deferred far into the future. The rigorous demands of publishing in highly ranked journals, lengthy gestation periods to publication and the cumulative nature of accruing citations all take time and career benefits will only accrue subsequently. Arguably, time preferences will be reflected in the rate at which academic benefits are discounted over time. A higher rate of time discounting will be consistent with a preference for the smaller but more immediate rewards of teaching, and a lower rate of time discounting will be consistent with a willingness to defer the larger reward of progression through research.

**Hypothesis 4: (Time Preference)** those academics with a lower rate of time discounting will allocate more hours to research.

**Gender Differences in Research and Teaching Preferences.**

It is notable that all three preferences discussed above may be gender-specific. If this is the case, gendered preferences may be an important lens for understanding potential gender differences in research time allocations which, arguably, produce well-defined gender differences in the career outcomes of academics. In this context, lower research productivity has long been associated with lesser career outcomes for female academics (Larivière, Vignola-Gagné, Villeneuve, Gélinas, & Gingras, 2011; Leahey, 2006; Shauman & Xie, 2003).

A key argument in studying gender differences in preferences is that both male and female academics have power over their allocation of time (Fumasoli et al., 2014; Nokkala & Bacevic, 2014; Zgaga, 2012). Systemic cultural sexism in a male-centered academic system may produce gendered role expectations with increased gender inequalities at the
intersectionality of gender and motherhood limiting discretion (Heijstra et al., 2015; O'Hagan, 2018; Savigny, 2014; Savigny, 2017). Female academics face different societal expectations for care-giving which may lead to gendered differences in time usage and power over time (Beddoes, 2015; Heijstra et al., 2015; Rafnsdóttir & Heijstra, 2013). Further, gender differences in volunteering for non-promotable administration/service tasks may curtail the effect of research and teaching preferences on female academics’ time allocations (Babcock et al., 2017; Vesterlund et al., 2014). I develop these and other broader qualifications in the Sociological Considerations section of the General Discussion. Subject to these considerations, I will build on hypotheses 2 to 4, to formulate hypotheses for the mediating role of gender-specific research, teaching and time preferences on time allocated to research in turn.

I anticipate a greater male preference for research, and a greater female preference for teaching. Publishing research in high ranking journals is a competitive process (Brembs, Button, & Munafò, 2013; Fanelli, 2010; Fanelli, 2012; Van Dalen & Henkens, 2012). Studies have found that men are more willing to engage in competitive interactions and self-select into competitive environments (Gneezy, Niederle, & Rustichini, 2003; Niederle & Vesterlund, 2007). Male role preferences for challenge, autonomy and influence, in line with gender stereotypes, may be an additional contributory factor to a male preference for research (Konrad et al., 2000). Further, neoliberalised academia has been associated with individualism and an increasingly international job market in higher education (Canaan & Shumar, 2008). Male academics, constrained by fewer domestic responsibilities relative to female academics, may be better placed to leverage the career advantages afforded by greater research productivity (Ivancheva, Lynch, & Keating, 2019; Loveday, 2018; Rafnsdóttir & Heijstra, 2013), also shaping a male preference for research.
At the same time, fewer constraints on male academics meeting the demands of increasing workload accountability may reinforce traditional academic stereotypes of the male researcher and the female teacher (MacNell et al., 2015; Powell, Ah-King, & Hussénius, 2018). Gendered preferences, again fashioned at the social structural level, may also encourage a female preference for teaching and care-giving in higher education (Ashencaen Crabtree & Shiel, 2018). A range of studies suggest greater female other-regarding preferences, and more socially oriented role preferences (Azmat & Petrongolo, 2014; Croson & Gneezy, 2009; Konrad et al., 2000; Tonin & Vlassopoulos, 2010). A female preference for affiliation and social interaction may work against a more autonomous research role in favour of a preference for a more socially oriented teaching role (Barbezat, 2006; Callaghan & Coldwell, 2014; Diener & Fujita, 1995). Whilst it can be argued that collaborative research roles can provide social interaction, at least part of the time is engaged on solitary activities such as data analysis and writing. Conversely, teaching involves both one-to-many and one-to-one experiences, providing oneself with more episodes of social interaction. The preferences discussed here may be reinforced by the findings that “women have a higher number of life goals, ..., associate more negative outcomes with high-power positions, perceive power as less desirable though equally attainable, and are less likely to take advantage of opportunities for professional advancement” (Gino et al., 2015, p. 12358). So, female academics may recognise the importance of research productivity to progression but place a lower value on academic progression, and thus on research as a means to progression, compared to male academics.

Given scarce personal resources, I expect a conflict between allocating time to research versus teaching, all else remaining equal. Hence I would expect a male preference for research to both positively impact time allocated to research and negatively impact time
allocated to teaching and a female preference for teaching to both positively impact time
allocated to teaching and negatively impact time allocated to research. To summarize,

**Hypothesis 5:** *(Mediation by (greater male) preference for research)* the effect of
gender on hours allocated to research will be positively mediated by
gender differences in the preference for research. In particular, the
positive effect of male gender on time allocated to research will be
mediated by a greater male preference for research.

**Hypothesis 6:** *(Mediation by a (greater female) preference for teaching under a
research and teaching conflict)*: the effect of gender on hours
allocated to research will be negatively mediated by gender differences
in the preference for teaching. In particular, the positive effect of male
gender on time allocated to research will be mediated by a lower male
preference for teaching.

There is some empirical evidence for less time being devoted to research and more to
teaching and service by female academics (Bellas, 1999; Toutkoushian & Bellas, 1999).

Without changing total working hours, time devoted to research, teaching and service
are conflicting alternatives. Indeed, this view of activities as alternatives is consistent with
evidence that many academics have stronger preferences for either research or teaching such
that these preferences are likely to be in conflict (Callaghan & Coldwell, 2014). However, it
is important to note that the view of research as an activity conflicting with teaching, as taken
above, is contested by some researchers. Theoretical models have been built that assume the
relationship between the two activities to be conflicting, complementary or for the two to be
unrelated academic activities (Hattie & Marsh, 1996; Marsh & Hattie, 2002). Thus, I would
have to test empirically how preferences for research and teaching/service relate to each other. To date, a meta-analysis of empirical studies could not find evidence to support research and teaching being complementary activities (Hattie & Marsh, 1996).

**Gender Differences in Time Preferences**

I anticipate higher male time discounting rates. A greater female ability to delay gratification has been theorised based on the social pressures applied to women in child-rearing (Bjorklund & Kipp, 1996). Greater delay gratification is consistent with lower female time discounting rates. Conversely, lower male delay gratification will lead to higher male time discounting rates which may lead to lower research time allocations. To summarize:

*Hypothesis 7: (Mediation by (lower female) rate of time preference) the effect of gender on hours allocated to research will be mediated by gender differences in time preferences. In particular, the negative effect of male gender on time allocated to research will be mediated by a higher male rate of time discounting.*

Empirical evidence for gender differences in delay gratification and time preferences is mixed. In the psychology literature, a greater female ability to delay gratification has been found (Bjorklund & Kipp, 1996; Silverman, 2003). The related field of economic time preferences has also provided some evidence for a greater female ability to defer gratification, consistent with lower female time discounting rates (Bauer & Chytilová, 2013; Dittrich & Leipold, 2014). Conversely, a series of studies have found higher, not lower, female time discounting rates (Beck & Triplett, 2009; Martorano, Handa, Halpern, Pettifor, & Thirumurthy, 2015; Reynolds, Ortengren, Richards, & de Wit, 2006).
All in all, I am unable to hypothesise the overall effect of gender on research hours allocated. I expect a positive effect of male gender on time allocated to research from hypotheses 5 and 6, but a negative effect from hypothesis 7, such that the combined effect could be either positive or negative.

**Study Design**

**Participants and Design:**

The data was gathered as part of a survey sent to 12,272 academics at UK Universities in July 2016, with 1,418 respondents (11.6%). The integrity of responses was ensured by sending personalized email links to the Qualtrics based survey. Respondents starting but dropping out of the survey, and the restriction of the dataset to full-time staff only, reduced the number of respondents completing the detailed information required to 1,102 (Mage = 45.00, SDage = 9.59, 36.8% female) as set out at Table 1. Restricting respondents to full-time staff only was to avoid gendered preferences for research or teaching being shaped by a respondent’s part-time employment status and domestic responsibilities (Fagan, 2001). The mean duration for completed surveys was 21 minutes 57 seconds.

The number of academics invited to participate was determined by the requirements of an experiment embedded within the survey, which was used in chapter three of this thesis. Only academics with titles implying both research and teaching duties were invited to participate. The survey was sent to academics in 15 common departments at 27 pre-1992 universities with both a research and teaching focus.

I checked for selection bias by benchmarking responses (Montaquila & Olson, 2012; Pedersen, 2015). Female academics made up 38.5% of responses to the survey compared to a
benchmark for UK higher education of 37.3% (Higher Education Statistics Agency, 2017). Respondents at professor level at 27.5% exceeded a benchmark of 11.9%, but mirror over-representation at 25.8% in a comparable study (Blackaby, Booth, & Frank, 2005).

Materials:

Full details of the invitation e-mail and all materials and measures within the survey follow at Appendix A.

Pilot Study: I conducted a pilot study with 45 respondents (Mage = 42.80, SDage = 10.45, 44.4% female) at the University of Surrey to test likely response rates, which were in line with similar surveys (Abreu, Grinevich, Hughes, & Kitson, 2009). The format of the invitation email was subsequently amended based on feedback.

Procedure:

Survey invitations were sent out in two batches on Wednesdays and Fridays at 15:00 to maximise responses (Van Dessel, 2015). Non-responders were tracked and up to two reminder emails were sent, as appropriate, one week and two weeks respectively after the original invitation.

Measures:

Research Hours per Week. Participants were asked to provide a measure of their actual weekly hours worked (as opposed to contractual hours). They were also asked to provide their average percentage split between research, teaching and administration/service. Participant’s average research hours per week were then calculated by applying the percentage of time engaged in research to their average weekly hours.
**Research Preferences.** Participants were asked five questions about their preferences for research (The questions were “In order to be successful in academia, I need to focus on my research.”, “My first priority is research.”, “I am active (in the scholarship of) research.”, “I have recently participated in professional development in research.”, “I enjoy participating in research”). Four of these questions were drawn from a previous study questionnaire which did not report a Cronbach’s $\alpha$ (Matthews et al., 2014). The responses were on 5-point Likert-type scales anchored by 1 (strongly disagree) to 5 (strongly agree). The responses were averaged together to form a single composite score, where higher scores indicated a greater preference for research (Cronbach’s $\alpha = .66$). Given this relatively low value of Cronbach’s $\alpha$, I performed factor analysis to check whether the questions loaded on multiple factors representing separate concepts (see Table 2) (Tavakol & Dennick, 2011). The two largest positive factors were 1.69202 and 0.05782 respectively, so under the Kaiser criterion the model is unidimensional loading on a single factor (Kaiser, 1960).

**Teaching Preferences.** Participants were asked five questions about their preferences for teaching (The questions were “In order to be successful in academia, I need to focus on my teaching.”, “My first priority is teaching.”, “I am active (in the scholarship of) teaching.”, “I have recently participated in professional development in teaching.”, “I enjoy participating in teaching”). Four of these questions were drawn from a previous study questionnaire which did not report a Cronbach’s $\alpha$ (Matthews et al., 2014). The responses were on 5-point Likert-type scales anchored by 1 (strongly disagree) to 5 (strongly agree). The responses were averaged together to form a single composite score, where higher scores indicated a greater preference for teaching (Cronbach’s $\alpha = .56$). Given this low value for Cronbach’s $\alpha$, I again performed factor analysis (see Table 2). The two largest positive factors were 1.00293 and 0.08081 respectively, so under the Kaiser criterion the model is also unidimensional loading on a single factor (Kaiser, 1960).
**Time Preferences.** Participants were presented with choice pairs of either a declining amount paid in two days’ time or £80 at a later date. They completed seven payment choices in each of four tables derived from a previous study of inter-temporal discount rates (Burks, Carpenter, Götte, & Rustichini, 2012). Choice patterns implying negative or inconsistent interest rates were excluded from the analysis. A value for time preference was derived for each participant based on a regression over their choices. A value of 1 indicated perfect future focus with the larger future amount of £80 being chosen every time. Descending values from 1 indicated progressively increasing present focus with more choices of smaller financial amounts in two days’ time.

Control variables. I controlled for gender and also for experience, in the form of years since completion of a PhD. A control for research quality was based on a weighted index of published articles in the previous three years. A control for teaching quality was based on the achievement of teaching evaluation scores in excess of 4 out of 5 in the last academic year. Controls for domestic responsibilities measured by number of children, either one child under 5 years old or two children under 5 years old and mobility were also included.

Incentives and checks

Lotteries were used to incentivise the completion of the survey (Göritz, 2006). A prize of £300 was awarded to the respondent most accurately predicting the behavioural norms of other academics towards teaching duties. This was calculated as the response that minimises the sum of absolute deviations from the mean of each response within each academic discipline. A prize of up to £80 was awarded randomly, with a 1 in 300 chance of winning, as part of the end-of-survey questionnaire to elicit the time preferences of responders. In
addition to incentivising the completion of the survey, these specific awards were designed to make the questions they referred to more salient.

To test for selection bias arising from the use of incentives, I used two formats of survey invitation email. In one format I described the incentives whilst in the other format, I omitted this information. There were no significant differences in the characteristics of respondents between the two formats.

**Results**

Descriptive statistics for study variables are shown in Table 3. Although male and female academics report similar working hours (50.99 hours and 50.26 hours per week respectively), their allocation of time between roles differs. Overall, male academic respondents allocate 19.30 hours per week on average to research, female academics 16.68 hours per week. The effect size of the gender difference, measured by Cohen’s D is .27. Male academics allocate 18.05 hours per week on average to teaching, female academics 19.28 hours per week. The effect size of the gender difference, measured by Cohen’s D is -.16.

Within this dataset there is little gender difference in the burden of administration/service. Male academic respondents allocate 13.64 hours per week on average and female academics 14.30 hours per week. The effect size of the gender difference, measured by Cohen’s D is -.08. At academic levels below professor, the gender difference is very small (Male respondents 13.10 hours per week, female respondents 12.94 hours per week and Cohen’s D is .02). At professor level, the gender difference is large (Male respondents 15.19 hours per week, female respondents 19.12 hours per week and Cohen’s D is -.40).

Insert Table 3 about here

------------------------------------------------------------------
Correlations between research and teaching variables are shown in Table 4. Hours allocated to research are positively correlated with a preference for research and negatively correlated with a preference for teaching. The relationship between research and teaching variables is always negative with a preference for research negatively correlated with a preference for teaching.

To test Hypothesis 1 (Academic Progression), I conducted a probit regression with Professor as the dependent variable, shown in Table 5. Consistent with my hypothesis I found that hours allocated to research per week are correlated with the academic title of professor ($\beta = .02, p < .01$). However, given this result is obtained with cross-sectional data this may be better described as an association between variables with no implications for causality, as discussed later.

To test hypotheses 2 to 4 (Preferences), I regressed hours allocated to research on research, teaching and time preferences, as shown in Table 6. Consistent with hypothesis 2, a research preference is positively correlated with hours allocated to research ($\beta = 5.08, p < .01$). Consistent with hypothesis 3, a teaching preference is negatively correlated with hours allocated to research ($\beta = -3.07, p < .01$). However, the rate of time discounting is not correlated with hours allocated to research, so I do not find support for hypothesis 4 ($\beta = -175.93, ns$).
To test hypotheses 5 to 7 (mediation), I analysed whether research preferences, teaching preferences and time preferences acted as mediators between the gender of the academic and the time allocation to research, as shown at Table 7 (D. Kenny, 2016). For the individual regressions, there is a significant gender difference for research preferences ($\beta = - .08, p < .05$) but only weakly significant gender differences for teaching preferences ($\beta = .08, p < .10$) and for time preferences ($\beta = .00, p < .10$).

For hypothesis 5 (Mediation by (greater male) preference for research) the indirect effect of a preference for research was correctly signed and significant (standardized path coefficient = -.02, $p < .05$). The indirect effect of gender mediated by a preference for research on time allocated to research remained significant when bootstrapping standard errors to allow for kurtosis (Preacher & Hayes, 2008). Thus I have support for hypothesis 5.

For hypothesis 6 (Mediation by a (greater female) preference for teaching under a research and teaching conflict) the indirect effect of a preference for teaching was correctly signed but just outside of significance at $p = .10$ (standardized path coefficient = -.01, $ns$). For hypothesis 7 (Mediation by rate of time preference) the indirect effect of time preferences was not significant (standardized path coefficient = .00, $ns$). Thus I do not find support for either hypothesis 6 or hypothesis 7.

However, the overall (negative) indirect effect of the three mediators; research preferences, teaching preferences and time preferences; is significant (standardized path coefficient = -.03, $p < .05$). The direct effect of gender on time allocated to research remains significant indicating only partial mediation (standardized path coefficient = -.09, $p < .01$).

Insert Table 7 about here

--------------------------------------------
To further analyse the combined indirect effect of the three mediators; research preferences, teaching preferences and time preferences, I recoded the data. The recoding sought to identify those conditions most aligned with my theoretical prediction of greater time allocated to research using a technique devised to provide a holistic view of the alignment of characteristics associated with a positive career outcome (Foti & Hauenstein, 2007). I grouped individuals by whether they scored higher than average research preferences, lower than average teaching preferences and lower than average time discount rates as a pattern most aligned, theoretically, with greater time allocated to research. I also grouped them according to a misaligned pattern (being the opposite pattern) and partially aligned pattern (i.e. some preferences aligned only). As shown in Table 8, there is an overall (negative) indirect effect of the aligned pattern together with the misaligned pattern (standardized path coefficient = -.02, p < .05). This verified that gender differences in the allocation of time to research appear to be explained by a greater likelihood of encountering individuals with an aligned pattern of preferences, with my theoretical prediction of preferences associated with greater time allocated to research, among male academics, and a greater likelihood of encountering individuals with a misaligned pattern of preferences among female academics.

-------------------------------
Insert Table 8 about here
-------------------------------

Robustness Analysis

I conducted a number of tests of key assumptions and findings. Firstly, I verified whether the allocation of time to administration/service represents an omitted variable in the regressions of time allocated to research (Clarke, 2005; Wooldridge, 2009). Including administration/service hours in the probit regression for the professor dummy variable does not change the sign or the statistical significance for any of the measures, as shown in Table
9. Including administration hours in the regression for research hours per week on preferences and gender does not change the sign or the statistical significance for any of the measures, as shown in Table 10.

Secondly, I tested an alternative specification of the model, treating gender as a moderator, to determine whether it simplified and clarified the proposed conceptual framework. The interaction between female academics, research preferences and teaching preferences does not show significant effects, as shown in Table 11. Hence this formulation does not offer any additional insights as preferences do not lead to different time allocations for male and female academics.

**General Discussion**

With a unique dataset, based on a survey of UK academics actual time allocations, I determined the role of individual preferences in allocating time to research and the association of research time with seniority. As hypothesised, I found that a preference for research was positively correlated with, and a preference for teaching negatively correlated with, time allocated to career-enhancing research activities. Also as hypothesised, I found that gender differences in the preference for research produce gender differences in time allocated to research. These new findings may help explain the career advantage enjoyed by male academics.
I highlight the theoretical and practical implications of my results below, together with sociological considerations, the limitations of, and future directions for my research.

**Theoretical Implications**

I found that a preference for research was positively correlated with, and a preference for teaching negatively correlated with, time allocated to research. These findings provide a better understanding of the motivations of academics and contribute directly to the literature on teaching and research preferences in higher education (Callaghan & Coldwell, 2014; Matthews et al., 2014).

Theoretical models have been built that assume the relationship between the roles of research and teaching to be conflicting, complementary or for the two to be unrelated academic activities. To date, a meta-analysis could not find evidence to support research and teaching being complementary activities (Hattie & Marsh, 1996). Indeed, in my sample, a preference for teaching is negatively correlated with hours allocated to research. However, recently, research productivity has been linked to better teaching (Healey et al., 2010; Shin, 2011; Zhang & Shin, 2015). Hence, if there is a complementarity between the two roles, those academics who prefer teaching over research are likely to devote more hours to teaching and fewer hours to research so that their teaching may be under-informed by an active research agenda. In the light of the ongoing debate about the relationship between research and teaching, this is an area that merits further analysis.

Developing my analysis, I used gender differences in research, teaching and time preferences as a lens to examine gender differences in career outcomes. I found a male preference for research, a female preference for teaching and, contrary to my hypothesis, higher female rates of time discounting. I found that a strong male preference for research mediates time allocated to research, providing one explanation for why men allocate more time to research. Given studies linking time spent on research to research productivity, and
the importance of research productivity to career outcomes, this finding is important and adds to the literature on the impact of gender differences in preferences (Croson & Gneezy, 2009; Gino et al., 2015). Having argued that rather than differ in the extent that either research or teaching is preferred, female academics have to respond more or less strongly to a given level of preference for research or teaching, I tested an alternative specification of the model in the robustness analysis. Treating gender as a moderator did not show significant effects.

I also find that the combined effect of gender differences in research, teaching and time preferences mediates time allocated to research. This gender difference was confirmed by grouping individuals according to their alignment or misalignment with the pattern of higher than average research preferences but lower than average teaching and time preferences associated with allocating most hours to research. A particular pattern of preferences in combination across roles may give rise to the additional time allocation to research (Foti & Hauenstein, 2007) and this finding contributes to the study of academic preferences. My findings contribute directly to work on gender differences in role preferences (Azmat & Petrongolo, 2014; Konrad et al., 2000).

My findings, in a higher education context, also add to the understanding of gender differences in time preferences. I hypothesised higher time discounting rates for men but found higher rates of time discounting for women. However, my result is consistent with other studies based on hypothetical money choices at different times (Beck & Triplett, 2009; Martorano et al., 2015) and a laboratory study of impulsive behaviour (Reynolds et al., 2006). The study of time preferences is complicated by the likelihood that time discounting rates are domain specific, such that preferences expressed in my study for hypothetical money choices may differ from preferences relevant to time allocated to research and career outcomes (Ubfal, 2016; Urminsly & Zauberman, 2016; Winer, 1997).
I also contribute to studies of progression in higher education, by finding that holding a senior academic title is more likely for those devoting more time to research. This association was found across a range of academic departments with different gender ratios, academic cultures and traditions. The links between allocating time to research and research productivity (Chen et al., 2006; Hattie & Marsh, 1996; Stack, 2004; Tower et al., 2007) and research productivity and career outcomes (Cadez et al., 2017; Parker, 2008) are established in existing literature. My finding is a first step in directly linking time allocated to research to career outcomes. I discuss below the limitations of my approach and future directions in research to provide insight into causal relationships.

**Practical Implications**

The practical implications of my findings depend on how individual preferences, and gender differences in preferences, are formed. Preferences are based on direct experiences with more abstract preferences developed and refined with age from interactions between individuals and their environment (Druckman & Lupia, 2000; Sameroff, 2010). The polarization of preferences that I found for either research or teaching roles will also reflect the culture and norms of higher education institutions in the UK and the incentives provided for each role. Academics may be encouraged by institutions to consider research and teaching as conflicting activities as teaching becomes more focused on employability concerns in the neoliberal academy (Canaan & Shumar, 2008). The introduction of the Teaching Excellence Framework (Bhardwa, 2017; Wild & Berger, 2016) and findings for research-driven teaching (Healey et al., 2010; Shin, 2011; Zhang & Shin, 2015), changing the incentives for each role, may modify preferences going forwards.

Gender differences in preferences have been widely studied (Croson & Gneezy, 2009). Social experiences, education and parental gender-role attitudes are likely to affect the early development of gender-role attitudes leading to gender differences in preferences.
Much research confirms the gender stereotypes of women being more communal and men more competent (Biernat & Fuegen, 2001; Diekman & Eagly, 2000; Eagly & Johannesen-Schmidt, 2001; Lockheed, Harris, & Nemceff, 1983) and there is evidence of gender differences in role preferences broadly following these stereotypes (Corrigall & Konrad, 2007; Konrad et al., 2000).

Gender differences in role preferences may be reinforced by the finding that women have more and wider life goals, and place more weight on the negative effects of time constraints and trade-offs, than men (Gino et al., 2015). Female academics may prefer more flexible, less research-intensive departments to accommodate the demands of childbirth and family (Rothausen-Vange, Marler, & Wright, 2005). However, changing societal gender stereotypes and institutional cultures will influence the preferences of future female academics (Carli, 2001). This may be crucial in equalising career outcomes derived from research in higher education (Eagly & Carli, 2007).

**Sociological Considerations**

My focus is confined to gender differences in research, teaching and time preferences as shaping the time allocated to research and gender differences in academic outcomes. Gender differences in preferences may be shaped, in turn, by social expectations. However whether those preferences are reflected in time allocations will depend on whether both male and female academics have power over their allocation of time. Female academics, especially when combining a career and motherhood, face different societal expectations for care-giving which may lead to gendered differences in time usage and power over time (Amsler & Motta, 2017; Beddoes, 2015; Heijstra et al., 2015; Rafnsdóttir & Heijstra, 2013). Further, gender differences in volunteering for non-promotable administration/service tasks, may reduce the effect of research and teaching preferences on female academics’ time allocations, hence I
control for administration/service in the robustness analysis (Babcock et al., 2017; Vesterlund et al., 2014).

In my study, the mediating effect of gender differences in preferences on time allocated to research is relatively small (the direct effect of gender is three times as large in Table 6). This may reflect both time usage and wider sociological constraints on female academics in allocating time to research. Whilst I do not find strongly significant effects controlling for domestic responsibilities (number of children, young children), my sample was restricted to full-time academics with both research and teaching responsibilities who may have fewer domestic constraints. Male academics are likely to be more flexible with fewer constraints in allocating time to research (Ivancheva et al., 2019; Rafnsdóttir & Heijstra, 2013). Further, cultural sexism in a male-centered academic system may lead to gendered expectations of the roles performed and how performance in these roles is evaluated and supported (Heijstra et al., 2015; Savigny, 2014; Savigny, 2017).

**Limitations and Future Directions**

In terms of methodological limitations, my study, like other studies of research and teaching preferences, is based on cross-sectional data (Callaghan & Coldwell, 2014; Matthews et al., 2014). From my cross-sectional data, I am not able to establish causality in relationships, only associations between variables, although in each case the hypothesised relationships are supported by theory. The issue is most pertinent in associating time allocated to research with seniority, where it is difficult to rule out rival hypotheses. Further cross-sectional data issues include omitted variable bias, recall bias and the potential for two way causality also apply both to this relationship and to the study of preferences for research and teaching where additional time devoted to either research or teaching may build self-efficacy which reinforce preferences (Andrews, 2018; Sedgwick, 2014). The direction of causality in a mediation model cannot be assessed by statistical methods (D. Kenny, 2016).
outline the steps taken to ensure the robustness of my survey data in more detail in chapter three and make proposals for further analysis using a longitudinal study below.

In conducting my survey, I collected a large sample with 1,102 viable responses. Unlike a comparable survey based cross-sectional study, I tracked responses and sent multiple reminders to maximise uptake (Blackaby et al., 2005). In line with best practice, I conducted benchmarking to ensure that a representative sample, in terms of gender and seniority, was obtained (Montaquila & Olson, 2012; Pedersen, 2015). I also compared signs and coefficients of control variables, where comparable, with similar studies. My study benefitted from a wealth of additional data with which I conducted a large number of internal consistency checks as detailed in footnote 4 of chapter three.

In addressing reverse causality between time allocated to research and seniority, two considerations may support the direction of the relationship. Firstly, I would expect senior academics to have some flexibility to reduce both research and teaching effort. Hence, the risk of examining cross-sectional data would be in finding a negative relationship where a positive one would hold longitudinally. Instead, I see a positive relationship that is unlikely to be the artefact of analysing cross-sectional data. Secondly, I find that time discount rates are negatively correlated with seniority. There is both theoretical and empirical support for patience being greatest in middle age and declining into old age, such that I might expect to see higher discount rates for older, more senior academics (Martorano et al., 2015; Read & Read, 2004; Sozou & Seymour, 2003). Here, the risk of examining cross-sectional data would be in finding a positive relationship, and mistaking its importance, where a negative relationship would hold longitudinally. Instead, I see a negative relationship that is again unlikely to be the artefact of analysing cross-sectional data. I also find that time allocated to teaching affects the likelihood of holding a senior academic title negatively but this finding
could be the artefact of cross-sectional data, with senior academics being shielded from teaching duties.

My findings for time preferences are based on the argument that rates of time preference may act as a mediator for time allocated to research, given the deferred benefits of research productivity to academic careers. However, it can be questioned whether academics consider the extrinsic rewards from research outputs as a primary motivation to engage in research. Some activities have been argued to provide their own intrinsic reward (Deci, 1971; Kahneman & Thaler, 2006). So, the intrinsic reward from performing research may undermine the relationship which I argue may exist between time preferences and research hours. Indeed, I did not find a significant relationship between these variables. However, this caveat does not affect my finding of higher rates of time discounting for female academics.

Preferences for administration/service were not captured as part of the research design but may yield interesting results as a subject for further research. From the robustness analysis, including hours allocated to administration/service in regressions did not affect either the sign or significance of any of the key variables. So, the findings in respect of research, teaching and time preferences on time allocated to research appear unaffected. In contrast to literature on gender differences in accepting administration/service tasks with low promotability, I do not find gender differences for time allocated to service below professor level (Babcock et al., 2017; Pyke, 2011; Vesterlund et al., 2014). However, I do find a large gender effect at professor level, consistent with previous research on the disproportionate burden of administration and service activities falling on underrepresented senior female academics (Grove, 2016; Guarino & Borden, 2017).

My findings, especially those relating to individual research and teaching preferences, represent a springboard for further research. Conducting a longitudinal study would enable researchers to track time allocation to research and individual research, teaching and time
preferences over a period of time. This would provide some insight into the stability of individual preferences over time and whether these preferences change with age, experience and academic progression. It would also enable the testing of proposed cause and effect relationships over time compared to merely identifying associations within a snapshot of cross-sectional data. Associating individual preferences, and gender differences in those preferences with career outcomes, merits further investigation using such advanced data-analytic methods.
## Tables and Figures

**TABLE 1**  
*Deductions from the survey sample in the study*

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Original survey sample</td>
<td>1,418</td>
<td>100.00%</td>
</tr>
<tr>
<td>Part time staff excluded</td>
<td>75</td>
<td>1,343</td>
</tr>
<tr>
<td></td>
<td></td>
<td>94.71%</td>
</tr>
<tr>
<td>Other deductions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inconsistent time prefs</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Incomplete data</td>
<td>226</td>
<td></td>
</tr>
<tr>
<td>Net Sample</td>
<td>1,102</td>
<td>77.72%</td>
</tr>
</tbody>
</table>
TABLE 2
Confirmatory Factor Analysis for Preference Scales

(a) Factor Analysis for Research Preference scale

<table>
<thead>
<tr>
<th>Factor</th>
<th>Eigenvalue</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td>1.69202</td>
<td>1.2531</td>
</tr>
<tr>
<td>Factor 2</td>
<td>0.05782</td>
<td>0.0428</td>
</tr>
<tr>
<td>Factor 3</td>
<td>-0.05130</td>
<td>-0.0380</td>
</tr>
<tr>
<td>Factor 4</td>
<td>-0.14094</td>
<td>-0.1044</td>
</tr>
<tr>
<td>Factor 5</td>
<td>-0.20734</td>
<td>-0.1536</td>
</tr>
</tbody>
</table>


(b) Factor Loadings for Research Preference scale

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>In order to succeed in academia, I need to focus on my teaching.</td>
<td>.57</td>
<td>-.07</td>
</tr>
<tr>
<td>My first priority is teaching.</td>
<td>.60</td>
<td>.02</td>
</tr>
<tr>
<td>I am active in (the scholarship of) teaching.</td>
<td>.72</td>
<td>.12</td>
</tr>
<tr>
<td>I have recently participated in professional development in teaching.</td>
<td>.28</td>
<td>.22</td>
</tr>
<tr>
<td>I enjoy participating in teaching.</td>
<td>.62</td>
<td>.03</td>
</tr>
</tbody>
</table>

Notes. Method: principal factors, rotation method: orthogonal varimax (Kaiser off). Loadings larger than .40 are in bold.

(c) Factor Analysis for Teaching Preference scale

<table>
<thead>
<tr>
<th>Factor</th>
<th>Eigenvalue</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td>1.00293</td>
<td>1.5019</td>
</tr>
<tr>
<td>Factor 2</td>
<td>0.08081</td>
<td>0.1210</td>
</tr>
<tr>
<td>Factor 3</td>
<td>-0.03994</td>
<td>-0.0598</td>
</tr>
<tr>
<td>Factor 4</td>
<td>-0.16059</td>
<td>-0.2405</td>
</tr>
<tr>
<td>Factor 5</td>
<td>-0.21543</td>
<td>-0.3226</td>
</tr>
</tbody>
</table>


(d) Factor Loadings for Teaching Preference scale

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>In order to succeed in academia, I need to focus on my teaching.</td>
<td>.26</td>
<td>.16</td>
</tr>
<tr>
<td>My first priority is teaching.</td>
<td>.35</td>
<td>.36</td>
</tr>
<tr>
<td>I am active in (the scholarship of) teaching.</td>
<td>.53</td>
<td>.22</td>
</tr>
<tr>
<td>I have recently participated in professional development in teaching.</td>
<td>.49</td>
<td>.11</td>
</tr>
<tr>
<td>I enjoy participating in teaching.</td>
<td>.22</td>
<td>.34</td>
</tr>
</tbody>
</table>

Notes. Method: principal factors, rotation method: orthogonal varimax (Kaiser off). Loadings larger than .40 are in bold.
TABLE 3
*Means (and standard deviations in brackets) for study variables*

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Professors</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Total Hours per week</td>
<td>50.26</td>
<td>50.99</td>
<td>53.63</td>
</tr>
<tr>
<td></td>
<td>(8.79)</td>
<td>(9.50)</td>
<td>(9.30)</td>
</tr>
<tr>
<td>Research Hours per week</td>
<td>16.68</td>
<td>19.30</td>
<td>18.64</td>
</tr>
<tr>
<td></td>
<td>(8.54)</td>
<td>(10.29)</td>
<td>(9.57)</td>
</tr>
<tr>
<td>Teaching Hours per week</td>
<td>19.28</td>
<td>18.05</td>
<td>15.87</td>
</tr>
<tr>
<td></td>
<td>(8.16)</td>
<td>(7.73)</td>
<td>(9.50)</td>
</tr>
<tr>
<td>Admin/Service Hours per week</td>
<td>14.30</td>
<td>13.64</td>
<td>19.12</td>
</tr>
<tr>
<td></td>
<td>(8.93)</td>
<td>(8.53)</td>
<td>(10.37)</td>
</tr>
<tr>
<td>Research Preference Index</td>
<td>3.04</td>
<td>3.07</td>
<td>3.07</td>
</tr>
<tr>
<td></td>
<td>(0.71)</td>
<td>(0.61)</td>
<td>(0.58)</td>
</tr>
<tr>
<td>Teaching Preference Index</td>
<td>2.29</td>
<td>2.19</td>
<td>2.06</td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
<td>(0.67)</td>
<td>(0.73)</td>
</tr>
<tr>
<td>High Research Quality (Note 1)</td>
<td>0.50</td>
<td>0.55</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(0.50)</td>
<td>(0.47)</td>
</tr>
<tr>
<td>High Teaching Quality (Note 2)</td>
<td>0.39</td>
<td>0.38</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>(0.49)</td>
<td>(0.49)</td>
<td>(0.47)</td>
</tr>
<tr>
<td>Years Post PHD</td>
<td>13.12</td>
<td>17.80</td>
<td>22.39</td>
</tr>
<tr>
<td></td>
<td>(8.73)</td>
<td>(10.59)</td>
<td>(7.67)</td>
</tr>
<tr>
<td>Observations</td>
<td>498</td>
<td>828</td>
<td>99</td>
</tr>
</tbody>
</table>

Standard deviations in brackets. Note 1 – a control for research quality was based on a weighted index of published articles in the past 3 years. Note 2 – a control for teaching quality was based on the achievement of teaching evaluation scores in excess of 4 out of 5 in the last academic year.
<table>
<thead>
<tr>
<th></th>
<th>Research Hours per Week</th>
<th>Research Preference Index</th>
<th>High Research Quality</th>
<th>Teaching Hours per week</th>
<th>Teaching Preference Index</th>
<th>High Teaching Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Hours per Week</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Preference Index</td>
<td>0.38</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Research Quality</td>
<td>0.15</td>
<td>0.05</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teaching Hours per week</td>
<td>-0.36</td>
<td>-0.18</td>
<td>-0.07</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teaching Preference Index</td>
<td>-0.31</td>
<td>-0.13</td>
<td>-0.04</td>
<td>0.25</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>High Teaching Quality</td>
<td>-0.09</td>
<td>-0.01</td>
<td>-0.23</td>
<td>0.10</td>
<td>0.07</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Note: N= 1,339, values in **bold** = significant at p<0.05.*
<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Professor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Hours Per Week</td>
<td>0.02***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
</tr>
<tr>
<td>Teaching Hours Per Week</td>
<td>-0.04***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
</tr>
<tr>
<td>Research Quality</td>
<td>0.47***</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
</tr>
<tr>
<td>Teaching Quality</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
</tr>
<tr>
<td>Female Academic</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
</tr>
<tr>
<td>Years Post PHD</td>
<td>0.11***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
</tr>
<tr>
<td>Children Number</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
</tr>
<tr>
<td>One Under 5 (Note 1)</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
</tr>
<tr>
<td>Two Under 5s (Note 1)</td>
<td>-0.29</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
</tr>
<tr>
<td>Research Preference Index</td>
<td>0.21**</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
</tr>
<tr>
<td>Teaching Preference Index</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
</tr>
<tr>
<td>Time Preference</td>
<td>117.15**</td>
</tr>
<tr>
<td></td>
<td>(45.51)</td>
</tr>
<tr>
<td>Mobility</td>
<td>0.27**</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
</tr>
<tr>
<td>Bath</td>
<td>-0.51</td>
</tr>
<tr>
<td></td>
<td>(0.45)</td>
</tr>
<tr>
<td>Birmingham</td>
<td>-0.83*</td>
</tr>
<tr>
<td></td>
<td>(0.42)</td>
</tr>
<tr>
<td>Bristol</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>(0.41)</td>
</tr>
<tr>
<td>Cambridge</td>
<td>-0.24</td>
</tr>
<tr>
<td></td>
<td>(0.42)</td>
</tr>
<tr>
<td>Cardiff</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>(0.36)</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>-0.10</td>
</tr>
<tr>
<td></td>
<td>(0.38)</td>
</tr>
<tr>
<td>Essex</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>(0.41)</td>
</tr>
<tr>
<td>University</td>
<td>Score</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Imperial College London</td>
<td>-0.07</td>
</tr>
<tr>
<td>Kent</td>
<td>0.06</td>
</tr>
<tr>
<td>Leeds</td>
<td>-0.21</td>
</tr>
<tr>
<td>Leicester</td>
<td>0.08</td>
</tr>
<tr>
<td>LSE</td>
<td>0.11</td>
</tr>
<tr>
<td>Loughborough</td>
<td>-0.05</td>
</tr>
<tr>
<td>Manchester</td>
<td>-0.15</td>
</tr>
<tr>
<td>Nottingham</td>
<td>-0.20</td>
</tr>
<tr>
<td>Oxford</td>
<td>-0.47</td>
</tr>
<tr>
<td>Queens Belfast</td>
<td>-0.07</td>
</tr>
<tr>
<td>Reading</td>
<td>0.16</td>
</tr>
<tr>
<td>Royal Holloway College</td>
<td>0.27</td>
</tr>
<tr>
<td>Southampton</td>
<td>0.26</td>
</tr>
<tr>
<td>St Andrews</td>
<td>-0.50</td>
</tr>
<tr>
<td>Surrey</td>
<td>0.00</td>
</tr>
<tr>
<td>UCL</td>
<td>0.24</td>
</tr>
<tr>
<td>UEA</td>
<td>0.20</td>
</tr>
<tr>
<td>Warwick</td>
<td>0.30</td>
</tr>
<tr>
<td>York</td>
<td>-0.50</td>
</tr>
<tr>
<td>English</td>
<td>0.81*</td>
</tr>
<tr>
<td>Business/Management</td>
<td>1.53***</td>
</tr>
<tr>
<td>Chemistry</td>
<td>0.61*</td>
</tr>
<tr>
<td>Field</td>
<td>Coefficient</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Computer Science</td>
<td>0.53</td>
</tr>
<tr>
<td>Economics</td>
<td>1.21**</td>
</tr>
<tr>
<td>Geography</td>
<td>0.98**</td>
</tr>
<tr>
<td>History</td>
<td>1.13***</td>
</tr>
<tr>
<td>Law</td>
<td>1.75***</td>
</tr>
<tr>
<td>Mathematics</td>
<td>0.95***</td>
</tr>
<tr>
<td>Physics</td>
<td>0.75**</td>
</tr>
<tr>
<td>Politics/International relations</td>
<td>0.91**</td>
</tr>
<tr>
<td>Psychology</td>
<td>1.08***</td>
</tr>
<tr>
<td>Sociology</td>
<td>1.02**</td>
</tr>
<tr>
<td>Other</td>
<td>1.19***</td>
</tr>
<tr>
<td>Constant</td>
<td>-121.37***</td>
</tr>
</tbody>
</table>

Observations: 1,102  
Pseudo r-squared: 0.44

Standard errors in parentheses  *** p<0.01, ** p<0.05, * p<0.1

Note 1: the control variables for children under 5 are mutually exclusive
TABLE 6
Regression for research hours per week on preferences and gender.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Research Hours Per Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor Level</td>
<td>2.94***</td>
</tr>
<tr>
<td></td>
<td>(0.71)</td>
</tr>
<tr>
<td>Female Academic</td>
<td>-1.79***</td>
</tr>
<tr>
<td></td>
<td>(0.56)</td>
</tr>
<tr>
<td>Years Post PHD</td>
<td>-0.12***</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
</tr>
<tr>
<td>Research Quality</td>
<td>2.54***</td>
</tr>
<tr>
<td></td>
<td>(0.53)</td>
</tr>
<tr>
<td>Teaching Quality</td>
<td>-1.44***</td>
</tr>
<tr>
<td></td>
<td>(0.55)</td>
</tr>
<tr>
<td>Children Number</td>
<td>-0.27</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
</tr>
<tr>
<td>One Under 5</td>
<td>-1.27*</td>
</tr>
<tr>
<td></td>
<td>(0.75)</td>
</tr>
<tr>
<td>Two Under 5s</td>
<td>-0.51</td>
</tr>
<tr>
<td></td>
<td>(1.20)</td>
</tr>
<tr>
<td>Research Preference Index</td>
<td>5.08***</td>
</tr>
<tr>
<td></td>
<td>(0.41)</td>
</tr>
<tr>
<td>Teaching Preference Index</td>
<td>-3.07***</td>
</tr>
<tr>
<td></td>
<td>(0.38)</td>
</tr>
<tr>
<td>Time Preference</td>
<td>-175.93</td>
</tr>
<tr>
<td></td>
<td>(197.26)</td>
</tr>
<tr>
<td>Mobility</td>
<td>-0.90*</td>
</tr>
<tr>
<td></td>
<td>(0.51)</td>
</tr>
<tr>
<td>Bath</td>
<td>-0.95</td>
</tr>
<tr>
<td></td>
<td>(2.10)</td>
</tr>
<tr>
<td>Birmingham</td>
<td>1.48</td>
</tr>
<tr>
<td></td>
<td>(1.96)</td>
</tr>
<tr>
<td>Bristol</td>
<td>1.18</td>
</tr>
<tr>
<td></td>
<td>(1.97)</td>
</tr>
<tr>
<td>Cambridge</td>
<td>2.72</td>
</tr>
<tr>
<td></td>
<td>(2.01)</td>
</tr>
<tr>
<td>Cardiff</td>
<td>-1.17</td>
</tr>
<tr>
<td></td>
<td>(1.71)</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>-1.80</td>
</tr>
<tr>
<td></td>
<td>(1.86)</td>
</tr>
<tr>
<td>Essex</td>
<td>-2.01</td>
</tr>
<tr>
<td></td>
<td>(1.95)</td>
</tr>
<tr>
<td>Imperial College London</td>
<td>-0.40</td>
</tr>
<tr>
<td></td>
<td>(3.12)</td>
</tr>
<tr>
<td>University</td>
<td>Score</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Kent</td>
<td>0.92</td>
</tr>
<tr>
<td>Leeds</td>
<td>-1.52</td>
</tr>
<tr>
<td>Leicester</td>
<td>-1.41</td>
</tr>
<tr>
<td>LSE</td>
<td>-1.50</td>
</tr>
<tr>
<td>Loughborough</td>
<td>-2.01</td>
</tr>
<tr>
<td>Manchester</td>
<td>-0.54</td>
</tr>
<tr>
<td>Nottingham</td>
<td>-1.22</td>
</tr>
<tr>
<td>Oxford</td>
<td>2.65</td>
</tr>
<tr>
<td>Queens Belfast</td>
<td>-0.27</td>
</tr>
<tr>
<td>Reading</td>
<td>-0.43</td>
</tr>
<tr>
<td>Royal Holloway College</td>
<td>-1.59</td>
</tr>
<tr>
<td>Southampton</td>
<td>3.82</td>
</tr>
<tr>
<td>St Andrews</td>
<td>1.39</td>
</tr>
<tr>
<td>Surrey</td>
<td>-0.28</td>
</tr>
<tr>
<td>UCL</td>
<td>0.73</td>
</tr>
<tr>
<td>UEA</td>
<td>-1.20</td>
</tr>
<tr>
<td>Warwick</td>
<td>1.07</td>
</tr>
<tr>
<td>York</td>
<td>-0.48</td>
</tr>
<tr>
<td>English</td>
<td>-5.78***</td>
</tr>
<tr>
<td>Business/Management</td>
<td>-1.95</td>
</tr>
<tr>
<td>Chemistry</td>
<td>0.95</td>
</tr>
<tr>
<td>Computer Science</td>
<td>-0.15</td>
</tr>
<tr>
<td>Field</td>
<td>Coefficient</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Economics</td>
<td>-0.78</td>
</tr>
<tr>
<td>Geography</td>
<td>-2.21</td>
</tr>
<tr>
<td>History</td>
<td>-5.95***</td>
</tr>
<tr>
<td>Law</td>
<td>-3.75**</td>
</tr>
<tr>
<td>Mathematics</td>
<td>-1.95</td>
</tr>
<tr>
<td>Physics</td>
<td>-1.68</td>
</tr>
<tr>
<td>Politics/International relations</td>
<td>-3.63**</td>
</tr>
<tr>
<td>Psychology</td>
<td>-3.33**</td>
</tr>
<tr>
<td>Sociology</td>
<td>-3.38*</td>
</tr>
<tr>
<td>Other</td>
<td>-1.02</td>
</tr>
<tr>
<td>Constant</td>
<td>189.78</td>
</tr>
</tbody>
</table>

**Observations**: 1,102  
**R-squared**: 0.32

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

*Note 1*: the control variables for children under 5 are mutually exclusive
TABLE 7

*Standardized Mediation Effects between gender and research hours per week.*

<table>
<thead>
<tr>
<th></th>
<th>Indirect effect of IV on DV</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direct Effect of IV on DV</td>
<td>Via Research Preferences</td>
<td>Via Teaching Preferences</td>
<td>Via Delta (Time Preference)</td>
<td>Total Effect IV on DV</td>
</tr>
<tr>
<td>Gender</td>
<td>-0.09</td>
<td>-0.02</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

*Note: N= 1,102, DV = dependent variable, IV = independent variable (Gender), Values in bold = significant at p < .05.*
TABLE 8
Standardized Mediation Effects between gender and research hours per week using pattern analysis.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Direct Effect of IV on DV</th>
<th>Via Aligned Pattern</th>
<th>Via Misaligned Pattern</th>
<th>Total Indirect Effect IV on DV</th>
<th>Total Effect IV on DV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.09</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.11</td>
</tr>
</tbody>
</table>

Note: N= 1,102, DV = dependent variable, IV = independent variable (Gender), Values in **bold** = significant at p < .05.
### TABLE 9
*Probit regression for Professor dummy variable (including administration hours)*

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Professor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Hours Per Week</td>
<td>0.03***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
</tr>
<tr>
<td>Teaching Hours Per Week</td>
<td>-0.03***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
</tr>
<tr>
<td>Administration Hours Per Week</td>
<td>0.03***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
</tr>
<tr>
<td>Research Quality</td>
<td>0.41***</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
</tr>
<tr>
<td>Teaching Quality</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
</tr>
<tr>
<td>Female Academic</td>
<td>-0.10</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
</tr>
<tr>
<td>Years Post PHD</td>
<td>0.11***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
</tr>
<tr>
<td>Children Number</td>
<td>0.11*</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
</tr>
<tr>
<td>One Under 5 (Note 1)</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
</tr>
<tr>
<td>Two Under 5s (Note 1)</td>
<td>-0.30</td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
</tr>
<tr>
<td>Research Preference Index</td>
<td>0.24**</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
</tr>
<tr>
<td>Teaching Preference Index</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
</tr>
<tr>
<td>Time Preference</td>
<td>115.39**</td>
</tr>
<tr>
<td></td>
<td>(46.10)</td>
</tr>
<tr>
<td>Mobility</td>
<td>0.26**</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
</tr>
<tr>
<td>Bath</td>
<td>-0.54</td>
</tr>
<tr>
<td></td>
<td>(0.46)</td>
</tr>
<tr>
<td>Birmingham</td>
<td>-0.88*</td>
</tr>
<tr>
<td></td>
<td>(0.43)</td>
</tr>
<tr>
<td>Bristol</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.41)</td>
</tr>
<tr>
<td>Cambridge</td>
<td>-0.30</td>
</tr>
<tr>
<td></td>
<td>(0.43)</td>
</tr>
<tr>
<td>Cardiff</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>(0.36)</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>-0.14</td>
</tr>
<tr>
<td></td>
<td>(0.39)</td>
</tr>
<tr>
<td>Institution</td>
<td>R-value</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Essex</td>
<td>0.47</td>
</tr>
<tr>
<td>Imperial College London</td>
<td>0.11</td>
</tr>
<tr>
<td>Kent</td>
<td>0.07</td>
</tr>
<tr>
<td>Leeds</td>
<td>-0.28</td>
</tr>
<tr>
<td>Leicester</td>
<td>0.05</td>
</tr>
<tr>
<td>LSE</td>
<td>0.13</td>
</tr>
<tr>
<td>Loughborough</td>
<td>-0.11</td>
</tr>
<tr>
<td>Manchester</td>
<td>-0.13</td>
</tr>
<tr>
<td>Nottingham</td>
<td>-0.17</td>
</tr>
<tr>
<td>Oxford</td>
<td>-0.48</td>
</tr>
<tr>
<td>Queens Belfast</td>
<td>-0.02</td>
</tr>
<tr>
<td>Reading</td>
<td>0.15</td>
</tr>
<tr>
<td>Royal Holloway College</td>
<td>0.28</td>
</tr>
<tr>
<td>Southampton</td>
<td>0.24</td>
</tr>
<tr>
<td>St Andrews</td>
<td>-0.60</td>
</tr>
<tr>
<td>Surrey</td>
<td>-0.01</td>
</tr>
<tr>
<td>UCL</td>
<td>0.25</td>
</tr>
<tr>
<td>UEA</td>
<td>0.12</td>
</tr>
<tr>
<td>Warwick</td>
<td>0.33</td>
</tr>
<tr>
<td>York</td>
<td>-0.55</td>
</tr>
<tr>
<td>English</td>
<td>0.72</td>
</tr>
<tr>
<td>Business/Management</td>
<td>1.52***</td>
</tr>
<tr>
<td>Subject</td>
<td>Coefficient</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Chemistry</td>
<td>0.54</td>
</tr>
<tr>
<td>Computer Science</td>
<td>0.47</td>
</tr>
<tr>
<td>Economics</td>
<td>1.24***</td>
</tr>
<tr>
<td>Geography</td>
<td>0.94**</td>
</tr>
<tr>
<td>History</td>
<td>1.14***</td>
</tr>
<tr>
<td>Law</td>
<td>1.79***</td>
</tr>
<tr>
<td>Mathematics</td>
<td>1.00***</td>
</tr>
<tr>
<td>Physics</td>
<td>0.78**</td>
</tr>
<tr>
<td>Politics/International relations</td>
<td>0.84**</td>
</tr>
<tr>
<td>Psychology</td>
<td>1.12***</td>
</tr>
<tr>
<td>Sociology</td>
<td>0.98**</td>
</tr>
<tr>
<td>Other</td>
<td>1.14***</td>
</tr>
<tr>
<td>Constant</td>
<td>-120.54***</td>
</tr>
</tbody>
</table>

Observations: 1,102
Pseudo R-squared: 0.45

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

*Note 1:* the control variables for children under 5 are mutually exclusive.
TABLE 10
Regression for research hours per week (including administration hours).

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Research Hours Per Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor Level</td>
<td>3.79***</td>
</tr>
<tr>
<td></td>
<td>(0.67)</td>
</tr>
<tr>
<td>Female Academic</td>
<td>-1.33**</td>
</tr>
<tr>
<td></td>
<td>(0.52)</td>
</tr>
<tr>
<td>Administration Hours Per Week</td>
<td>-0.38***</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
</tr>
<tr>
<td>Years Post PHD</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
</tr>
<tr>
<td>Research Quality</td>
<td>2.84***</td>
</tr>
<tr>
<td></td>
<td>(0.49)</td>
</tr>
<tr>
<td>Teaching Quality</td>
<td>-1.37***</td>
</tr>
<tr>
<td></td>
<td>(0.51)</td>
</tr>
<tr>
<td>Children Number</td>
<td>-0.43</td>
</tr>
<tr>
<td></td>
<td>(0.24)</td>
</tr>
<tr>
<td>One Under 5 (Note 1)</td>
<td>-0.97*</td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
</tr>
<tr>
<td>Two Under 5s (Note 1)</td>
<td>-0.23</td>
</tr>
<tr>
<td></td>
<td>(1.11)</td>
</tr>
<tr>
<td>Research Preference Index</td>
<td>4.28***</td>
</tr>
<tr>
<td></td>
<td>(0.39)</td>
</tr>
<tr>
<td>Teaching Preference Index</td>
<td>-2.78***</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
</tr>
<tr>
<td>Time Preference</td>
<td>-61.32</td>
</tr>
<tr>
<td></td>
<td>(183.12)</td>
</tr>
<tr>
<td>Mobility</td>
<td>-0.61*</td>
</tr>
<tr>
<td></td>
<td>(0.47)</td>
</tr>
<tr>
<td>Bath</td>
<td>-0.54</td>
</tr>
<tr>
<td></td>
<td>(1.95)</td>
</tr>
<tr>
<td>Birmingham</td>
<td>2.55</td>
</tr>
<tr>
<td></td>
<td>(1.82)</td>
</tr>
<tr>
<td>Bristol</td>
<td>1.82</td>
</tr>
<tr>
<td></td>
<td>(1.83)</td>
</tr>
<tr>
<td>Cambridge</td>
<td>3.36</td>
</tr>
<tr>
<td></td>
<td>(1.86)</td>
</tr>
<tr>
<td>Cardiff</td>
<td>-0.75</td>
</tr>
<tr>
<td></td>
<td>(1.59)</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>-0.25</td>
</tr>
<tr>
<td></td>
<td>(1.73)</td>
</tr>
<tr>
<td>Essex</td>
<td>-1.82</td>
</tr>
<tr>
<td>University</td>
<td>Rank</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Imperial College London</td>
<td>-0.42</td>
</tr>
<tr>
<td>Kent</td>
<td>1.24</td>
</tr>
<tr>
<td>Leeds</td>
<td>-0.35</td>
</tr>
<tr>
<td>Leicester</td>
<td>-0.36</td>
</tr>
<tr>
<td>LSE</td>
<td>-0.49</td>
</tr>
<tr>
<td>Loughborough</td>
<td>-0.80</td>
</tr>
<tr>
<td>Manchester</td>
<td>-0.37</td>
</tr>
<tr>
<td>Nottingham</td>
<td>-0.56</td>
</tr>
<tr>
<td>Oxford</td>
<td>2.12</td>
</tr>
<tr>
<td>Queens Belfast</td>
<td>-0.85</td>
</tr>
<tr>
<td>Reading</td>
<td>1.09</td>
</tr>
<tr>
<td>Royal Holloway College</td>
<td>-1.14</td>
</tr>
<tr>
<td>Southampton</td>
<td>3.58</td>
</tr>
<tr>
<td>St Andrews</td>
<td>2.06</td>
</tr>
<tr>
<td>Surrey</td>
<td>0.49</td>
</tr>
<tr>
<td>UCL</td>
<td>0.97</td>
</tr>
<tr>
<td>UEA</td>
<td>0.13</td>
</tr>
<tr>
<td>Warwick</td>
<td>1.15</td>
</tr>
<tr>
<td>York</td>
<td>0.75</td>
</tr>
<tr>
<td>English</td>
<td>-3.99**</td>
</tr>
<tr>
<td>Business/Management</td>
<td>-1.44</td>
</tr>
<tr>
<td>Chemistry</td>
<td>1.47</td>
</tr>
<tr>
<td>Field</td>
<td>Coefficient</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Computer Science</td>
<td>1.08</td>
</tr>
<tr>
<td>Economic</td>
<td>-0.43</td>
</tr>
<tr>
<td>Geography</td>
<td>-1.87</td>
</tr>
<tr>
<td>History</td>
<td>-5.28***</td>
</tr>
<tr>
<td>Law</td>
<td>-3.44**</td>
</tr>
<tr>
<td>Mathematics</td>
<td>-2.19*</td>
</tr>
<tr>
<td>Physics</td>
<td>-1.72</td>
</tr>
<tr>
<td>Politics/International relations</td>
<td>-2.27</td>
</tr>
<tr>
<td>Psychology</td>
<td>-3.10**</td>
</tr>
<tr>
<td>Sociology</td>
<td>-2.36</td>
</tr>
<tr>
<td>Other</td>
<td>-0.68</td>
</tr>
<tr>
<td>Constant</td>
<td>79.82</td>
</tr>
<tr>
<td>Observations</td>
<td>1,102</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Note 1: the control variables for children under 5 are mutually exclusive.
<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Hours Allocated to Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor Level</td>
<td>2.82***</td>
</tr>
<tr>
<td></td>
<td>(0.72)</td>
</tr>
<tr>
<td>Female Academic</td>
<td>4.93</td>
</tr>
<tr>
<td></td>
<td>(7.93)</td>
</tr>
<tr>
<td>Research Preference Index</td>
<td>6.67***</td>
</tr>
<tr>
<td></td>
<td>(1.58)</td>
</tr>
<tr>
<td>Female x Research Pref.</td>
<td>-3.51</td>
</tr>
<tr>
<td></td>
<td>(2.48)</td>
</tr>
<tr>
<td>Teaching Preference Index</td>
<td>-2.77</td>
</tr>
<tr>
<td></td>
<td>(2.05)</td>
</tr>
<tr>
<td>Female x Teaching Preference</td>
<td>-0.27</td>
</tr>
<tr>
<td></td>
<td>(3.03)</td>
</tr>
<tr>
<td>Research Preference x Teaching Preference</td>
<td>-0.32</td>
</tr>
<tr>
<td></td>
<td>(0.65)</td>
</tr>
<tr>
<td>Female x Research Preference x Teaching Preference</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>(0.97)</td>
</tr>
<tr>
<td>Years Post PHD</td>
<td>-0.12***</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
</tr>
<tr>
<td>Research Quality</td>
<td>2.43***</td>
</tr>
<tr>
<td></td>
<td>(0.53)</td>
</tr>
<tr>
<td>Teaching Quality</td>
<td>-1.50***</td>
</tr>
<tr>
<td></td>
<td>(0.55)</td>
</tr>
<tr>
<td>Children Number</td>
<td>-0.27</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
</tr>
<tr>
<td>One Under 5 (Note 1)</td>
<td>-1.44*</td>
</tr>
<tr>
<td></td>
<td>(0.75)</td>
</tr>
<tr>
<td>Two Under 5s (Note 1)</td>
<td>-0.53</td>
</tr>
<tr>
<td></td>
<td>(1.20)</td>
</tr>
<tr>
<td>Time Preference</td>
<td>-123.06</td>
</tr>
<tr>
<td></td>
<td>(197.20)</td>
</tr>
<tr>
<td>Mobility</td>
<td>-0.90*</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
</tr>
<tr>
<td>Bath</td>
<td>-0.48</td>
</tr>
<tr>
<td></td>
<td>(2.10)</td>
</tr>
<tr>
<td>Birmingham</td>
<td>1.76</td>
</tr>
<tr>
<td></td>
<td>(1.95)</td>
</tr>
<tr>
<td>Bristol</td>
<td>1.23</td>
</tr>
<tr>
<td></td>
<td>(1.97)</td>
</tr>
<tr>
<td>Cambridge</td>
<td>3.17</td>
</tr>
<tr>
<td></td>
<td>(2.00)</td>
</tr>
<tr>
<td>Cardiff</td>
<td>-0.92</td>
</tr>
<tr>
<td></td>
<td>(1.71)</td>
</tr>
<tr>
<td>Institution</td>
<td>Score</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>-1.63</td>
</tr>
<tr>
<td>Essex</td>
<td>-1.69</td>
</tr>
<tr>
<td>Imperial College London</td>
<td>-0.18</td>
</tr>
<tr>
<td>Kent</td>
<td>1.05</td>
</tr>
<tr>
<td>Leeds</td>
<td>-1.49</td>
</tr>
<tr>
<td>Leicester</td>
<td>-0.96</td>
</tr>
<tr>
<td>LSE</td>
<td>-1.41</td>
</tr>
<tr>
<td>Loughborough</td>
<td>-1.83</td>
</tr>
<tr>
<td>Manchester</td>
<td>-0.35</td>
</tr>
<tr>
<td>Nottingham</td>
<td>-1.08</td>
</tr>
<tr>
<td>Oxford</td>
<td>2.90</td>
</tr>
<tr>
<td>Queens Belfast</td>
<td>0.10</td>
</tr>
<tr>
<td>Reading</td>
<td>-0.75</td>
</tr>
<tr>
<td>Royal Holloway College</td>
<td>-1.25</td>
</tr>
<tr>
<td>Southampton</td>
<td>3.85</td>
</tr>
<tr>
<td>St Andrews</td>
<td>1.41</td>
</tr>
<tr>
<td>Surrey</td>
<td>0.15</td>
</tr>
<tr>
<td>UCL</td>
<td>0.93</td>
</tr>
<tr>
<td>UEA</td>
<td>-1.08</td>
</tr>
<tr>
<td>Warwick</td>
<td>1.27</td>
</tr>
<tr>
<td>York</td>
<td>-0.18</td>
</tr>
<tr>
<td>English</td>
<td>-5.61***</td>
</tr>
<tr>
<td>Business/Management</td>
<td>-1.68</td>
</tr>
<tr>
<td>Category</td>
<td>Coefficient</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Chemistry</td>
<td>1.36</td>
</tr>
<tr>
<td>Computer Science</td>
<td>0.02</td>
</tr>
<tr>
<td>Economics</td>
<td>-0.44</td>
</tr>
<tr>
<td>Geography</td>
<td>-1.60</td>
</tr>
<tr>
<td>History</td>
<td>-5.67***</td>
</tr>
<tr>
<td>Law</td>
<td>-3.49**</td>
</tr>
<tr>
<td>Mathematics</td>
<td>-1.48</td>
</tr>
<tr>
<td>Physics</td>
<td>-1.69</td>
</tr>
<tr>
<td>Politics/International relations</td>
<td>-3.44**</td>
</tr>
<tr>
<td>Psychology</td>
<td>-2.98*</td>
</tr>
<tr>
<td>Sociology</td>
<td>-3.14*</td>
</tr>
<tr>
<td>Other</td>
<td>-1.10</td>
</tr>
<tr>
<td>Constant</td>
<td>133.19</td>
</tr>
<tr>
<td>Observations</td>
<td>1,102</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

*Note 1:* the control variables for children under 5 are mutually exclusive.
Appendix A


Original (Reminder) Cover un incentivised Email

Date: (Today’s Date)

Dear (First Name),

(Reminder) Higher Education Survey

We would like to invite you to take part today in a short survey relating to incentives and motivation in the Higher Education Sector. (You should have already received an invitation to take part in the national Higher Education Survey. We have noticed that you have yet to participate in this survey.) This survey is a part of a major national research project. The benefit from completing this survey will come from studying how academics approach multi-tasking conflicts in the context of increasing demands on academics for both research productivity and teaching quality.

The deadline for completion, after which the personalized link below will expire, is 21 (14) days from receipt of this invitation. Completing this survey should take 15 minutes {, and you will be eligible for two prizes as follows:

- Prize of £500 for the best prediction of your colleagues' views.
- Multiple prizes of up to £80 drawn at random for every 300 submissions received.}

If you have any queries, please feel free to contact our survey team at deliveringbetter@surrey.ac.uk

To participate please click on the personalised link below.

Follow this link to the Survey:
$:!://SurveyLink?d=Take the Survey

Or copy and paste the URL below into your internet browser:
$:!://SurveyURL

This link will be valid for 21 (14) days from the date of receipt.

Thank You
Professor Graham Cookson,
Surrey Business School,
University of Surrey.
Graham Cookson Profile
Delivering Better For Less Website
The study has been reviewed and received a Favourable Ethical Opinion (FEO) from the University of Surrey Ethics Committee.

Data protection notice:
We obtained your contact details from publically available data from your academic institution’s website. Please see the Participant Information Sheet for details of our ethics policy.
If you do not wish to be contacted again, please click here to unsubscribe.

[Introductory Screen]

Welcome to the Higher Education Survey. Please press I CONSENT and NEXT if you agree with the terms below to continue to the survey.

Informed Consent - Higher Education Survey

- I have read and understood the Participant Information Sheet provided (version 1.0, dated 08/02/2016).
- I agree for my anonymised data to be used for this study.
- I understand that data will be collected using Qualtrics software. This software is used by academic institutions worldwide and Qualtrics give assurances that data collected in the EU will remain in the EU. I understand that all the data collected from my responses will be anonymised and the collected data points will not be linked to my identity.
- I understand that all project data will be held for at least 6 years and all research data for at least 10 years in accordance with University policy and that my personal data is held and processed in the strictest confidence, and in accordance with the Data Protection Act (1998).
- I understand that I can request for my data to be withdrawn up to one month after completing the survey and that following my request all data already collected from me will be deleted.
- I confirm that I have read and understood the above and freely consent to participating in this study. I have been given adequate time to consider my participation.

I CONSENT and wish to complete the survey.

I DO NOT CONSENT and wish to exit the survey now.

[Initial question Block]

Please tell us about yourself under the following headings:

Personal Information
Are you Male or Female?

What is your age?

**Academic Information**

Which University do you work at?

Are you a full-time or part-time member of staff?

How many hours per week on average do you work in reality (rather than contracted)?

Which School or Department do you associate yourself most with at your University?

Over a full year what percentage of your time do you spend on the following?

Research (%)
Teaching (%)
Administration and other (%)

How many years is it since you completed your PhD or other Highest Degree?

How many different universities have you been employed by?

What best describes your Academic Title?

[Own Norms Screens]

On a scale of 0% to 100%, what percentage of YOUR OWN working time over a full year do you spend on each of the following activities?

% of working time on preparation for lectures.
% of working time on additional assessment of learning, e.g. essays and quizzes
% of working time with your door open to indicate availability for interruption.
% of working time with your door locked whilst in your office to avoid interruptions.
% of working time focusing on low achieving or disadvantaged students.

[Peer Norms Screen]

**Prize draw for most accurate guess of other academics' responses**

There will be a prize of a £500 for the academic who most accurately guesses the mean of responses to the following questions. In the event of a tie, the winner will be selected at random. If you win, you will be able to select whether you would like payment to be credited directly to your bank account or to be in the form of gift cards to the value of £500 from any major online retailer. *
On a scale of 0% to 100%, what percentage of THEIR WORKING TIME DO OTHER ACADEMICS in your discipline and in comparable institutions CONSIDER PROFESSIONALLY APPROPRIATE to spend on each of the following activities over a full year?

If you would like to know the means of actual and guessed answers of other academics in your discipline, please choose this option in the box below, and the information will be sent to you after the completion of the study.

% of working time on preparation for lectures.
% of working time on additional assessment of learning, e.g. essays and quizzes
% of working time with your door open to indicate availability for interruption.
% of working time with your door locked whilst in your office to avoid interruptions.
% of working time focusing on low achieving or disadvantaged students.

* The accuracy of your guesses will be judged by the lowest sum of absolute deviations from the mean of each response within your discipline.

Research Preferences

On a scale from 1 (strongly disagree) to 5 (strongly agree), please indicate your views on each statement below.

In order to be successful in academia, I need to focus on my research.

My first priority is research

I am active (in the scholarship of) research

I have recently participated in professional development in research

I enjoy participating in research

Teaching Preferences

On a scale from 1 (strongly disagree) to 5 (strongly agree), please indicate your views on each statement below.

In order to be successful in academia, I need to focus on my teaching

My first priority is teaching

I am active (in the scholarship of) teaching

I have recently participated in professional development in teaching

I enjoy participating in teaching
[Trait Guilt Questions]

In this short questionnaire you will read about situations that people are likely to encounter in day-to-day life, followed by common reactions to those situations. As you read each scenario, try to imagine yourself in that situation. Then indicate the likelihood that you would react in the way described.

Scale of 1 (very unlikely) to 5 (very likely)

After realizing you have received too much change at a shop, you decide to keep it because the sales assistant does not notice. What is the likelihood that you would feel uncomfortable about keeping the money?

You secretly commit a crime. What is the likelihood that you would feel remorse about breaking the law?

At a co-worker’s housewarming party, you spill red wine on their new cream-coloured carpet. You cover the stain with a chair so that nobody notices your mess. What is the likelihood that you would feel that the way you acted was pathetic?

You lie to people but they never find out about it. What is the likelihood that you would feel terrible about the lies you told?

Out of frustration, you break the photocopier at work. Nobody is around and you leave without telling anyone. What is the likelihood you would feel bad about the way you acted?

[Measure of Time Preference]

Scale of 1 (very unwilling) to 5 (very willing)

In comparison to others, are you a person who is willing to give up something today in order to benefit in the future?

[Time Preference Tables]

In this page you are asked to complete 6 tables. Each row of each table has two choice options, and you have to pick one option for each row. You have a 1 in 300 chance of being paid according to your choices below in one randomly chosen row from one randomly chosen table.

Table 1

For each of the following seven choice pairs, would you prefer to receive the amount in the left hand column in 2 days’ time or the fixed amount of £80 in the right hand column in 9 days’ time? Make your selections for each row by choosing either the amount payable in 2 days’
time or the fixed amount of £80 payable in 9 days' time. **YOU MAY SWITCH BETWEEN COLUMNS ONLY ONCE** in making your choices.

<table>
<thead>
<tr>
<th></th>
<th>2 days’ time</th>
<th>9 days’ time</th>
<th>30 days’ time</th>
</tr>
</thead>
<tbody>
<tr>
<td>£75 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 9 days’ time</td>
</tr>
<tr>
<td>£70 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 9 days’ time</td>
</tr>
<tr>
<td>£65 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 9 days’ time</td>
</tr>
<tr>
<td>£60 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 9 days’ time</td>
</tr>
<tr>
<td>£55 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 9 days’ time</td>
</tr>
<tr>
<td>£50 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 9 days’ time</td>
</tr>
<tr>
<td>£45 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 9 days’ time</td>
</tr>
</tbody>
</table>

Table 2

For each of the following seven choice pairs, would you prefer to receive the amount in the left hand column in 2 days’ time or the fixed amount of £80 in the right hand column in 90 days’ time? Make your selections for each row by choosing either the amount payable in 2 days’ time or the fixed amount of £80 payable in 90 days' time. **YOU MAY SWITCH BETWEEN COLUMNS ONLY ONCE** in making your choices.

<table>
<thead>
<tr>
<th></th>
<th>2 days’ time</th>
<th>30 days’ time</th>
<th>90 days’ time</th>
</tr>
</thead>
<tbody>
<tr>
<td>£75 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 90 days’ time</td>
</tr>
<tr>
<td>£70 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 90 days’ time</td>
</tr>
<tr>
<td>£65 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 90 days’ time</td>
</tr>
<tr>
<td>£60 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 90 days’ time</td>
</tr>
<tr>
<td>£55 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 90 days’ time</td>
</tr>
<tr>
<td>£50 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 90 days’ time</td>
</tr>
<tr>
<td>£45 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 90 days’ time</td>
</tr>
</tbody>
</table>

Table 3

For each of the following seven choice pairs, would you prefer to receive the amount in the left hand column in 2 days’ time or the fixed amount of £80 in the right hand column in 90 days’ time? Make your selections for each row by choosing either the amount payable in 2 days' time or the fixed amount of £80 payable in 90 days' time. **YOU MAY SWITCH BETWEEN COLUMNS ONLY ONCE** in making your choices.

<table>
<thead>
<tr>
<th></th>
<th>2 days’ time</th>
<th>90 days’ time</th>
<th>90 days’ time</th>
</tr>
</thead>
<tbody>
<tr>
<td>£75 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 90 days’ time</td>
</tr>
<tr>
<td>£70 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 90 days’ time</td>
</tr>
<tr>
<td>£65 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 90 days’ time</td>
</tr>
<tr>
<td>£60 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 90 days’ time</td>
</tr>
<tr>
<td>£55 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 90 days’ time</td>
</tr>
<tr>
<td>£50 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 90 days’ time</td>
</tr>
<tr>
<td>£45 in 2 days’ time</td>
<td>o</td>
<td>o</td>
<td>£80 in 90 days’ time</td>
</tr>
</tbody>
</table>

Table 4
For each of the following seven choice pairs, would you prefer to receive the amount in the left hand column in 2 days’ time or the fixed amount of £80 in the right hand column in 180 days’ time? Make your selections for each row by choosing either the amount payable in 2 days’ time or the fixed amount of £80 payable in 180 days’ time. **YOU MAY SWITCH BETWEEN COLUMNS ONLY ONCE** in making your choices.

<table>
<thead>
<tr>
<th></th>
<th>2 days’ time</th>
<th>180 days’ time</th>
</tr>
</thead>
<tbody>
<tr>
<td>£75 in 2 days’ time</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>£70 in 2 days’ time</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>£65 in 2 days’ time</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>£60 in 2 days’ time</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>£55 in 2 days’ time</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>£50 in 2 days’ time</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>£45 in 2 days’ time</td>
<td>o</td>
<td>o</td>
</tr>
</tbody>
</table>

**Table 5**

For each of the following 10 choice pairs, would you prefer to receive the lottery on the left or the lottery on the right? The lottery on the left will pay a combination of £50 and £35, each with the varying probabilities (PR) shown. The lottery on the right will pay a combination of £80 and £10, each with the probabilities (PR) shown. Make your selections for each row by choosing either the lottery on the left or the lottery on the right. **YOU MAY SWITCH BETWEEN COLUMNS ONLY ONCE** in making your choices.

<table>
<thead>
<tr>
<th>Lottery on the left</th>
<th>Lottery on the right</th>
</tr>
</thead>
<tbody>
<tr>
<td>£50 (PR 10%) or £35 (PR 90%)</td>
<td>o</td>
</tr>
<tr>
<td>£50 (PR 20%) or £35 (PR 80%)</td>
<td>o</td>
</tr>
<tr>
<td>£50 (PR 30%) or £35 (PR 70%)</td>
<td>o</td>
</tr>
<tr>
<td>£50 (PR 40%) or £35 (PR 60%)</td>
<td>o</td>
</tr>
<tr>
<td>£50 (PR 50%) or £35 (PR 50%)</td>
<td>o</td>
</tr>
<tr>
<td>£50 (PR 60%) or £35 (PR 40%)</td>
<td>o</td>
</tr>
<tr>
<td>£50 (PR 70%) or £35 (PR 30%)</td>
<td>o</td>
</tr>
<tr>
<td>£50 (PR 80%) or £35 (PR 20%)</td>
<td>o</td>
</tr>
<tr>
<td>£50 (PR 90%) or £35 (PR 10%)</td>
<td>o</td>
</tr>
<tr>
<td>£50 (PR 100%) or £35 (PR 0%)</td>
<td>o</td>
</tr>
</tbody>
</table>

**Table 6**

For each of the following 10 choice pairs, would you prefer to receive the lottery on the left or the lottery on the right? The lottery on the left will pay a combination of £60 and £30, each with the varying probabilities (PR) shown. The lottery on the right will pay a combination of £100 and £20, each with the probabilities (PR) shown. Make your selections for each row by choosing either the lottery on the left or the lottery on the right. **YOU MAY SWITCH BETWEEN COLUMNS ONLY ONCE** in making your choices.

<table>
<thead>
<tr>
<th>Lottery on the left</th>
<th>Lottery on the right</th>
</tr>
</thead>
<tbody>
<tr>
<td>£60 (PR 10%) or £30 (PR 90%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 20%) or £30 (PR 80%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 30%) or £30 (PR 70%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 40%) or £30 (PR 60%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 50%) or £30 (PR 50%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 60%) or £30 (PR 40%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 70%) or £30 (PR 30%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 80%) or £30 (PR 20%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 90%) or £30 (PR 10%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 100%) or £30 (PR 0%)</td>
<td>o</td>
</tr>
<tr>
<td>Lottery on the left</td>
<td>Lottery on the right</td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>£60 (PR 10%) or £30 (PR 90%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 20%) or £30 (PR 80%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 30%) or £30 (PR 70%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 40%) or £30 (PR 60%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 50%) or £30 (PR 50%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 60%) or £30 (PR 40%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 70%) or £30 (PR 30%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 80%) or £30 (PR 20%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 90%) or £30 (PR 10%)</td>
<td>o</td>
</tr>
<tr>
<td>£60 (PR 100%) or £30 (PR 0%)</td>
<td>o</td>
</tr>
</tbody>
</table>

**Remaining Academic Information**

**Please provide some final academic and personal information before finding out if you have won a prize**

For the last 3 years

Using metrics appropriate to your subject area, how many peer reviewed articles of 2*, 3* and 4* quality have you had published?

For the last 3 years

How many books have you had published?

For the last 3 years

How many unpublished discussion papers, or the equivalent for your subject area, have you produced?

How many scientific works in progress, or the equivalent in your subject area, do you currently have?

For the last 3 years

How many research grants have you obtained and what was their value?

For the last academic year

How many hours have you spent teaching in the classroom?

For the last academic year

How many students have you taught or assessed?

For the previous academic year
For how many teaching modules did you receive an average module evaluation score submitted by your students? Answer zero if your university does not operate a teaching module evaluation system with 5 as the highest score.

Remaining Personal Information

Where have you spent most of your life?

What is your Marital Status?

In which band is your total household income?

Up to £40,000
£40,001 to £70,000
£70,001 to £100,000
£100,001 to £150,000
£150,001 and above

How Many Children do you have?

[Prize Preamble]

Thank you for filling out the survey.

A number will now be randomly generated between 1 and 300. If the number is equal to 300, you will win a prize: the amount you could win will depend on your choice in {Randomised Table/Row}. You chose to receive £ {Corresponding Amount}. This entry has been randomly selected.

Press "Next" to generate a number between 1 and 300.

BAD LUCK - YOU HAVE NOT WON A PRIZE. Your number was {Random number between 1 & 300}. You still have a chance to win a prize of £500 for most accurately guessing the mean of other academics' responses in the "Prize draw for most accurate guess of other academics'

[CONGRATULATIONS - YOU HAVE WON A PRIZE. Your number was ${Random number between 1 & 300}.

You will be paid according to your choice in {Randomised Table/Row} which was £ {Corresponding Amount}. This entry was randomly selected.

You also have a chance to win a prize of £500 for most accurately guessing the mean of other academics' responses in the "Prize draw for most accurate guess of other academics responses".]

If you would like to be considered for [receive] a payment, please enter your email address below:
Please also choose whether you would like to be paid by direct credit to your bank account or by gift cards. You do not need to provide your bank details at this stage. We will contact you with further details.
Chapter 3
“Why give more? The role of student expectations and engagement in extra-curricular teaching effort.”
Abstract

Academics in higher education in the UK face increasing and conflicting demands for both research and teaching excellence. Career payoffs are higher for research excellence leading to the question of what motivates extra-curricular teaching effort. I test whether student expectations of support and student engagement motivates extra-curricular teaching effort, as well as examine gender differences in the responsiveness to expectations and engagement. I find that expectations and engagement affect extra-curricular teaching effort positively for both male and female academics.

Introduction

UK universities now compete in global markets for students who are likely to pay close attention to how different institutions perform. The publication of the latest national and international league tables often makes headlines and influences students’ choice of universities (Chevalier & Jia, 2015; Gibbons, Neumayer, & Perkins, 2015). The removal of caps on student numbers has exacerbated competition (Fazackerley, 2017). Teaching quality is under intense public scrutiny with a steep increase in tuition fees driving student value for money concerns, and the introduction of the Teaching Excellence Framework adding to the focus on teaching quality (Department for Business Innovation and Skills, 2016). An increasing focus on teaching to enhance employment prospects may have added to the conflict between time spent on research versus teaching (Canaan & Shumar, 2008; Gautier & Wauthy, 2007; Hattie & Marsh, 1996).

As discussed in chapter two, despite reduced academic autonomy associated with the marketization of higher education (J. Kenny & Fluck, 2014; J. Kenny, 2017; Olssen & Peters, 2005; Olssen, 2016; Vardi, 2009), the remaining discretion in time allocation may lead to a
difference in focus between research and teaching (Allgood & Walstad, 2013; Boyd & Smith, 2016; Gautier & Wauthy, 2007; Kossi et al., 2015; Wilkin & Tavernier, 2002). The value of research activities to career progression in the higher education sector is frequently highlighted (Barbezat, 2006; Hattie & Marsh, 1996; Rey, 2001). However, the lack of career incentives to devote extra-curricular teaching effort to students raises the question of what motivates academics in going the extra mile in practice.

This paper examines the multi-tasking conflict academics face between research and teaching from a behavioural point of view, given the importance of research activities to career progression. With a randomised experiment, embedded in an online survey, I test academics’ responses to varying levels of student expectations of teaching support and student engagement. I use the social exchange theory (Blau, 1964; Emerson, 1976) and the theory of guilt aversion (Battigalli & Dufwenberg, 2007) to predict positive effects of these variables on the academics’ willingness to support students when academics experience a conflict between research and teaching responsibilities. I also test whether female academics are more responsive to student expectations and engagement. This follows from previous research suggesting that women may be more other-regarding than men, including a greater sensitivity to social cues (Croson & Gneezy, 2009; Kahn, Hottes, & Davis, 1971) and a greater disposition to experience guilt (Else-Quest, Higgins, Allison, & Morton, 2012). Whilst gender differences have not previously been identified in guilt-driven behaviour in one-off economic games, the ongoing relationship between academic and student is a setting in which gender differences could manifest.

My paper is the first to apply the behavioural economic concept of guilt aversion, in response to the effects of student expectations of support and student engagement, as a motivation for the discretionary allocation of time between research and teaching in the higher education sector. I find that high student expectations of support and student
engagement give rise to extra-curricular teaching effort. The findings are mixed with regard to the mechanism for these effects. On the one hand, I control for individual differences in guilt proneness (T. Cohen et al., 2014), and show a significant correlation of guilt proneness with extra-curricular teaching effort. This provides support for guilt aversion possibly being a mechanism for in-the-moment effects of student expectations and engagement on extra-curricular teaching effort. On the other hand, I find that female academics are more prone to guilt than male academics. Yet, there are no gender differences in how female versus male academics respond to student expectations of support and student engagement. I discuss these seemingly conflicting findings and suggest promising avenues for future research. My work contributes to the literature on social exchange theory (Blau, 1964; Emerson, 1976) and guilt aversion (Battigalli & Dufwenberg, 2007; Charness & Dufwenberg, 2006).

In what follows, I set out the theoretical framework for my research hypotheses. I describe next my experimental design, data analysis, experimental validity checks, and conclude with the discussions of all findings.

Theory Development

In what follows, I provide the rationale for why academics may engage in extra-curricular teaching effort and what is the role of student expectations of support and student engagement. I formulate predictions that differ for male and female academics. Central to my hypotheses will be the social exchange theory (Blau, 1964; Emerson, 1976) and the concept of guilt aversion (Battigalli & Dufwenberg, 2007; Charness & Dufwenberg, 2006), topics which I discuss below.

Social Exchange Theory
According to social exchange theory, parties to a relationship remain satisfied and committed as long as the benefits of the relationship outweigh its costs (Blau, 1964; Emerson, 1976). In the higher education context, part of the academic’s job is to engage with students for the purpose of forming and maintaining positive and mutually satisfying relationships (DeShields Jr, Kara, & Kaynak, 2005; Voss, Gruber, & Szmigin, 2007). In this respect, it is important then to ensure that the relationship between the academic and the student yields equitable outcomes whereby what one party takes from the relationship and what they give to it is comparable to what the other party takes from the relationship and what they, in turn, give to it (Adams, 1963; Rawls, 2009).

The expectations of students and the engagement of students may underpin the academic-student relationship as independent constructs. Promises that induce expectations in the other party need to be fulfilled both as an indicator of valued social status for the other party and because such consistency may signal a commitment to equity as a norm for the relationship (Cropanzano, Byrne, Bobocel, & Rupp, 2001; Tyler & Blader, 2001). Independently of expectations induced through promises, an academic’s decision to devote extra-curricular teaching effort may depend on a student’s level of engagement as a more engaged student may be presumed to give more to the relationship with the academic than a less engaged student. In this alternative case, student expectations are also likely to be higher, the higher their level of engagement (Blau, 1964; Emerson, 1976). I develop these ideas below with reference to the phenomenon of guilt aversion. In the model, I discuss the effects of higher expectations on the part of the student (due to promises versus student engagement) only when the distinction is relevant to discuss possible interaction effects. I then manipulate expectations both ways to empirically test my theoretical predictions.

Guilt Aversion
Guilt aversion can be defined in the following terms: “if people feel guilt for hurting their partners…..and for failing to live up to their expectations, they will alter their behaviour (to avoid guilt) in ways that seem likely to maintain and strengthen the relationship (Baumeister et al., 1994). So, empathic (interpersonal) guilt can be seen as one mechanism for the interdependent exchanges described in social exchange theory. In making a promise, one party has intentionally induced the other party to rely on that behaviour. By breaking a promise the promisor disappoints the entitled expectation of the promisee to receive what is deserved, which is harmful to the relationship. Similarly, guilt may arise as a result of violating the principle of equity in the relationship when, for example, someone who gives more and someone who gives less to the relationship receives from it the same amount of the corresponding reward. In this sense, either student expectations or student engagement are likely to affect the behaviour of the academic independently of each other through the mechanism of guilt aversion (Battigalli & Dufwenberg, 2007; Charness & Dufwenberg, 2006). The application of guilt aversion to the academic-student relationship is a novel development. Increasing student expectations of extra-curricular teaching support in the neoliberal academy and the relationship that may exist between the two parties provide a setting that could give rise to guilt aversion. I define extra-curricular teaching effort as both the likelihood that an academic will respond to a student and how much time they will allocate if they respond (DellaVigna, List, & Malmendier, 2012).

The Model and Research Hypotheses

I adapt a formalisation of guilt aversion, which applies to theoretic games with financial payoffs (Battigalli & Dufwenberg, 2007). I modify the dictator game to represent the relationship between an academic and a student, with the academic deriving utility from having more time to finalize a research grant application. There is no retribution from the student affecting the academic’s payoff as the timeframe is set subsequent to the student
completing the teaching evaluation. In total, she has a time resource of $x$ units which she can allocate between working on her research grant application and helping a student who asks for support before his upcoming final exam.\(^1\) Both tasks have the same deadline creating a conflict. The basic idea of a guilt averse academic is that she will suffer from guilt to the extent that she believes the student receives less support than he expects to get. Thus she will increase her teaching effort to avert these feelings of guilt. A psychological utility function for the academic $\mu_A$ can be defined as:

$$
\mu_A(z, \alpha_S) = \pi_A(z) - \theta_1 \max \{0, E_{\alpha_S}[\pi_S] - \pi_S(z)\} - \theta_2 \max \{0, \pi_S(z) - E_{\alpha_S}[\pi_S]\} \quad (1)
$$

Where $z$ is the eventual time allocation of the academic to her research grant, $\pi_A(z)$ is the payoff (from the research grant application) to the academic at $z$, $\alpha_S$ is the student’s expectation of receiving help from the academic and $E_{\alpha_S}[\pi_S]$ is the student’s expected payoff (in this case the student’s final grade) given his beliefs of receiving help from the academic and $\theta_1$ and $\theta_2$ are exogenously given constants. The academic then maximises her utility subject to the constraint that $z \leq x$, where $x$ is the academic’s time resource.

Further discussions regarding the mathematical details of the model follow in the formalization of guilt aversion (Battigalli & Dufwenberg, 2007). Here I concentrate on the interpretation of (1) in the higher education context. Equation (1) says that in a conflict of interests, as in dictator games, the increase in the academic’s payoff from spending more time on the grant application [$\pi_A(z_2) - \pi_A(z_1)$] > 0 may be offset by the cost of her guilt due to the failure to meet the expectations of the student caused by his lower grade.

$$
\max \{0, E_{\alpha_S}[\pi_S] - \pi_S(z_2)\} - \max \{0, E_{\alpha_S}[\pi_S] - \pi_S(z_1)\} \geq 0 \quad (2)
$$

\(^1\) In this subsection I will refer to the academic as “she” and the student as “he”.

79
Where inequality (2) is strict if the student initially expects to get more help than $\pi_S (z_2)$. The extent of the psychological cost is given by $\theta_1$ and $\theta_2$ which measure the academic’s sensitivity to guilt by not meeting the student’s high or low expectations of support, respectively.

Note that student expectations may be high either due to explicit promises by the academic (direct induction of expectations) or because the student gives more (versus less) to the relationship in terms of his/her engagement (a prediction derived from social exchange theory (Blau, 1964; Emerson, 1976)). Thus, the model illustrates the effects on extra-curricular teaching effort that are possible as a function of either promise-induced expectations or student engagement. I summarize these as my first two hypotheses:

**Hypothesis 1: (Student Expectations)** There will be a positive association between student expectations and higher extra-curricular teaching efforts by academics.

**Hypothesis 2: (Student Engagement)** There will be a positive association between student engagement and higher extra-curricular teaching efforts by academics.

I further note that there may be interactive effects between student expectations and student engagement. That is academics may be more responsive to expectations when students are more engaged. This type of interaction has previously been theorized between a guilt averse response to another party’s expectations and the engagement of the other party in the relationship, measured by their effort exerted in response to differing incentives (Adams, 1963; Ghidoni & Ploner, 2014). Also, as previously discussed, high levels of engagement may reduce the social distance between the academic and the student, which has also been
shown to more likely give rise to guilt aversion when expectations are not met (Balafoutas & Sutter, 2016; Morell, 2016).

Thus,

**Hypothesis 3: (Student Expectation and Engagement)**. The positive relationship between student expectations induced through promises and extra-curricular teaching effort will be stronger when student engagement is high.

**The Role of Gender**

In this section, I discuss the role of the academic’s gender in changing the effects predicted in Hypotheses 1 to 3. As discussed in sociological research studies, female academics face different societal expectations for care-giving which may affect the roles female academics are expected to adopt (Beddoes, 2015; Heijstra et al., 2015; Rafnsdóttir & Heijstra, 2013). The neoliberalism of academia may reinforce traditional academic stereotypes of the male researcher, who is freer to meet workflow accountability constraints, and the caring female teacher (Lu, 2018; MacNell et al., 2015; Powell et al., 2018).

Gendered preferences associated with teaching effort may also be fashioned by social expectations. First, female academics, like other women, may be more other-regarding than men with a greater sensitivity to social cues (Croson & Gneezy, 2009; Kahn et al., 1971). Second, men may be more responsive to financial incentives and invest in types of activities that lead to valued financial rewards from their career, compared to women who may respond to social incentives more (Murad, Stavropoulou, & Cookson, 2016; Tonin & Vlassopoulos, 2014). Third, there may be female self-selection into roles and sectors with greater social motivation and affiliation (Konrad et al., 2000; Tonin & Vlassopoulos, 2010).
Women have long been gender stereotyped as more emotional, and more prone to feelings of guilt, than men (Barrett & Bliss-Moreau, 2009; Plant, Hyde, Keltner, & Devine, 2000). A meta-analysis of studies of self-reported guilt proneness found that women are more prone to feelings of guilt than men (Else-Quest et al., 2012). The gender difference was small but statistically significant. I also test for gender differences in guilt-proneness for academics:

**Hypothesis 4: (Gender and Guilt)** Female academics will tend to be more guilt-prone than male academics.

Following from hypothesis 4, if female academics are more prone to guilt than male academics I would expect this to change how Hypotheses 1 to 3 above apply to female versus male academics. In particular, I expect all hypothesized relationships to be stronger for female academics (as they are more prone to the key trigger of the effects I predict). This leads to the following hypothesis:

**Hypothesis 5: (Gender, Expectations and Engagement)** The positive relationship between extra-curricular teaching effort and student expectations of support, student engagement and the interaction of student expectations and engagement, will be stronger for female versus male academics.

**Study Design**

**Participants and Design:**

The data was gathered from an experiment embedded in a survey sent to 12,272 academics at UK Universities in July 2016, with 1,418 respondents (M age = 45.61, SD age =
9.83, 38.5% female, 44.3% at senior academic levels (associate or full professor or equivalent), 790 mainly lived in the UK). The survey experiment approach has an extensive history and was dictated by the impossibility of observing student expectations in a natural setting (Cruces, Perez-Truglia, & Tetaz, 2013; Di Tella, Galiani, & Schargrodsky, 2012; Finseraas & Jakobsson, 2014; Hainmueller & Hiscox, 2010; Jakobsson, Kotsadam, Syse, & Øien, 2015).

The study consisted of a 3 (expectations: high vs. low vs. unknown) x 3 (student engagement: high vs. low vs. unknown) between-subjects design, with participants randomly assigned to one of 9 vignettes representing each combination of treatments. By manipulating the student’s expectation and engagement conditions in my vignettes, I hoped to observe the hypothesized differences between the academics time allocations, i.e. both the likelihood of responding and the time allocated to the student. Full details of the 9 vignettes, together with a table showing the link between vignettes and hypotheses, follows at Appendix A.

**Student Expectations.** Participants were randomly allocated to one of three student expectation conditions: high, low or unknown. For high expectations of support, the wording was: “In your lectures, you have made a point of telling your students that you would be prepared to provide support for them during the exam period. As a result, your students' expectations of gaining support from you are high.” For low expectations of support, the wording was: “In your lectures, you have made a point that you might be involved in a bid process in the same period as the exam takes place. As a result, this specific student's expectation of receiving support from you before the exam is low.” For unknown student expectations, no statement describing student expectations was provided.

**Student Engagement.** Participants were randomly allocated to one of three student engagement conditions: high, low or unknown. For high/ (low) student engagement, the wording was: “The scenarios that follow relate to a specific student of yours who is very
engaged/ (not very engaged) with your module.” For unknown student engagement, no statement describing student engagement was provided.

To conceal the purpose of the survey experiment, it was embedded within a wide-ranging survey of academics preferences, productivity measures and other control measures, which have already been detailed in chapter two.

The number of academics polled for the study was determined based on an estimated response rate of 10% and on a-priori power analysis with anticipated small effect sizes (i.e., Cohen’s $f = .15$; (J. Cohen, 1992)) requiring a minimum sample size of 1,100 to be powered at 95%. All power calculations were conducted using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007).

Materials

In my experiment, formulated as a dictator game, the academic (the dictator) has to decide whether to allocate time and, if so, how much time to devote to a student (the recipient). Each vignette describes a multi-tasking conflict between a research grant application and supporting the student for an exam. Both the research grant application and the exam have identical deadlines, in one week’s time. It is made clear to the academic that helping the student will benefit the student’s exam result but may diminish the quality of the grant application, which is critical to the academic’s next review and that there is no penalty for withholding assistance to the student$^2$.

Procedure

$^2$ To ensure a true dictator game, a key assumption is that there can be no retribution by the student in the form of a poor teaching evaluation. I contacted all 27 universities, whose staff constituted the polling sample, to verify this assertion. Where no university wide policy applied, I contacted 3 departments within the university to determine their policies. Based on the replies received, in 92.3% of cases teaching evaluations were conducted more than one week before the end of teaching and 3-4 weeks before the exam. Thus my assertion that no retribution is possible is supported in the vast majority of cases.
Participants were randomly assigned to one of the nine experimental conditions, and read the corresponding vignette (see Appendix A). They proceeded to record their likelihood of responding to a student and the minutes they would devote to the student during the meeting, based on the vignette.

The survey experiment was conducted using the Qualtrics online survey tool. Invitations were sent to academics’ university e-mail accounts with access to the survey by unique individualised URL. The software enabled full response tracking and prevented the duplication of responses. For all questions that involved sensitive personal information respondents were able to either answer the question, in full or in part, or leave it blank if they do not want to submit these details. Ethical approval of the proposals and data security arrangements was granted by the University of Surrey Ethics Committee.

Twenty-seven of the top fifty UK universities, from the Times 2015 UK higher education ranking (Bothwell, 2015), were selected as having similar research and teaching characteristics. Common departments were chosen across these universities to allow further analysis of the data at a departmental level. The selection of academics was by title with the aim of identifying those with both research and teaching responsibilities. Responses were screened to verify these dual responsibilities. Further data was sought, with completion optional, to provide controls and additional insights into academics’ willingness to provide additional support to students.

**Measures**

**Extra-Curricular Teaching Effort.** Participants were asked to provide two measures of extra-curricular teaching effort. These were their likelihood of responding to a student, and

---

3 In the UK Education system the titles Assistant Professor, which is equivalent to a Lecturer or Senior Lecturer, and Associate Professor, which is equivalent to a Reader can be used.
if they responded, how many minutes they would devote to a student. Where the likelihood of responding to a student was entered as zero, any entries for minutes devoted to a student were removed from subsequent regressions. These measures are summarised below and detailed at Appendix A.

(a) **Likelihood of Responding to a Student.** Participants were asked to provide a percentage likelihood, on a scale of 0% to 100%, of responding to a student using two items. One item measured the likelihood of responding to a student email and the other item measured the likelihood of responding to a knock on the door (Spearman-Brown = .66).

(b) **Minutes Devoted to a Student.** Participants were asked how many minutes they would devote to responding to a student, on a scale of 0 minutes to 100 minutes, using 4 items for both an email approach and a knock on the door approach. A sample item was “This specific student of yours emails you (knocks on your door outside your office hours) with a significant number of questions. If you respond without any information that it will substantially affect the student’s grade (If, according to your student, your meeting could improve their overall grade from a 2.2 to a 2.1/ If, according to your student, your meeting could improve their overall grade from a 2.1 to a 1st/ If, according to your student, your meeting could avoid the student failing their degree), how many minutes would you dedicate to responding to your student?”. These items probed 4 impacts of the intervention on the student’s overall grade for exploratory purposes, as I speculated that they may influence extra-curricular teaching effort. However, given a Cronbach’s $\alpha = .90$ across the 8 measures (4 impacts x 2 approaches), they were consolidated to a single measure.
Control Variables. I controlled for the academic’s guilt proneness, gender, seniority, type of department, research preference, teaching preference, time preference, years post PhD, risk aversion and altruism.

Since I hypothesised that extra-curricular teaching effort would be motivated by guilt aversion, I investigated the effect of guilt proneness, using a psychometric index of guilt proneness (T. Cohen et al., 2014). The justification for controlling for gender, seniority and type of department was that each variable may make a difference to extra-curricular teaching effort, so reducing the residual variance of the regression and improving the precision of the estimates. I also elicited, and controlled for, academic’s validated teaching and research preferences (Callaghan & Coldwell, 2014; Matthews et al., 2014). A measure of time preference was also elicited and controlled for, given that allocating time to extra-curricular teaching effort may reflect a desire for short-term benefits over the long-term career benefits accruing from the research grant application (Burks et al., 2012). I controlled for years post PhD as a measure of experience. I controlled for the respondent’s degree of risk aversion which may influence decision-making. I also controlled for altruism, the principle of concern for other human beings which may be linked to helpfulness to students, by asking how much of a £500 win from a prize draw the respondent would donate to charity (Falk, Becker, Dohmen, Huffman, & Sunde, 2016). I ensured the integrity of controls by conducting a series of internal data validity checks.4

Incentives

---

4 Internal validity checks included the following regressions: (a) Household income on age and years post PhD (b) Years post PhD to seniority (c) Teaching hours on teaching percentage of academic’s time (d) Aggregated publications on research percentage of academic’s time (e) Teaching preference on teaching percentage (f) teaching preference on teaching productivity index (g) research preference on research percentage (h) Research preference on research productivity index. All regressions produced the anticipated positive coefficients significant at 1%.
To improve survey response rates a prize of £300 was awarded to the respondent most accurately predicting the behavioural norms of other academics towards teaching duties. This was calculated as the response that minimises the sum of absolute deviations from the mean of each response within each academic discipline. A prize of up to £80 was also awarded randomly, with a 1 in 300 chance of winning, as part of the end-of-survey questionnaire to elicit the time preferences of responders. The latter was aimed at studying an unrelated research question, which is the subject of chapter two. In addition to incentivising the completion of the survey, these awards were designed to make the questions they referred to more salient.

**Experimental Validity Checks**

Randomized economic experiments are commonly referred to as a benchmark, in terms of their empirical credibility, against which non-experimental studies can be evaluated (Angrist & Pischke, 2010). My use of a survey experiment to determine economic treatment effects follows established methodology (Cruces et al., 2013; Di Tella et al., 2012; Finseraas & Jakobsson, 2014; Hainmueller & Hiscox, 2010; Jakobsson et al., 2015). The survey was completed by academics actively engaged in both research and teaching in higher education lending weight to the effects obtained compared with results from a non-representative population (Mullinix, Leeper, Druckman, & Freese, 2015). Given the typically low response rates obtained, selection bias may have altered the magnitude of treatment effects (Jakobsson et al., 2015). However, non-response does not necessarily predict selection bias (Groves, 2006).

I followed a comparable higher education survey and recommended techniques to determine the robustness of my results (Abreu et al., 2009; Montaquila & Olson, 2012; Pedersen, 2015). Benchmarking to industry data (Higher Education Statistics Agency, 2017),
female academics made up 38.5% of responses to the survey compared to the HESA benchmark 37.3%. Respondents at professor level at 27.5% exceeded the HESA benchmark of 11.9%, mirroring a typical over-representation of senior academic levels in surveys (Abreu et al., 2009; Blackaby et al., 2005).

In my regressions, a teaching preference (positive), a research preference (negative) and guilt proneness (positive) are all correlated with extra-curricular teaching efforts. So, if the average values of these three controls varied between experimental conditions it could distort my results. However, I found only insignificant differences in the average value of these controls across the three experimental conditions, as shown at Table 1 (A). Repeating this comparison for different waves of respondents (to the initial email, and to the first and second reminders), did not show significant differences either, as shown at Table 1 (B).

I also compared respondents receiving incentivised emails, describing the incentives to complete the survey, or unincentivised emails which omitted this information. Again, if the average value of any variable correlated with time allocations varied between the incentivised and unincentivised emails it could distort my results. Those receiving the incentivised email had significantly higher research preferences, compared to those receiving the unincentivised email, as shown at Table 1 (C). If this difference was due to self-selection, based on the incentives offered, this would have increased the average research preference scores in my sample and reduced time allocated to students, due to the negative correlation between these variables. If the reduction was proportional to time allocated, then in line with hypothesis one, which proposed higher time allocations to students with high expectations of support, the reduction would be greater for the high expectations condition. Hence the treatment effects, the difference between high and unknown expectations, would be reduced but the findings would not be undermined.
Many studies have found no difference between responses in laboratory experiments compared to real-life due to self-selection or experimenter demand effects (Abeler & Nosenzo, 2015; Anderson et al., 2013; Cleave, Nikiforakis, & Slonim, 2013; Exadaktylos, Espin, & Branas-Garza, 2013; Falk, Meier, & Zehnder, 2013). However I employed a number of measures to guard against possible experimenter demand effects, responses prompted by cues about what constitutes acceptable behaviour (Zizzo, 2010). Firstly, the experiment was embedded within a lengthy series of survey questions to mitigate possible cues. Secondly, financial incentives in the survey did not relate to the experimental vignettes. The ordering of vignettes was also randomized to avoid spill-over effects. Thirdly, the online format avoided cues from the invigilator or other subjects in a laboratory setting. Finally, it can be argued that the social cues in the vignettes have real-world validity by mirroring the social cues experienced by the academics balancing the demands of research and teaching.

Results

Summary Statistics

A summary of the lecturer’s extra-curricular teaching effort, being the likelihood of responding (as a percentage) and time devoted (minutes and seconds), follows at Table 2, with effect sizes at Table 3. The results can be seen graphically at figure 1.

Insert Table 2 about here
The mean likelihood of the lecturer responding to a student is 67.88% and the mean amount of time devoted is 20 minutes 53 seconds. Comparing high to low expectations, the average likelihood of responding to a student increases by 3.02 percentage points, an effect size of .11, and the amount of time devoted increases by 2 minutes 29 seconds, an increase of 12.6% and an effect size of .20. Comparing high to low engagement, the average likelihood of responding to a student increases by 2.07 percentage points, an effect size of .08, and the amount of time devoted reduces by 11 seconds, an effect size of -.01. The relatively small effect sizes are typical of vignette experiments (Alexander & Becker, 1978; Atzmüller & Steiner, 2010). However, in real-world terms, the difference for high compared to low expectations of roughly 2 ½ minutes per student for 20 students, a larger tutorial group, equates to almost an hour of time devoted to students arising from one conflict, an economically significant figure (Gibbs, Lucas, & Simonite, 1996; Kingsbury & Lymn, 2008).

Turning to gender differences in time allocations, the average likelihood of responding to, and the minutes allocated to, the student for male and female academics individually can be seen graphically in Figure 2.
In the next section I analyse extra-curricular teaching effort motivated by student expectations and student engagement, covering the first three hypotheses, and gender differences in guilt-proneness and extra-curricular teaching effort, covering the final two hypotheses.

**Time Allocations motivated by Student Expectations and Student Engagement**

I used Tobit regressions to investigate the hypotheses, dictated by the likelihood of response and time donations being clumped at specific values and censored both at zero and at 100% and 100 minutes respectively. I show the results of the regression of the likelihood of response and time donations to students, with and without controls, in Table 4. Angrist and Pischke (2008) raise concerns over the approach used by the Tobit model. Recognising these concerns, I test for the robustness of the results that follow using an OLS model, in Table 5. There are no differences in the significance of high (compared to unknown) student expectations and high (compared to unknown) student engagement compared to Table 4.

Turning to Hypothesis 1 (Student Expectations), when students have high (compared to unknown) expectations of support from their lecturer, the lecturer is significantly more likely both to respond (β = 4.98, p < .05) and to devote additional time (β = 2.04, p < .01). I also performed post-estimation Wald tests to compare positive extra-curricular teaching effort between the high and low expectations of support. For the likelihood of response, the Wald test was significant at 10% and for the minutes devoted it was significant at 1%.

--------------------------------------------

Insert Table 4 about here

--------------------------------------------

Insert Table 5 about here
The regressions for extra-curricular teaching effort included controls for research preference, teaching preference and a measure of guilt proneness (T. Cohen et al., 2014), all which are significant ($p < .01$). The significance for the measure of guilt proneness is consistent with the proposed transmission effect for high student expectations through the desire to avert guilt. In summary, the Tobit regressions and Wald tests in Table 4 provide support for Hypothesis 1 (Student Expectations) for both the likelihood of responding and for minutes devoted.

Turning to Hypothesis 2 (Student Engagement), a high level of student engagement (compared to unknown student engagement) increases the likelihood of responding ($\beta = 5.50$, $p < .05$). However, the donation of additional time was not significant ($\beta = 0.43$, $ns$). The Wald tests in Table 4, comparing high engagement to low engagement, were not significant. Thus I have limited support for Hypothesis 2, with academics being more likely to respond to a student with high (compared to unknown) engagement.

Turning to hypothesis 3 (Student expectation and engagement), I hypothesised that the positive relationship between student expectations and extra-curricular teaching effort will be stronger when student engagement is high. However, the interaction between high student expectations and high student engagement was not significant for either the likelihood of responding ($\beta = -2.15$, $ns$) or for minutes devoted ($\beta = 0$, $ns$), as shown in model 3 in Table 4. I recoded the experimental conditions to further investigate these findings and for illustrative purposes. The recoding sought to identify those conditions most aligned with the theoretical prediction of high extra-curricular teaching effort using a technique devised to provide a holistic view of the alignment of characteristics associated with a positive career outcome (Foti & Hauenstein, 2007). The conditions were coded as aligned, where high student expectations were aligned to high student engagement; partially aligned, where either
student expectations or engagement was high but the other variable was low; and unaligned, where both student expectations and engagement were low. All other combinations of student expectations and engagement were recoded as ambiguous. Following this recoding, I show the Tobit regressions in Table 6.

---------------------------------------------------
Insert Table 6 about here
---------------------------------------------------

Only the aligned high expectations-high engagement condition is significant for both the likelihood of responding to the student ($\beta = 7.54, p < .01$), and the minutes devoted to the student ($\beta = 2.17, p < .05$). Other combinations of expectations and engagement were not significant for either the likelihood of responding to a student or for minutes devoted to the student. So, in summary I do not find support for hypothesis 3, that the relationship between high expectations and extra-curricular teaching effort will be stronger with high engagement but find that only the combination of high expectations and high engagement gives rise to extra-curricular teaching effort.

**Gender Differences in Guilt and Time Allocations**

To test hypothesis 4 (Gender and Guilt), whether guilt proneness differs by gender in the respondents, I performed a t-test. This showed that female academics were more guilt prone than male academics ($t = 3.56, p < .01$, Cohen’s D =-.19). This result provides support for hypothesis 4.

To test hypothesis 5 (Gender, Expectations and Engagement), I performed difference in differences tests for each of hypothesis 1 to 3 by gender, which follow in Table 7.

---------------------------------------------------
The interaction between high expectations and gender was not significant (see Table 7, Model 1). The interaction between high engagement and gender for minutes devoted was negative and significant (see Table 7, Model 2, $\beta = -3.96$, $p<0.05$). To ease the interpretation of gender differences, I examined the significance of the interaction terms between gender and aligned/partially aligned/unaligned coding of the experimental conditions. The interaction between aligned high expectations/high engagement and gender for minutes devoted was negative and weakly significant (see Table 7, Model 3, $\beta = -3.25$, $p<0.10$). Thus, I do not have support for hypothesis 5.

**General Discussion**

In this paper, I have subjected a large sample of academics in UK higher education to a multi-tasking vignette experiment. Over 60% of UK academics face both research and teaching demands on a regular basis so that studying a conflict between these demands is of interest (Higher Education Statistics Agency, 2018). In line with my hypothesis, I found that a high level of student expectations of support led to higher extra-curricular teaching effort. Also in line with my hypothesis, I found that academics were more likely to engage in extra-curricular teaching efforts, in terms of being more likely to respond but not to devote extra minutes, to help students whose engagement is high. Contrary to my hypothesis, I did not find the positive relationship between high expectations and extra-curricular teaching effort to be stronger with high student engagement. However, to analyse the interaction of student expectations and engagement, I recoded data for illustrative purposes and found that
academics will be more responsive when both student expectations of support and student engagement are high.

Whilst I found that female academics were more guilt prone than male academics, this was not reflected in greater extra-curricular teaching effort by female academics compared to male academics in response to student expectations and engagement. I discuss the theoretical implications of my findings below, together with the limitations of, and the future directions for, my research.

**Theoretical Implications**

My findings that high student expectations and high student engagement lead to extra-curricular teaching effort were in accordance with social exchange theory and the phenomenon of guilt aversion (Battigalli & Dufwenberg, 2007; Blau, 1964; Charness & Dufwenberg, 2006; Emerson, 1976).

I theorised that both student expectations and student engagement were likely to affect the behaviour of the academic through the mechanism of guilt aversion (Battigalli & Dufwenberg, 2007; Charness & Dufwenberg, 2006). A correlation between the size of monetary donations to another player and the perceived expectations of the amount to be received by the other player has been found in economic experiments and attributed to guilt aversion. In my model, guilt could have resulted from the academic anticipating breaking a promise or from violating the principle of equity in the academic student relationship.

In my model, I controlled for motivations of extra-curricular teaching effort other than guilt aversion, including preferences for research and teaching. By manipulating and conveying student expectations of support directly to academics in the vignette descriptions, the design addressed concerns of a false consensus of expectations (Ellingsen, Johannesson, Tjøtta, & Torsvik, 2010). My results were obtained across many universities and
departments, adding to the robustness of findings. However, the findings are mixed with regard to the mechanism for the extra-curricular teaching effort recorded. On the one hand, I control for individual differences in guilt proneness (T. Cohen et al., 2014), and show a significant correlation of guilt proneness with extra-curricular teaching effort. This provides support to the idea that guilt aversion may be the mechanism for the effects of student expectations and engagement on extra-curricular teaching effort. On the other hand, I find that female academics are more prone to guilt than male academics. Yet, there are no gender differences in how female versus male academics respond to student expectations of support and student engagement. This raises questions as to whether high student expectations and engagement generated feelings of guilt for the academics when they contemplated not engaging in the extra-curricular teaching effort. Perhaps, the mechanism for the effect of expectations and engagement did not involve the anticipation of, and aversion to, such potential feelings of guilt.

**Limitations and Future Directions**

I have actively sought to address potential limitations to this study. Firstly, there is a risk of self-selection into the survey and observer effects increasing the extra-curricular teaching effort recorded. To counter this risk, I adopted design and experimental validity checks, as previously detailed, to check for these biases. Secondly, the model assumes that academics will not respond to a conflict between research and teaching by simply working longer hours. The scenarios were carefully constructed to mirror that assumption and create a genuine conflict. To the extent that not all conflict may be resolved by working longer hours, my study is instructive regarding the trade-offs academics are willing to make and the role of student expectations in those trade-offs.

The model using Tobit regressions, for both the likelihood of responding to a student and the amount of time to be allocated, can be criticised, since a non-zero likelihood of
response is a pre-requisite for allocating time to a student. An alternative approach could be Heckman’s selection model with a two stage approach, firstly to responding, then to time allocated, since the regressions for time allocated were conducted only for those academics who responded (Heckman, 1979).

I did not plan to test for guilt as a mediating variable with a measure of state guilt, although my final theory development does build on guilt aversion as a mediating mechanism. It is a point of learning that a measure of state guilt could be used to test the mechanism more directly than a measure of guilt proneness as a trait. Employing a measure of state guilt and testing for a mediating role in extra-curricular teaching effort would be the subject of further research to clarify the effects found.

In chapter two, I noted that gender differences may exist in time usage and power over time and in volunteering for non-promotable administration and service tasks (Babcock et al., 2017; Rafnsdóttir & Heijstra, 2013). These gender differences may also affect the allocation of extra-curricular teaching effort. So, even if female academics feel more guilt than male academics in a specific situation, it may limit their practical ability to supply extra-curricular teaching effort (O'Hagan, 2018; Toffoletti & Starr, 2016).

At a more detailed level, the email sent out to academics, inviting them to take part in the survey, had several references to the “Delivering Better For Less” website. A number of recipients submitted complaints, having associated the name with punitive measures to increase productivity in higher education. Although the email was subsequently amended, with hindsight, this issue could have been anticipated, benefitting overall response rates. Further, the survey sent out to UK academics was too long, requiring a minimum of 15 to 20 minutes to complete. In retrospect, separate surveys could have been used, each closely aligned with the design of the corresponding study rather than trying to capture additional data possibly required in one long questionnaire that some of the respondents did not answer
in full. In mitigation, as in chapter two, data from partially completed surveys was used where appropriate.

An extension of the model within the higher education setting could be to study responsiveness to student expectations of support and student engagement specifically within hard science departments (compared to other departments). I define hard science departments as including the biological sciences, chemistry and physics. This is motivated by hard science departments tending to work in larger teams producing multi-authored papers (Jones, Wuchty, & Uzzi, 2008; Wuchty, Jones, & Uzzi, 2007). This, in turn, is likely to provide a greater incentive to work on site compared to other departments, more contact with students in laboratories and, potentially, a greater aversion to guilt should students be let down in terms of their expectations.

Finally, in response to the overall question posed by this thesis of whether gender plays a role in the handling of multiple demands in academia, I find that both male and female academics respond positively to student expectations of support and student engagement. However, unlike in chapter two, I do not find a significant difference in responses by gender.
### TABLE 1

**Experimental Validity Checks.**

(A) Differences in key variables across control and treatment groups

<table>
<thead>
<tr>
<th></th>
<th>High Expectations</th>
<th>Low Expectations</th>
<th>Unknown Expectations</th>
<th>High v Unknown</th>
<th>High v Low</th>
<th>Low v Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching Preference Index</td>
<td>2.2386</td>
<td>2.2486</td>
<td>2.2133</td>
<td>0.64</td>
<td>0.67</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>(0.6981)</td>
<td>(0.6781)</td>
<td>(0.6940)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Preference Index</td>
<td>3.0398</td>
<td>3.0399</td>
<td>3.0454</td>
<td>0.71</td>
<td>0.95</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>(0.6859)</td>
<td>(0.6685)</td>
<td>(0.6246)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guilt Index</td>
<td>3.9215</td>
<td>3.8938</td>
<td>3.8951</td>
<td>0.18</td>
<td>0.57</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>(0.6589)</td>
<td>(0.6655)</td>
<td>(0.6022)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard deviations in parentheses

(B) Differences in key variables across waves of responders

<table>
<thead>
<tr>
<th></th>
<th>Initial Email</th>
<th>R1 (1st Reminder)</th>
<th>R2 (2nd Reminder)</th>
<th>R1 v Initial</th>
<th>R2 v Initial</th>
<th>R2 v R1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching Preference Index</td>
<td>2.1351</td>
<td>2.2486</td>
<td>2.2613</td>
<td>0.07</td>
<td>0.03</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>(0.6996)</td>
<td>(0.6913)</td>
<td>(0.6831)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Preference Index</td>
<td>3.0498</td>
<td>2.9827</td>
<td>3.0852</td>
<td>0.60</td>
<td>0.38</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(0.6060)</td>
<td>(0.7474)</td>
<td>(0.6021)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guilt Index</td>
<td>3.8767</td>
<td>3.8673</td>
<td>3.9444</td>
<td>0.90</td>
<td>0.38</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>(0.6657)</td>
<td>(0.7001)</td>
<td>(0.5797)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard deviations in parentheses

(C) Differences in key variables by email incentive/no incentive

<table>
<thead>
<tr>
<th></th>
<th>Incentivised Email</th>
<th>Unincentivised Email</th>
<th>Mann Whitney p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching Preference Index</td>
<td>2.1750</td>
<td>2.2140</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>(0.6755)</td>
<td>(0.6845)</td>
<td></td>
</tr>
<tr>
<td>Research Preference Index</td>
<td>3.0970</td>
<td>2.9330</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>0.6555</td>
<td>0.72105</td>
<td></td>
</tr>
<tr>
<td>Guilt Index</td>
<td>3.9125</td>
<td>3.8295</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>(0.6092)</td>
<td>(0.6958)</td>
<td></td>
</tr>
</tbody>
</table>

Standard deviations in parentheses
<table>
<thead>
<tr>
<th></th>
<th>Likelihood of Response %</th>
<th>Time Devoted:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall</strong></td>
<td>67.88%</td>
<td>20m 53s</td>
</tr>
<tr>
<td></td>
<td>(26.67)</td>
<td>(12.72)</td>
</tr>
<tr>
<td><strong>By Expectations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>70.22%</td>
<td>22m 12s</td>
</tr>
<tr>
<td></td>
<td>(25.02)</td>
<td>(13.46)</td>
</tr>
<tr>
<td>Low</td>
<td>67.20%</td>
<td>19m 43s</td>
</tr>
<tr>
<td></td>
<td>(28.60)</td>
<td>(11.56)</td>
</tr>
<tr>
<td>Unknown</td>
<td>66.08%</td>
<td>20m 34s</td>
</tr>
<tr>
<td></td>
<td>(26.35)</td>
<td>(12.81)</td>
</tr>
<tr>
<td><strong>By Engagement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>69.15%</td>
<td>20m 53s</td>
</tr>
<tr>
<td></td>
<td>(26.01)</td>
<td>(12.51)</td>
</tr>
<tr>
<td>Low</td>
<td>67.08%</td>
<td>21m 4s</td>
</tr>
<tr>
<td></td>
<td>(27.62)</td>
<td>(13.24)</td>
</tr>
<tr>
<td>Unknown</td>
<td>67.38%</td>
<td>20m 43s</td>
</tr>
<tr>
<td></td>
<td>(26.39)</td>
<td>(12.43)</td>
</tr>
<tr>
<td><strong>By Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>67.28%</td>
<td>20m 41s</td>
</tr>
<tr>
<td></td>
<td>(26.85)</td>
<td>(12.46)</td>
</tr>
<tr>
<td>Male</td>
<td>68.30%</td>
<td>21m 6s</td>
</tr>
<tr>
<td></td>
<td>(26.46)</td>
<td>(12.92)</td>
</tr>
</tbody>
</table>

Note: Each row shows the average percentage likelihood of responding to a student and the average time devoted. Standard deviations appear in brackets.
TABLE 3

Effect sizes (Cohen's $D$) for academics' likelihood of response and time donations

<table>
<thead>
<tr>
<th>Effect Size - Cohen's D</th>
<th>Likelihood of Response</th>
<th>Time Devoted</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Expectations : Low Expectations</td>
<td>.11</td>
<td>.20</td>
</tr>
<tr>
<td>High Engagement: Low Engagement</td>
<td>.08</td>
<td>-.01</td>
</tr>
<tr>
<td>Male Academic : Female Academic</td>
<td>.04</td>
<td>.03</td>
</tr>
</tbody>
</table>

Note: Each row shows Cohen’s d, a measure of effect size in standard deviations between two data samples. Negative signs indicate that the second term in the comparison is greater than the first term.
<table>
<thead>
<tr>
<th></th>
<th>Model 1 Likelihood</th>
<th>Model 1 Minutes</th>
<th>Model 2 Likelihood</th>
<th>Model 2 Minutes</th>
<th>Model 3 Likelihood</th>
<th>Model 3 Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low Expectations</strong></td>
<td>2.25</td>
<td>-0.80</td>
<td>0.94</td>
<td>-1.04</td>
<td>8.23*</td>
<td>-1.38</td>
</tr>
<tr>
<td></td>
<td>(2.16)</td>
<td>(0.71)</td>
<td>(2.33)</td>
<td>(0.78)</td>
<td>(4.02)</td>
<td>(1.34)</td>
</tr>
<tr>
<td><strong>High Expectations</strong></td>
<td>6.70***</td>
<td>1.78***</td>
<td>4.98**</td>
<td>2.04***</td>
<td>4.44</td>
<td>2.09*</td>
</tr>
<tr>
<td></td>
<td>(2.08)</td>
<td>(0.69)</td>
<td>(2.24)</td>
<td>(0.74)</td>
<td>(3.75)</td>
<td>(1.26)</td>
</tr>
<tr>
<td><strong>Low Engagement</strong></td>
<td>1.13</td>
<td>0.23</td>
<td>1.44</td>
<td>0.06</td>
<td>2.68</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>(2.14)</td>
<td>(0.70)</td>
<td>(2.31)</td>
<td>(0.77)</td>
<td>(3.87)</td>
<td>(1.30)</td>
</tr>
<tr>
<td><strong>High Engagement</strong></td>
<td>2.98</td>
<td>0.04</td>
<td>5.50**</td>
<td>0.43</td>
<td>9.92***</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>(2.10)</td>
<td>(0.69)</td>
<td>(2.26)</td>
<td>(0.75)</td>
<td>(3.78)</td>
<td>(1.26)</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td>-0.26</td>
<td>-0.42</td>
<td>-0.68</td>
<td>-0.85</td>
<td>-0.40</td>
<td>-0.86</td>
</tr>
<tr>
<td></td>
<td>(1.79)</td>
<td>(0.59)</td>
<td>(2.12)</td>
<td>(0.70)</td>
<td>(2.12)</td>
<td>(0.70)</td>
</tr>
<tr>
<td><strong>Low Expect x Low Eng</strong></td>
<td>-8.77</td>
<td>0.66</td>
<td>(5.75)</td>
<td>(1.92)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Low Expect x High Eng</strong></td>
<td>-12.74**</td>
<td>0.36</td>
<td>(5.69)</td>
<td>(1.89)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High Expect x Low Eng</strong></td>
<td>3.83</td>
<td>-0.14</td>
<td>(5.58)</td>
<td>(1.86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High Expect x High Eng</strong></td>
<td>-2.15</td>
<td>0.00</td>
<td>(5.35)</td>
<td>(1.79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Research Preference</strong></td>
<td>-6.47***</td>
<td>-2.12***</td>
<td>-6.57***</td>
<td>-2.11***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.55)</td>
<td>(0.51)</td>
<td>(1.55)</td>
<td>(0.51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Teaching Preference</strong></td>
<td>7.50***</td>
<td>1.11*</td>
<td>7.44***</td>
<td>1.11*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.44)</td>
<td>(0.48)</td>
<td>(1.44)</td>
<td>(0.48)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time Preference</strong></td>
<td>1210.72</td>
<td>-411.76</td>
<td>1136.17</td>
<td>-408.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(744.30)</td>
<td>(247.35)</td>
<td>(748.14)</td>
<td>(249.15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Guilt Proneness</strong></td>
<td>5.33**</td>
<td>2.00***</td>
<td>5.39***</td>
<td>2.00***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.62)</td>
<td>(0.54)</td>
<td>(1.62)</td>
<td>(0.54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Years Post PHD</strong></td>
<td>0.14</td>
<td>0.02</td>
<td>0.14</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.04)</td>
<td>(0.14)</td>
<td>(0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Risk Aversion</strong></td>
<td>-0.05</td>
<td>0.26</td>
<td>0.09</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(0.33)</td>
<td>(1.00)</td>
<td>(0.33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Altruism</strong></td>
<td>-0.70</td>
<td>-0.05</td>
<td>-0.69</td>
<td>-0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.42)</td>
<td>(0.14)</td>
<td>(0.42)</td>
<td>(0.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bath</strong></td>
<td>9.07</td>
<td>-2.68</td>
<td>10.04</td>
<td>-2.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.87)</td>
<td>(2.64)</td>
<td>(7.88)</td>
<td>(2.65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Birmingham</strong></td>
<td>9.23</td>
<td>-3.88</td>
<td>9.81</td>
<td>-3.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.32)</td>
<td>(2.45)</td>
<td>(7.32)</td>
<td>(2.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bristol</strong></td>
<td>-0.11</td>
<td>-3.82</td>
<td>1.53</td>
<td>-3.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.40)</td>
<td>(2.49)</td>
<td>(7.41)</td>
<td>(2.50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cambridge</strong></td>
<td>8.41</td>
<td>-1.48</td>
<td>8.97</td>
<td>-1.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.80)</td>
<td>(2.58)</td>
<td>(7.80)</td>
<td>(2.59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University</td>
<td>3.55</td>
<td>-3.26</td>
<td>4.15</td>
<td>-3.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiff</td>
<td>(6.42)</td>
<td>(2.15)</td>
<td>(6.42)</td>
<td>(2.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edinburgh</td>
<td>9.04</td>
<td>-3.45</td>
<td>10.22</td>
<td>-3.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Essex</td>
<td>0.50</td>
<td>-4.07</td>
<td>1.83</td>
<td>-4.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imperial College London</td>
<td>9.29</td>
<td>5.59</td>
<td>11.42</td>
<td>5.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kent</td>
<td>6.29</td>
<td>-4.77</td>
<td>7.36</td>
<td>-4.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leeds</td>
<td>10.08</td>
<td>-4.34</td>
<td>11.66</td>
<td>-4.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leicester</td>
<td>12.79</td>
<td>1.68</td>
<td>13.75</td>
<td>1.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSE</td>
<td>4.80</td>
<td>-6.09*</td>
<td>6.11</td>
<td>-6.11*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loughborough</td>
<td>17.87*</td>
<td>0.29</td>
<td>18.92*</td>
<td>0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manchester</td>
<td>11.55</td>
<td>-5.94*</td>
<td>12.74</td>
<td>-5.99*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nottingham</td>
<td>5.29</td>
<td>-2.27</td>
<td>5.73</td>
<td>-2.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxford</td>
<td>10.08</td>
<td>0.40</td>
<td>11.38</td>
<td>0.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Queens Belfast</td>
<td>12.01</td>
<td>-4.93*</td>
<td>12.94</td>
<td>-4.96*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>14.29</td>
<td>2.34</td>
<td>15.83*</td>
<td>2.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Royal Holloway College</td>
<td>5.07</td>
<td>-6.63*</td>
<td>6.63</td>
<td>-6.70*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southampton</td>
<td>4.12</td>
<td>-2.68</td>
<td>4.33</td>
<td>-2.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>St Andrews</td>
<td>16.04*</td>
<td>-5.89*</td>
<td>17.07*</td>
<td>-5.92*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surrey</td>
<td>25.38***</td>
<td>-0.97</td>
<td>26.39***</td>
<td>-1.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCL</td>
<td>11.95</td>
<td>-3.84</td>
<td>13.22</td>
<td>-3.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UEA</td>
<td>20.66**</td>
<td>-1.44</td>
<td>21.34**</td>
<td>-1.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warwick</td>
<td>2.35</td>
<td>-5.41*</td>
<td>2.43</td>
<td>-5.41*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>York</td>
<td>15.28*</td>
<td>-4.01</td>
<td>16.44*</td>
<td>-4.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.27)</td>
<td>(2.42)</td>
<td>(7.28)</td>
<td>(2.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discipline</td>
<td>Mean 1</td>
<td>StdDev 1</td>
<td>Mean 2</td>
<td>StdDev 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------</td>
<td>----------</td>
<td>---------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>8.48</td>
<td>0.57</td>
<td>7.54</td>
<td>0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.77)</td>
<td>(1.90)</td>
<td>(5.77)</td>
<td>(1.90)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business/Management</td>
<td>2.71</td>
<td>0.17</td>
<td>2.54</td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.14)</td>
<td>(1.38)</td>
<td>(4.14)</td>
<td>(1.38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemistry</td>
<td>-4.98</td>
<td>1.87</td>
<td>-5.02</td>
<td>1.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.92)</td>
<td>(1.63)</td>
<td>(4.91)</td>
<td>(1.63)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Science</td>
<td>-2.05</td>
<td>0.05</td>
<td>-2.86</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.02)</td>
<td>(1.67)</td>
<td>(5.02)</td>
<td>(1.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economics</td>
<td>-7.32</td>
<td>-1.07</td>
<td>-7.61</td>
<td>-1.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.74)</td>
<td>(1.59)</td>
<td>(4.75)</td>
<td>(1.59)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geography</td>
<td>6.32</td>
<td>-1.19</td>
<td>6.39</td>
<td>-1.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.09)</td>
<td>(1.68)</td>
<td>(5.10)</td>
<td>(1.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>History</td>
<td>15.49**</td>
<td>-1.97</td>
<td>15.74**</td>
<td>-1.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.90)</td>
<td>(1.60)</td>
<td>(4.90)</td>
<td>(1.60)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Law</td>
<td>2.91</td>
<td>0.42</td>
<td>3.19</td>
<td>0.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.81)</td>
<td>(1.60)</td>
<td>(4.83)</td>
<td>(1.61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematics</td>
<td>1.10</td>
<td>1.00</td>
<td>1.27</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.46)</td>
<td>(1.49)</td>
<td>(4.46)</td>
<td>(1.49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physics</td>
<td>2.92</td>
<td>1.29</td>
<td>2.92</td>
<td>1.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.57)</td>
<td>(1.52)</td>
<td>(4.57)</td>
<td>(1.52)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Politics/Int. relations</td>
<td>-4.12</td>
<td>-3.32*</td>
<td>-4.40</td>
<td>-3.34*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.86)</td>
<td>(1.63)</td>
<td>(4.86)</td>
<td>(1.63)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychology</td>
<td>-2.20</td>
<td>-0.50</td>
<td>-2.06</td>
<td>-0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.45)</td>
<td>(1.49)</td>
<td>(4.44)</td>
<td>(1.49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sociology</td>
<td>-2.01</td>
<td>-0.56</td>
<td>-1.74</td>
<td>-0.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.53)</td>
<td>(1.86)</td>
<td>(5.52)</td>
<td>(1.86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>3.83</td>
<td>3.13</td>
<td>4.08</td>
<td>3.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.15)</td>
<td>(1.72)</td>
<td>(5.15)</td>
<td>(1.72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Associate Professor</td>
<td>-3.85</td>
<td>-1.58</td>
<td>-4.02</td>
<td>-1.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.48)</td>
<td>(1.49)</td>
<td>(4.48)</td>
<td>(1.49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reader</td>
<td>5.92</td>
<td>0.49</td>
<td>5.49</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.86)</td>
<td>(1.27)</td>
<td>(3.85)</td>
<td>(1.27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assistant professor</td>
<td>-0.29</td>
<td>-1.16</td>
<td>-0.19</td>
<td>-1.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.83)</td>
<td>(1.95)</td>
<td>(5.82)</td>
<td>(1.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior Lecturer</td>
<td>-3.75</td>
<td>-0.07</td>
<td>-3.90</td>
<td>-0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.23)</td>
<td>(1.07)</td>
<td>(3.23)</td>
<td>(1.07)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecturer</td>
<td>-2.79</td>
<td>1.29</td>
<td>-3.14</td>
<td>1.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.42)</td>
<td>(1.13)</td>
<td>(3.42)</td>
<td>(1.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>-4.87</td>
<td>-2.55</td>
<td>-5.31</td>
<td>-2.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(10.65)</td>
<td>(3.53)</td>
<td>(10.63)</td>
<td>(3.53)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continental Europe</td>
<td>-1.60</td>
<td>0.35</td>
<td>-1.48</td>
<td>0.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.50)</td>
<td>(0.83)</td>
<td>(2.50)</td>
<td>(0.83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>13.76*</td>
<td>3.11</td>
<td>14.21*</td>
<td>3.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.67)</td>
<td>(2.18)</td>
<td>(6.67)</td>
<td>(2.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>Coefficient</td>
<td>Standard Error</td>
<td>Coefficient</td>
<td>Standard Error</td>
<td>Coefficient</td>
<td>Standard Error</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>----------------</td>
<td>-------------</td>
<td>----------------</td>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(12.06)</td>
<td></td>
<td>(4.06)</td>
<td></td>
<td>(12.06)</td>
</tr>
<tr>
<td>Africa</td>
<td>-9.40</td>
<td>5.17</td>
<td>-9.75</td>
<td>5.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(12.06)</td>
<td>(4.06)</td>
<td>(12.06)</td>
<td>(4.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>2.94</td>
<td>-0.66</td>
<td>3.02</td>
<td>-0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.99)</td>
<td>(1.33)</td>
<td>(4.00)</td>
<td>(1.34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South America</td>
<td>-8.30</td>
<td>-0.91</td>
<td>-8.42</td>
<td>-0.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(10.11)</td>
<td>(3.45)</td>
<td>(10.10)</td>
<td>(3.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australasia</td>
<td>-1.58</td>
<td>1.91</td>
<td>-1.86</td>
<td>1.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.67)</td>
<td>(2.20)</td>
<td>(6.67)</td>
<td>(2.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>70.39***</td>
<td>20.45***</td>
<td>-1165.29</td>
<td>429.56</td>
<td>-1093.67</td>
<td>426.23</td>
</tr>
<tr>
<td></td>
<td>(1.99)</td>
<td>(0.66)</td>
<td>(743.72)</td>
<td>(247.16)</td>
<td>(747.61)</td>
<td>(248.98)</td>
</tr>
<tr>
<td>Controls</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Observations</td>
<td>2,802</td>
<td>2,802</td>
<td>2,350</td>
<td>2,350</td>
<td>2,350</td>
<td>2,350</td>
</tr>
<tr>
<td>Pseudo R-squared</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Wald Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Hi – Lo expectations)</td>
<td>0.04</td>
<td>0.00</td>
<td>0.09</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Hi – Lo engagement)</td>
<td>0.39</td>
<td>0.79</td>
<td>0.80</td>
<td>0.63</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Model 1 = without controls, Model 2 = with controls, Model 3 = with controls and interaction effects. The Tobit regressions shown are for the percentage likelihood of responding to a student and the time devoted in minutes. The coefficients for high and low expectations are relative to the control treatment where student expectations are unknown to the academic. Robust standard errors are shown in parentheses *** p<0.01, ** p<0.05, * p<0.1
<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Model 1 Likelihood</th>
<th>Model 1 Minutes</th>
<th>Model 2 Likelihood</th>
<th>Model 2 Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Expectations</td>
<td>2.25</td>
<td>-0.80</td>
<td>-0.03</td>
<td>-0.99</td>
</tr>
<tr>
<td></td>
<td>(2.16)</td>
<td>(0.71)</td>
<td>(1.68)</td>
<td>(0.76)</td>
</tr>
<tr>
<td>High Expectations</td>
<td>6.70***</td>
<td>1.78***</td>
<td>3.72***</td>
<td>2.04***</td>
</tr>
<tr>
<td></td>
<td>(2.08)</td>
<td>(0.69)</td>
<td>(1.62)</td>
<td>(0.73)</td>
</tr>
<tr>
<td>Low Engagement</td>
<td>1.13</td>
<td>0.23</td>
<td>0.05</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>(2.14)</td>
<td>(0.70)</td>
<td>(1.67)</td>
<td>(0.76)</td>
</tr>
<tr>
<td>High Engagement</td>
<td>2.98</td>
<td>0.04</td>
<td>3.89**</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>(2.10)</td>
<td>(0.69)</td>
<td>(1.63)</td>
<td>(0.74)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.26</td>
<td>-0.42</td>
<td>-1.61</td>
<td>-0.93</td>
</tr>
<tr>
<td></td>
<td>(1.79)</td>
<td>(0.59)</td>
<td>(1.52)</td>
<td>(0.69)</td>
</tr>
<tr>
<td>Research Preference</td>
<td>-4.70***</td>
<td>-2.08***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.11)</td>
<td>(0.50)</td>
<td>(1.03)</td>
<td>(0.47)</td>
</tr>
<tr>
<td>Teaching Preference</td>
<td>5.40***</td>
<td>0.93***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.03)</td>
<td>(0.47)</td>
<td>(1.03)</td>
<td>(0.47)</td>
</tr>
<tr>
<td>Time Preference</td>
<td>722.53</td>
<td>-415.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(536.69)</td>
<td>(243.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guilt Proneness</td>
<td>3.93***</td>
<td>2.07***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.17)</td>
<td>(0.53)</td>
<td>(1.07)</td>
<td>(0.47)</td>
</tr>
<tr>
<td>Years Post PHD</td>
<td>0.11</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.04)</td>
<td>(1.07)</td>
<td>(0.47)</td>
</tr>
<tr>
<td>Risk Aversion</td>
<td>-0.39</td>
<td>0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.72)</td>
<td>(0.33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altruism</td>
<td>-0.31</td>
<td>-0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td>(0.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bath</td>
<td>7.20</td>
<td>-2.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.73)</td>
<td>(2.60)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birmingham</td>
<td>7.72</td>
<td>-3.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.30)</td>
<td>(2.41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bristol</td>
<td>2.55</td>
<td>-3.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.40)</td>
<td>(2.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cambridge</td>
<td>5.97</td>
<td>-1.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.60)</td>
<td>(2.54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiff</td>
<td>3.13</td>
<td>-3.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.67)</td>
<td>(2.12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edinburgh</td>
<td>7.18</td>
<td>-3.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.03)</td>
<td>(2.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Essex</td>
<td>0.37</td>
<td>-4.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.32)</td>
<td>(2.42)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imperial College London</td>
<td>7.34</td>
<td>5.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University</td>
<td>Grade</td>
<td>Score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kent</td>
<td>4.76</td>
<td>8.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leeds</td>
<td>7.04</td>
<td>8.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leicester</td>
<td>9.24</td>
<td>10.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSE</td>
<td>2.79</td>
<td>6.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loughborough</td>
<td>15.02</td>
<td>20.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manchester</td>
<td>9.60</td>
<td>13.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nottingham</td>
<td>4.67</td>
<td>5.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxford</td>
<td>9.18</td>
<td>14.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Queens Belfast</td>
<td>9.66</td>
<td>17.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>9.92</td>
<td>14.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Royal Holloway College</td>
<td>3.15</td>
<td>6.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southampton</td>
<td>5.69</td>
<td>7.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>St Andrews</td>
<td>10.97</td>
<td>16.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surrey</td>
<td>18.35</td>
<td>26.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCL</td>
<td>8.13</td>
<td>11.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UEA</td>
<td>13.75</td>
<td>18.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Warwick</td>
<td>1.87</td>
<td>3.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>York</td>
<td>11.09</td>
<td>14.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>6.42</td>
<td>10.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business/Management</td>
<td>2.26</td>
<td>5.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemistry</td>
<td>4.27</td>
<td>7.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Science</td>
<td>-2.64</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discipline</td>
<td>Mean</td>
<td>Median</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>------</td>
<td>--------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economics</td>
<td>3.63</td>
<td>1.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geography</td>
<td>3.79</td>
<td>1.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>History</td>
<td>9.40</td>
<td>-1.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Law</td>
<td>1.12</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematics</td>
<td>1.26</td>
<td>1.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physics</td>
<td>2.30</td>
<td>1.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Politics/International relations</td>
<td>-3.13</td>
<td>-2.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychology</td>
<td>-0.88</td>
<td>-0.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sociology</td>
<td>-0.58</td>
<td>-0.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>4.31</td>
<td>3.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Associate Professor</td>
<td>-3.24</td>
<td>-1.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reader</td>
<td>3.51</td>
<td>0.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assistant professor</td>
<td>-1.08</td>
<td>-0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior Lecturer</td>
<td>-2.72</td>
<td>-0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecturer</td>
<td>-1.92</td>
<td>1.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>-2.24</td>
<td>-2.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continental Europe</td>
<td>-1.24</td>
<td>0.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>8.24</td>
<td>2.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa</td>
<td>-7.07</td>
<td>4.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>2.37</td>
<td>-0.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South America</td>
<td>-0.93</td>
<td>-0.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australasia</td>
<td>-3.00</td>
<td>2.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The regressions shown in Table 3 have been re-estimated in this table using an OLS regression as a comparator. Model 1 = without controls, Model 2 = with controls, Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>70.39***</td>
<td>1.99</td>
<td>35.74</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>20.45***</td>
<td>0.64</td>
<td>32.10</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>-675.13</td>
<td>536.28</td>
<td>-1.26</td>
<td>0.208</td>
</tr>
<tr>
<td></td>
<td>433.51*</td>
<td>243.59</td>
<td>1.77</td>
<td>0.079</td>
</tr>
<tr>
<td>Observations</td>
<td>2,802</td>
<td>0.01</td>
<td>242.65</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.01</td>
<td>0.01</td>
<td>0.08</td>
<td>0.672</td>
</tr>
<tr>
<td></td>
<td>2,350</td>
<td>0.06</td>
<td>0.06</td>
<td>0.712</td>
</tr>
</tbody>
</table>
TABLE 6
Tobit Regressions of response for Expectations-Engagement Conditions (Note 1)

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>Likelihood</th>
<th>Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expectations-Engagement aligned</td>
<td>7.54***</td>
<td>2.17**</td>
</tr>
<tr>
<td></td>
<td>(2.92)</td>
<td>(0.97)</td>
</tr>
<tr>
<td>Expectations-Engagement partially aligned</td>
<td>3.36</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>(2.42)</td>
<td>(0.80)</td>
</tr>
<tr>
<td>Expectations-Engagement unaligned</td>
<td>-2.71</td>
<td>-1.04</td>
</tr>
<tr>
<td></td>
<td>(3.18)</td>
<td>(1.06)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.55</td>
<td>-0.81</td>
</tr>
<tr>
<td></td>
<td>(2.12)</td>
<td>(0.70)</td>
</tr>
<tr>
<td>Research Preference</td>
<td>-6.42***</td>
<td>-2.09***</td>
</tr>
<tr>
<td></td>
<td>(1.55)</td>
<td>(0.51)</td>
</tr>
<tr>
<td>Teaching Preference</td>
<td>7.50***</td>
<td>1.15**</td>
</tr>
<tr>
<td></td>
<td>(1.44)</td>
<td>(0.48)</td>
</tr>
<tr>
<td>Time Preference</td>
<td>1231.71</td>
<td>-372.66</td>
</tr>
<tr>
<td></td>
<td>(747.36)</td>
<td>(248.77)</td>
</tr>
<tr>
<td>Guilt Proneness</td>
<td>5.44***</td>
<td>2.06***</td>
</tr>
<tr>
<td></td>
<td>(1.62)</td>
<td>(0.54)</td>
</tr>
<tr>
<td>Years Post PHD</td>
<td>0.15</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Risk Aversion</td>
<td>-0.05</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>Altruism</td>
<td>-0.68</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>(0.43)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Bath</td>
<td>8.88</td>
<td>-2.79</td>
</tr>
<tr>
<td></td>
<td>(7.89)</td>
<td>(2.65)</td>
</tr>
<tr>
<td>Birmingham</td>
<td>8.82</td>
<td>-3.97</td>
</tr>
<tr>
<td></td>
<td>(7.33)</td>
<td>(2.45)</td>
</tr>
<tr>
<td>Bristol</td>
<td>-0.45</td>
<td>-4.12</td>
</tr>
<tr>
<td></td>
<td>(7.40)</td>
<td>(2.49)</td>
</tr>
<tr>
<td>Cambridge</td>
<td>7.99</td>
<td>-1.36</td>
</tr>
<tr>
<td></td>
<td>(7.80)</td>
<td>(2.59)</td>
</tr>
<tr>
<td>Cardiff</td>
<td>3.23</td>
<td>-3.26</td>
</tr>
<tr>
<td></td>
<td>(6.42)</td>
<td>(2.16)</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>8.66</td>
<td>-3.48</td>
</tr>
<tr>
<td></td>
<td>(6.92)</td>
<td>(2.32)</td>
</tr>
<tr>
<td>Essex</td>
<td>0.73</td>
<td>-3.74</td>
</tr>
<tr>
<td></td>
<td>(7.28)</td>
<td>(2.45)</td>
</tr>
<tr>
<td>Imperial College London</td>
<td>8.00</td>
<td>4.86</td>
</tr>
<tr>
<td></td>
<td>(11.49)</td>
<td>(3.85)</td>
</tr>
<tr>
<td>Kent</td>
<td>6.20</td>
<td>-4.67</td>
</tr>
<tr>
<td></td>
<td>(7.45)</td>
<td>(2.50)</td>
</tr>
<tr>
<td>Leeds</td>
<td>9.56</td>
<td>-4.52</td>
</tr>
</tbody>
</table>

111
<table>
<thead>
<tr>
<th>University</th>
<th>Average Score</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leicester</td>
<td>(7.12)</td>
<td>(2.38)</td>
</tr>
<tr>
<td></td>
<td>12.97</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>(7.20)</td>
<td>(2.40)</td>
</tr>
<tr>
<td>LSE</td>
<td>4.00</td>
<td>-6.43*</td>
</tr>
<tr>
<td></td>
<td>(8.46)</td>
<td>(2.82)</td>
</tr>
<tr>
<td>Loughborough</td>
<td>17.72*</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>(7.65)</td>
<td>(2.56)</td>
</tr>
<tr>
<td>Manchester</td>
<td>11.62</td>
<td>-5.96*</td>
</tr>
<tr>
<td></td>
<td>(7.09)</td>
<td>(2.37)</td>
</tr>
<tr>
<td>Nottingham</td>
<td>5.14</td>
<td>-2.41</td>
</tr>
<tr>
<td></td>
<td>(7.17)</td>
<td>(2.40)</td>
</tr>
<tr>
<td>Oxford</td>
<td>10.06</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>(7.83)</td>
<td>(2.62)</td>
</tr>
<tr>
<td>Queens Belfast</td>
<td>11.80</td>
<td>-5.10*</td>
</tr>
<tr>
<td></td>
<td>(7.13)</td>
<td>(2.38)</td>
</tr>
<tr>
<td>Reading</td>
<td>14.47</td>
<td>2.10</td>
</tr>
<tr>
<td></td>
<td>(8.06)</td>
<td>(2.66)</td>
</tr>
<tr>
<td>Royal Holloway College</td>
<td>4.89</td>
<td>-6.80**</td>
</tr>
<tr>
<td></td>
<td>(7.87)</td>
<td>(2.63)</td>
</tr>
<tr>
<td>Southampton</td>
<td>4.04</td>
<td>-2.97</td>
</tr>
<tr>
<td></td>
<td>(7.76)</td>
<td>(2.61)</td>
</tr>
<tr>
<td>St Andrews</td>
<td>16.03*</td>
<td>-5.70*</td>
</tr>
<tr>
<td></td>
<td>(7.98)</td>
<td>(2.64)</td>
</tr>
<tr>
<td>Surrey</td>
<td>24.94***</td>
<td>-1.06</td>
</tr>
<tr>
<td></td>
<td>(7.36)</td>
<td>(2.44)</td>
</tr>
<tr>
<td>UCL</td>
<td>11.38</td>
<td>-3.78</td>
</tr>
<tr>
<td></td>
<td>(7.24)</td>
<td>(2.42)</td>
</tr>
<tr>
<td>UEA</td>
<td>20.33**</td>
<td>-1.66</td>
</tr>
<tr>
<td></td>
<td>(7.34)</td>
<td>(2.43)</td>
</tr>
<tr>
<td>Warwick</td>
<td>1.94</td>
<td>-5.84*</td>
</tr>
<tr>
<td></td>
<td>(7.28)</td>
<td>(2.44)</td>
</tr>
<tr>
<td>York</td>
<td>14.72*</td>
<td>-4.19</td>
</tr>
<tr>
<td></td>
<td>(7.27)</td>
<td>(2.42)</td>
</tr>
<tr>
<td>English</td>
<td>8.64</td>
<td>-0.71</td>
</tr>
<tr>
<td></td>
<td>(5.77)</td>
<td>(1.90)</td>
</tr>
<tr>
<td>Business/Management</td>
<td>2.77</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(4.15)</td>
<td>(1.38)</td>
</tr>
<tr>
<td>Chemistry</td>
<td>-4.83</td>
<td>1.66</td>
</tr>
<tr>
<td></td>
<td>(4.92)</td>
<td>(1.64)</td>
</tr>
<tr>
<td>Computer Science</td>
<td>-1.90</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>(5.02)</td>
<td>(1.67)</td>
</tr>
<tr>
<td>Economics</td>
<td>-7.94</td>
<td>-1.07</td>
</tr>
<tr>
<td></td>
<td>(4.75)</td>
<td>(1.59)</td>
</tr>
<tr>
<td>Geography</td>
<td>6.14</td>
<td>-1.09</td>
</tr>
<tr>
<td>Field</td>
<td>Estimate</td>
<td>Standard Error</td>
</tr>
<tr>
<td>------------------------------</td>
<td>----------</td>
<td>----------------</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>-1183.72</td>
<td>390.67</td>
</tr>
<tr>
<td>Observations</td>
<td>2,350</td>
<td>2,350</td>
</tr>
<tr>
<td><strong>Pseudo R-squared</strong></td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>History</td>
<td>15.55**</td>
<td>-1.99</td>
</tr>
<tr>
<td>(5.10) (1.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Law</td>
<td>3.04</td>
<td>0.16</td>
</tr>
<tr>
<td>(4.82) (1.61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematics</td>
<td>1.10</td>
<td>0.98</td>
</tr>
<tr>
<td>(4.47) (1.49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physics</td>
<td>2.95</td>
<td>1.28</td>
</tr>
<tr>
<td>(4.58) (1.52)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Politics/International relations</td>
<td>-3.73</td>
<td>-3.31*</td>
</tr>
<tr>
<td>(4.87) (1.63)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychology</td>
<td>-2.21</td>
<td>-0.53</td>
</tr>
<tr>
<td>(4.45) (1.49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sociology</td>
<td>-1.96</td>
<td>-0.60</td>
</tr>
<tr>
<td>(5.53) (1.86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>3.73</td>
<td>3.02</td>
</tr>
<tr>
<td>(3.85) (1.27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assistant professor</td>
<td>-0.07</td>
<td>-1.16</td>
</tr>
<tr>
<td>(5.82) (1.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior Lecturer</td>
<td>-3.95</td>
<td>-0.24</td>
</tr>
<tr>
<td>(3.23) (1.07)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecturer</td>
<td>-2.92</td>
<td>1.08</td>
</tr>
<tr>
<td>(3.42) (1.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>-4.47</td>
<td>-2.38</td>
</tr>
<tr>
<td>(10.63) (3.53)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continental Europe</td>
<td>-1.73</td>
<td>0.38</td>
</tr>
<tr>
<td>(2.49) (0.83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>14.17*</td>
<td>3.07</td>
</tr>
<tr>
<td>(6.67) (2.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa</td>
<td>-10.23</td>
<td>4.70</td>
</tr>
<tr>
<td>(12.06) (4.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>2.45</td>
<td>-0.75</td>
</tr>
<tr>
<td>(4.00) (1.34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South America</td>
<td>-8.53</td>
<td>-1.10</td>
</tr>
<tr>
<td>(10.11) (3.46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australasia</td>
<td>-2.00</td>
<td>2.09</td>
</tr>
<tr>
<td>(6.64) (2.19)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note 1:** The alignment of conditions, compared to high student expectations and high student engagement, follows a technique devised to provide a holistic view of the alignment of characteristics associated with an outcome (leadership) (Foti & Hauenstein, 2007). **Note 2:** Tobit regressions for the percentage likelihood of responding to a student and the time devoted in minutes are shown. The coefficients for high and low expectations are relative.
to the control treatment where student expectations are unknown to the academic. Robust standard errors are shown in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$
<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Model 1 Likelihood</th>
<th>Model 1 Minutes</th>
<th>Model 2 Likelihood</th>
<th>Model 2 Minutes</th>
<th>Model 3 Likelihood</th>
<th>Model 3 Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Expectations</td>
<td>3.05</td>
<td>-1.11</td>
<td>1.02</td>
<td>-0.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.94)</td>
<td>(0.98)</td>
<td>(2.34)</td>
<td>(0.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Expectations</td>
<td>6.00**</td>
<td>1.29</td>
<td>4.98**</td>
<td>2.03***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.83)</td>
<td>(0.94)</td>
<td>(2.24)</td>
<td>(0.74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1.97</td>
<td>-1.60</td>
<td>0.25</td>
<td>0.96</td>
<td>-1.02</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>(3.36)</td>
<td>(1.12)</td>
<td>(3.47)</td>
<td>(1.16)</td>
<td>(2.75)</td>
<td>(0.92)</td>
</tr>
<tr>
<td>Low Expectations x Female</td>
<td>-5.66</td>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.79)</td>
<td>(1.59)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Expectations x Female</td>
<td>-2.84</td>
<td>1.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.64)</td>
<td>(1.54)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Engagement</td>
<td>1.53</td>
<td>0.09</td>
<td>1.39</td>
<td>0.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.32)</td>
<td>(0.77)</td>
<td>(2.92)</td>
<td>(0.97)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Engagement</td>
<td>5.66**</td>
<td>0.45</td>
<td>6.49**</td>
<td>1.92**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.26)</td>
<td>(0.75)</td>
<td>(2.85)</td>
<td>(0.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Engagement x Female</td>
<td>0.04</td>
<td>-1.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.78)</td>
<td>(1.59)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Engagement x Female</td>
<td>-2.64</td>
<td>-3.96**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.68)</td>
<td>(1.55)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aligned Expect-Engage</td>
<td></td>
<td>7.29***</td>
<td>3.40***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.68)</td>
<td>(1.22)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aligned x Female</td>
<td>0.63</td>
<td>-3.25*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.95)</td>
<td>(1.97)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partial Aligned Expect-Engage</td>
<td>4.08</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.13)</td>
<td>(1.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partial Aligned x Female</td>
<td>-1.66</td>
<td>-0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.93)</td>
<td>(1.63)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unaligned Expect-Engage</td>
<td>-5.48</td>
<td>-0.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.06)</td>
<td>(1.36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unaligned x Female</td>
<td>6.91</td>
<td>-1.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.43)</td>
<td>(2.15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Preference</td>
<td>-6.42***</td>
<td>-2.14***</td>
<td>-6.45***</td>
<td>-2.08***</td>
<td>-6.44***</td>
<td>-2.06***</td>
</tr>
<tr>
<td></td>
<td>(1.55)</td>
<td>(0.51)</td>
<td>(1.55)</td>
<td>(0.51)</td>
<td>(1.55)</td>
<td>(0.51)</td>
</tr>
<tr>
<td>Teaching Preference</td>
<td>7.45***</td>
<td>1.08**</td>
<td>7.46***</td>
<td>1.07**</td>
<td>7.46***</td>
<td>1.14**</td>
</tr>
<tr>
<td></td>
<td>(1.44)</td>
<td>(0.48)</td>
<td>(1.44)</td>
<td>(0.48)</td>
<td>(1.45)</td>
<td>(0.48)</td>
</tr>
<tr>
<td>Time Preference</td>
<td>1168.08</td>
<td>-409.28</td>
<td>1213.74</td>
<td>-403.02</td>
<td>1244.86</td>
<td>-367.24</td>
</tr>
<tr>
<td></td>
<td>(744.96)</td>
<td>(247.47)</td>
<td>(744.56)</td>
<td>(247.11)</td>
<td>(747.27)</td>
<td>(248.71)</td>
</tr>
<tr>
<td>Guilt Proneness</td>
<td>5.35***</td>
<td>2.01***</td>
<td>5.29**</td>
<td>1.94***</td>
<td>5.45***</td>
<td>2.04***</td>
</tr>
<tr>
<td></td>
<td>(1.62)</td>
<td>(0.54)</td>
<td>(1.62)</td>
<td>(0.54)</td>
<td>(1.62)</td>
<td>(0.54)</td>
</tr>
<tr>
<td>Years Post PHD</td>
<td>0.14</td>
<td>0.02</td>
<td>0.15</td>
<td>0.03</td>
<td>0.15</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.04)</td>
<td>(0.14)</td>
<td>(0.04)</td>
<td>(0.14)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>City</td>
<td>Risk Aversion</td>
<td>Altruism</td>
<td>Bath</td>
<td>Birmingham</td>
<td>Bristol</td>
<td>Cambridge</td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------</td>
<td>----------</td>
<td>----------</td>
<td>------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>-0.10</td>
<td>-0.71</td>
<td>8.92</td>
<td>8.79</td>
<td>-0.26</td>
<td>7.98</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(0.42)</td>
<td>(7.88)</td>
<td>(7.34)</td>
<td>(7.39)</td>
<td>(7.81)</td>
</tr>
<tr>
<td></td>
<td>(0.33)</td>
<td>(0.14)</td>
<td>(2.64)</td>
<td>(2.45)</td>
<td>(2.49)</td>
<td>(2.59)</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(0.14)</td>
<td>(7.87)</td>
<td>(7.32)</td>
<td>(7.39)</td>
<td>(7.80)</td>
</tr>
<tr>
<td></td>
<td>(0.33)</td>
<td>(0.14)</td>
<td>(2.64)</td>
<td>(2.44)</td>
<td>(2.49)</td>
<td>(2.58)</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(0.14)</td>
<td>(2.64)</td>
<td>(2.44)</td>
<td>(2.49)</td>
<td>(2.58)</td>
</tr>
<tr>
<td></td>
<td>(0.33)</td>
<td>(0.14)</td>
<td>(2.64)</td>
<td>(2.44)</td>
<td>(2.49)</td>
<td>(2.58)</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(0.14)</td>
<td>(2.64)</td>
<td>(2.44)</td>
<td>(2.49)</td>
<td>(2.58)</td>
</tr>
<tr>
<td></td>
<td>(0.33)</td>
<td>(0.14)</td>
<td>(2.64)</td>
<td>(2.44)</td>
<td>(2.49)</td>
<td>(2.58)</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(0.14)</td>
<td>(2.64)</td>
<td>(2.44)</td>
<td>(2.49)</td>
<td>(2.58)</td>
</tr>
<tr>
<td>Subject</td>
<td>St Andrews</td>
<td>Surrey</td>
<td>UCL</td>
<td>UEA</td>
<td>Warwick</td>
<td>York</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------</td>
<td>--------</td>
<td>-------</td>
<td>------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>15.91*</td>
<td>25.10***</td>
<td>11.74</td>
<td>20.37**</td>
<td>2.20</td>
<td>14.87*</td>
</tr>
<tr>
<td></td>
<td>(7.97)</td>
<td>(2.64)</td>
<td>(7.24)</td>
<td>(7.33)</td>
<td>(2.44)</td>
<td>(7.27)</td>
</tr>
<tr>
<td></td>
<td>(2.63)</td>
<td>(2.43)</td>
<td>(1.21)</td>
<td>(2.22)</td>
<td>(1.79)</td>
<td>(1.17)</td>
</tr>
<tr>
<td></td>
<td>(7.99)</td>
<td>(7.34)</td>
<td>(3.58)</td>
<td>(2.05)</td>
<td>(4.27)</td>
<td>(7.29)</td>
</tr>
<tr>
<td></td>
<td>(2.65)</td>
<td>(2.43)</td>
<td>(1.36)</td>
<td>(1.79)</td>
<td>(1.68)</td>
<td>(1.42)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Coefficient Mean</td>
<td>Coefficient Median</td>
<td>Coefficient Minimum</td>
<td>Coefficient Maximum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------</td>
<td>--------------------</td>
<td>---------------------</td>
<td>--------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assistant professor</td>
<td>-0.57</td>
<td>-1.11</td>
<td>-0.37</td>
<td>-1.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.83)</td>
<td>(1.95)</td>
<td>(5.83)</td>
<td>(1.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior Lecturer</td>
<td>-3.80</td>
<td>-0.07</td>
<td>-3.71</td>
<td>-0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.23)</td>
<td>(1.07)</td>
<td>(3.23)</td>
<td>(1.07)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecturer</td>
<td>-2.87</td>
<td>1.37</td>
<td>-2.69</td>
<td>1.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.43)</td>
<td>(1.14)</td>
<td>(3.43)</td>
<td>(1.13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>-5.29</td>
<td>-2.63</td>
<td>-4.86</td>
<td>-2.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(10.65)</td>
<td>(3.53)</td>
<td>(10.65)</td>
<td>(3.52)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continental Europe</td>
<td>-1.54</td>
<td>0.25</td>
<td>-1.71</td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.51)</td>
<td>(0.84)</td>
<td>(2.51)</td>
<td>(0.84)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>13.50*</td>
<td>3.11</td>
<td>13.61*</td>
<td>2.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.67)</td>
<td>(2.18)</td>
<td>(6.68)</td>
<td>(2.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa</td>
<td>-9.59</td>
<td>5.40</td>
<td>-9.41</td>
<td>5.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(12.07)</td>
<td>(4.06)</td>
<td>(12.07)</td>
<td>(4.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>3.17</td>
<td>-0.74</td>
<td>2.83</td>
<td>-0.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.00)</td>
<td>(1.34)</td>
<td>(4.00)</td>
<td>(1.33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South America</td>
<td>-8.08</td>
<td>-0.87</td>
<td>-8.36</td>
<td>-0.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(10.11)</td>
<td>(3.45)</td>
<td>(10.11)</td>
<td>(3.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australasia</td>
<td>-1.58</td>
<td>1.83</td>
<td>-1.49</td>
<td>1.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.67)</td>
<td>(2.20)</td>
<td>(6.67)</td>
<td>(2.19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-1123.30</td>
<td>427.29*</td>
<td>-1168.42</td>
<td>420.47*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(744.35)</td>
<td>(247.28)</td>
<td>(744.00)</td>
<td>(246.93)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>2,350</td>
<td>2,350</td>
<td>2,350</td>
<td>2,350</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudo R-squared</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: Tobit regressions for the percentage likelihood of responding to a student and the time devoted in minutes are shown. Model 1 = interaction of Expectations and gender, Model 2 = interaction of engagement and gender, Model 3 = interaction of patterns (aligned, partly aligned, unaligned) and gender. Robust standard errors are shown in parentheses, *** p<0.01, ** p<0.05, * p<0.1
FIGURE 1
Likelihood of responding (%) and time devoted (in minutes) by expectation levels
FIGURE 2
Likelihood and Minutes Devoted by expectation levels by gender
Appendix A

Vignettes: text manipulations (see Explanatory Notes)

You are about to answer questions based on a multi-tasking scenario. Please read the scenario carefully.

As an academic you are engaged in a grant bid due in a week's time and you are under time pressure to complete this bid. You have also been lecturing to final year undergraduates who have an exam due for your course. Unfortunately the exam is on the same day as the bid deadline. Any interruptions by your students will benefit their exam results by extra marks but at a personal cost to you in terms of time lost by the interruption. This may diminish the quality of your bid, the success of which is critical for your next review.

{(The scenarios that follow relate to a specific student of yours who is {not} very engaged with your module. You may wish to imagine a specific student of yours who sounds like this.)} {(In your lectures you have made a point of telling your students that you would be prepared to provide support for them during the exam period. [You might be involved in a bid process in the same period as the exam takes place]. As a result, this specific student’s expectations of gaining support from you are high [low].}

Given this scenario, what would be your response to each situation below?

<Email> (Knock on the Door)

This specific student of yours <emails you with> (knocks on your door outside your office hours and asks to discuss) a significant number of questions about the topics that they are having problems with before the exam and asks for an immediate response. Given the scenario above, on a scale of 0% to 100%, <how likely are you to spend time on replying to the student's questions by email?> (how likely are you to break off from what you are doing to meet your student straight away?)

On a scale of 0% to 100%, how likely are you to respond to your student?  

In each case below how many MINUTES would you dedicate to responding to your student?

If you respond (to your student’s request to meet), without any information that it will substantially affect the student & grade, how many MINUTES would you dedicate to responding to your student?  

If, according to your student, <your email response to their questions> (your meeting) could improve their overall grade from a 2/2 to a 2/1, how many MINUTES would you dedicate to responding to your student?  

If, according to your student, <your email response to their questions> (your meeting) could improve their overall grade from a 2/1 to a first, how many MINUTES would you dedicate to responding to your student?
If, according to your student, <your email response to their questions> (your meeting) could avoid the student failing their degree, how many MINUTES would you dedicate to responding to your student? ¶¶**SLIDER SCALE 0 to 100 MINUTES**¶¶

Explanatory Notes:

The 3rd paragraph represents the 3 x 3 manipulations for student engagement and student expectations.

The section bounded by {{ }} is the wording for high engagement whilst the section bounded by { } also appears for low engagement. The entire section does not appear when student engagement is unknown.

The section bounded by [[ ]] is the wording for high expectations whilst the section bounded by [ ] is the alternative/additional wording for low engagement. The entire section does not appear when student expectations are unknown.

For the questions that follow, the specific wordings for <Email> or a (Knock on the Door) are indicated.

Where a slider scale appears in the survey this is indicated by the wording bounded by ¶¶ ¶¶.

Linking of Vignettes to hypothesis testing

The 9 vignettes (3 expectations conditions x 3 engagement conditions) are listed below. The vignettes are linked to hypotheses, with hypothesis 1 testing the effects of high student expectations on extra-curricular teaching effort, hypothesis 2 the effects of high student engagement and hypothesis 3 the interaction effect of high student expectations and engagement.

<table>
<thead>
<tr>
<th>Vignette</th>
<th>Expectations</th>
<th>Engagement</th>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High</td>
<td>High</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>2</td>
<td>High</td>
<td>Low</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>High</td>
<td>Unknown</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Low</td>
<td>High</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Low</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Low</td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Unknown</td>
<td>High</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>Unknown</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Unknown</td>
<td>Unknown</td>
<td></td>
</tr>
</tbody>
</table>
Survey questions relating to control variables

[Measure of Altruism]

Scale of 1 (very unlikely) to 5 (very likely)

How do you assess your willingness to share with others without expecting anything in return when it comes to charity?

[Measure of Risk Aversion]

Scale of 1 (very unwilling) to 5 (very willing)

Are you a person who is generally willing to take risks?

[Further measure of altruism]

Imagine that you won £500 in a prize draw. Considering your current situation, how much of this would you donate to charity?
Chapter 4
“Now you see it now you don’t: The effect of teaching style and seniority on gender bias in teaching evaluations.”
Abstract

Gender bias in teaching evaluations leads to unfair decisions during academics’ careers. In two controlled experiments, I examine whether gender bias is eliminated by an academic’s high warmth teaching style and by seniority. I find that gender bias lowers recommendations to hire female academics delivering identical content as male academics, with the effect mediated by evaluations of the academic’s warmth and/or competence. In Study 1, I test competing hypotheses regarding the effect of teaching style on gender bias. I find that a high warmth teaching style increases women’s perceived warmth, but decreases their perceived competence, so gender bias in hiring recommendations remains. In Study 2, I find that gender bias disappears for senior academics. Finally, I find no evidence of less biased evaluations by those who anticipate gender bias. I discuss my findings in the higher education context and make recommendations to mitigate gender bias in teaching evaluations.

Keywords:
Gender bias, teaching evaluations, teaching style, seniority, bias awareness.
Introduction

Recent decades have seen a surge of evidence in higher education settings pointing to a gender bias in teaching evaluations (Langbein, 1994; MacNell et al., 2015; Ottoboni et al., 2016; Pounder, 2007; Wagner, Rieger, & Voorvelt, 2016; Young, Rush, & Shaw, 2009). This is a problem because teaching evaluations are associated with educational outcomes and are used to make decisions on the careers of academics (Wild & Berger, 2016). To the extent that teaching evaluations assess academics in a biased way, based on their gender rather than specific behaviours, decisions that are key to academic careers may be unfair. For example, the under-representation of women in senior academic roles, especially in male-dominated disciplines, may be due to unfair decisions early in the careers of female academics (Dick & Nadin, 2006; Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012; Newsome, 2008; Sheltzer & Smith, 2014; Way, Larremore, & Clauset, 2016).

A recent quasi-experimental study of 19,920 teaching evaluations at Maastricht University in the Netherlands showed that it is female academics who tend to be evaluated less positively, especially if they are junior, in male-dominated disciplines, and rated by male students (Mengel et al., 2017). Interestingly, the gender bias in teaching evaluations mirrors a recent meta-analysis of gender bias in employment decision making (Koch, D’Mello, & Sackett, 2015). The latter typically examine decisions to hire a candidate, but have not looked at variables of relevance to the study of teaching evaluations, such as the teaching style or seniority of the candidate. What controlled experiments add to quasi-experimental or correlational designs is a ruling out of possible differences in such variables between female and male academics, and a possibility to estimate their effects in isolation and as moderators of the effect of gender (Arbuckle & Williams, 2003; Doubleday & Lee, 2016; MacNell et al., 2015; Ottoboni et al., 2016; Wagner et al., 2016).
In what follows, I build on the literatures in management, economics, and education to formulate testable research hypotheses regarding the role of teaching style and seniority on gender bias in teaching evaluations and hiring recommendations. I consider that making a hiring recommendation to a recruitment committee is a stronger and more salient measure than an overall approval rating of a candidate in a teaching evaluation. I perform controlled experiments testing competing hypotheses regarding the effects of teaching style and the debiasing role of seniority. In addition, I survey lay intuitions of experimental participants regarding a possible gender bias in teaching evaluations and hiring recommendations to a recruitment committee, and examine whether the bias is expected, and if so, what is the effect of bias awareness. Overall, my work contributes to a more fine-grained understanding of the gender bias in teaching evaluations allowing us to identify conditions under which the bias disappears.

To summarize, I find that a gender stereotypical ("warm") teaching style improves perceptions of warmth for female academics but may backfire by lowering perceptions of their competence. Hence, hiring recommendations (overall approval ratings) are lower for female academics (vs. male) irrespective of their style because of a double-bind nature of reactions to their teaching. If their teaching style is low in warmth, lower hiring recommendations are driven by lower perceptions of their warmth, and if their teaching style is warm, lower hiring recommendations are driven by lower perceptions of their competence. Fortunately, gender bias is sensitive to seniority, and I find no evidence of bias against senior female academics in hiring recommendations or warmth evaluations even when they teach in a low warmth style. In conjunction with findings from previous research, these results suggest a need to shield junior academics from decisions that rely on teaching evaluations, especially in the early stages of their careers (Mengel et al., 2017). Moreover, they highlight possible benefits from showcasing titles and other credentials that may indicate more senior
standing for female academics (Bohren, Imas, & Rosenberg, 2018). An unexpected finding was of lower perceptions of warmth for senior (vs. junior) male academics suggesting that senior male academics, unlike their female colleagues, may not need to be concerned with showcasing seniority. Last but not least, subject to considerations of bias blind spot (Page, 2009; Pronin, Lin, & Ross, 2002; Scopelliti et al., 2015), those aware of gender bias hurting female academics are no more likely to correct their evaluations, a result clarified by a recent field study (Boring & Arnaud, 2017). This result suggests caution in treating awareness alone as an effective remedy to the problem.

**Theory Development**

**The role of teaching style**

Academics are commonly evaluated on criteria that align with the two universal dimensions of social cognition: warmth and competence (Fiske, Cuddy, & Glick, 2007). For example, recommendations of research councils suggest assessments of warmth-related “enthusiasm”, “consideration” and “accessibility” and competence-related “class structure”, “mastery of material” and “level of preparation” (Hannover Research Council, 2009). Experimental evidence to date has found significant bias against female academics on both dimensions, including criteria such as enthusiasm, praise, respect and fairness (warmth) and promptness and professionalism (competence) (MacNell et al., 2015).

From a theoretical perspective, teaching evaluations are indeed ripe for gender bias. Teaching is a power relationship that highlights the dependence of the student on the goodwill, mastery and knowledge of the instructor (Schrodt, Witt, & Turman, 2007). The performance of the academic is highly salient to the student as the very reason why students enter the relationship. As a result, students are naturally inclined to judge various aspects of the academic’s performance in the classroom. Often, the judgment is made under time pressure and intuitively (Bassett, Cleveland, Acorn, Nix, & Snyder, 2017; Pinto & Mansfield,
Moreover, higher education is a credence-based service as students lack the knowledge necessary to confidently judge the academic, especially concerning competence (Darby & Karni, 1973; Kasnakoglu, 2016). This makes the evaluation of performance through teaching evaluations highly uncertain (Gruber & Frugone, 2011). Gender stereotypes and considerations of gender-role congruity become an important source of information that helps address the uncertainty in the teaching relationship (Davison & Burke, 2000; Kunda & Spencer, 2003). Yet, the reliance on gender stereotypes and considerations of gender-role congruity is likely to favour male as opposed to female academics because women are typically believed to be less competent than men and less fit to occupy positions of power (Eagly & Karau, 2002). This is particularly true of more male-dominated disciplines, which reinforce the stereotype and established gender roles, making them more salient in judgment (Cejka & Eagly, 1999; Koch et al., 2015).

However, research has also shown that one effective way of generating more positive and accepting evaluations of competent women, such as female academics, is for the women to show warmth, a stereotypically female characteristic associated with care and the pursuit of communal goals (Carli, 2001). Unlike men, women need to show pro-sociality in addition to self-confidence in order to influence others based on their higher performance (Guillén, Mayo, & Karelaia, 2018).

In what follows I formulate competing hypotheses regarding a possible effect of a teaching style that is high on warmth and, hence, stereotypically “female”. On the one hand, I suggest that in the context of teaching, evaluations of female academics may be enhanced if the style of lecture delivery is high rather than low on warmth, and more so than for male academics. Male academics who, from the start, are more likely to be perceived as fulfilling a gender-appropriate role, are simply less likely to be scrutinized in terms of their style. This prediction is supported indirectly by content analysis of qualitative data, including comments
on Ratemyprofessor.com. Adjectives that relate to high versus low warmth in teaching style (bossy, nice, caring, warm, etc.) are more likely to be mentioned in relation to female rather than male academics such that teaching style is more important in the assessment of female academics (Mitchell & Martin, 2018; Shen, 2015). So, if a male and a female academic teach the same content, a teaching style that is high on warmth is likely to raise the warmth and, together with it, competence evaluations for female academics more than for male academics (Carli, 2001). This may happen to the point of possibly eliminating the gender bias in these evaluations, as well as their associated downstream consequences, such as hiring recommendations. I formulate the following research hypotheses:

**Hypothesis 1: (gender bias in hiring/approval)** Hiring recommendations (overall approval ratings) will be lower for female than for male academics who teach the same content.

**Hypothesis 2: (mediation in hiring/approval)** Gender bias in hiring recommendations (overall approval ratings) will be mediated by warmth and competence perceptions.

**Hypothesis 3: (effect of style on warmth and competence):** Relative to a teaching style that is low on warmth, a teaching style that is high on warmth will raise perceptions of the academic’s warmth and competence, and more so for female rather than male academics.

Collectively, Hypotheses 2-3 imply the possibility of a reduction in gender bias under a teaching style that is high on warmth. So,

**Hypothesis 4: (effect of style on gender bias in hiring/approval):** Gender bias in hiring recommendations (overall approval ratings) will be reduced or eliminated under a teaching style that is high on warmth.
Although I predicted in Hypotheses 3-4 that a teaching style high on warmth may help overcome gender bias in the evaluations of female academics due to increasing perceptions of the female’s warmth and competence, competing hypotheses are also possible. To formulate competing hypotheses, I note the specificity of the teaching context in that it is relatively easier to assess the academics’ warmth rather than their competence. To the extent that competence assessments are highly uncertain, they may be affected in the direction of the stereotype especially when the style of teaching reinforces the stereotype. In particular, because women who behave warmly reinforce the gender stereotype, observers are likely to rely more heavily on the idea that women are less competent than men, and less fit to occupy positions of power. As a result, female academics may benefit from higher perceptions of their warmth but at the same time suffer a competence penalty associated with the alignment of the style and the stereotype of someone less knowledgeable. If this was the case, then I would predict that a teaching style that is high on warmth may not diminish or eliminate the gender bias, but rather affect competence perceptions differently for male versus female academics. For women, a warm teaching style could decrease competence perceptions whereas no such effect would be expected for men. Hence, a warm teaching style would increase gender bias in competence evaluations rather than help diminish it.

**Hypothesis 3A: (competing, effect of style on warmth and competence)** Relative to a teaching style that is low on warmth, a teaching style that is high on warmth will raise perceptions of the academic’s warmth, and more so for female rather than male academics. However, it will diminish perceptions of the academic’s competence, and more so for female rather than male academics.
As a result, female academics may continue to be recommended for hiring less than male academics because of their lower perceived competence and fit to the role. Depending on the weight placed on competence versus warmth as determinants of hiring recommendations, the bias may change either upward or downward, and I, therefore, limit my theorizing to the mediating role of warmth and competence for hiring recommendations.

The role of seniority

The fact that female academics may be doubted more in terms of their fit to the role than their male counterparts due to gender stereotypes and considerations of gender-role congruity invites the question of whether seniority has the potential to eliminate the gender bias. If, in a given setting, students require more convincing evidence to infer competence from female academics compared to male academics then a double standard exists (Rubin, 1981; Winocur, Schoen, & Sirowatka, 1989). Double standards are known to impede career advancement (Lyness & Thompson, 2000) but the attainment of a senior position implies, therefore, a higher level of skill or ability (Crocker & Major, 1989). Thus, where individuals reach senior positions despite the existence of double standards this may confer a positive advantage. Indeed, research shows that providing information that supports without ambiguity the high competence of candidates, gender bias disappears (Bohren et al., 2018; Koch et al., 2015).

As senior academics and especially in male-dominated disciplines, women may be judged unambiguously as highly competent. Moreover, it is likely that for senior female academics, both perceptions of warmth and competence will be high supporting their seniority proven fit to the role. Indeed leadership research has argued that, where warmth is perceived as advantageous in a role, women in top positions can be viewed as both warm and competent and enjoy an advantage in evaluations compared to male peers (Byron, 2007;
There has been a lengthy literature on the benefits to female students of female approaches to teaching and of a role model effect (Bettinger & Long, 2005; Carrell, Page, & West, 2010; Lockwood & Kunda, 1997). This could provide the basis for a female seniority advantage in academia.

**Hypothesis 5: (effect of seniority on gender bias on hiring recommendation/approval)**

Gender bias for junior academics will be reduced or eliminated for senior academics.

**Hypothesis 6: (effect of seniority on gender bias on warmth and competence)**

Relative to junior academics, senior academics delivering the same content will be perceived as more warm and more competent, and more so for female rather than male academics.

**Bias awareness**

A number of approaches have been suggested in the literature to overcome biases in decision-making including gender bias (Beshears & Gino, 2015). One important insight is that a more deliberate and thorough analysis of situations helps individuals control their tendency to rely on stereotypes or other faulty generalizations in judging an individual’s performance on a particular task. Bias awareness could help trigger a more deliberate analysis to overcome biased evaluations of male versus female academics. Even though students cannot “blind” themselves to the gender of the instructor, they may mentally simulate counterfactual scenarios. For example, they may consider evaluations they would have given if the same content was delivered by an academic of a different gender, examine the relevance of gender as a factor in their evaluations, and correct their evaluations accordingly. In fact, taking control over tacitly learned reactions to various stimuli in our daily environments, and developing skills of speculation, testing, and generalization has been advocated as a way to “educate” intuitive judgment and overcome biases (Hogarth, 2001;
Morewedge et al., 2015). Those who are aware of gender bias, may be in a better position to revise their judgment to more accurately reflect the quality of teaching and stray away from the considerations of the academic’s gender-role congruity. Consistent with this argument, a field experiment finds that a factual awareness of the gender bias in past evaluations of similar students leads to a reduction in gender bias (Boring & Arnaud, 2017). In the same study, there is a null effect of being merely reminded that one should not discriminate against female academics in teaching evaluations. Consequently,

**Hypothesis 7: (effect of bias awareness):** Those who are aware of the gender bias in teaching evaluations favouring male academics will be less likely to show gender bias in their teaching evaluations (warmth, competence, and hiring recommendations).

**Overview of studies**

I test the research hypotheses in two experimental studies. Study 1 tests for gender bias in the context of a male-dominated discipline (astronomy). I then examine evaluations of warmth, competence and hiring recommendations relative to male versus female academics who deliver the lecture in either a teaching style that is high or low on warmth.

In Study 2, I use the “low warmth” version of the same experimental materials to test the de-biasing effect of seniority. As in Study 1, I examine evaluations of the academic’s warmth, competence, and hiring recommendations. In addition, I elicit intuitions regarding a possible bias, and examine how bias awareness affects teaching evaluations.

**Study 1**

**Participants and Design**

I recruited 479 participants, aged between 18 and 30, on the Prolific.com academic website ($M_{age} = 24.07$, $SD_{age} = 3.17$, 50.2% female, 61.2% with undergraduate or
postgraduate degrees) for a study that asked them to assess a lecture by a candidate in the academic job market, and provide a hiring recommendation to the university. Participants, recruited from Prolific.com were primarily from countries with female representation of fewer than 20% in physics departments, validating the assumption that astronomy, as a branch of physics, is a male-dominated discipline. They were paid £1.40 for completing a 10 minute study (average completion time was 8 minutes 35 seconds). Data were gathered during September 2017.

The study consisted of a 2 (gender: male vs. female) x 2 (warmth: high vs. low) between-subjects design. Participants were randomly assigned to one of the four conditions. The number of participants required for the study was determined based on a-priori power analysis with anticipated small effect sizes (i.e., Cohen’s $f = .15$; (J. Cohen, 1992)) which would require a sample size of 460 to be powered at 90%. All power calculations were conducted using G*Power (Faul et al., 2007).

**Materials**

Participants read an astronomy lecture of around 900 words. The lecture was based on Professor Stephen Hawking’s first Reith Lecture entitled “Do Black Holes Have No Hair?” (Hawking, 2016). In the version of the lecture which was high on warmth, the candidate appeared warm and accessible as a teacher. In the version of the lecture which was low on warmth, the candidate appeared to be cold and patronizing. A silhouette of either a male or female head, together with the academic’s name (Steve Smith versus Sue Smith), was shown on each of the five screens of the lecture text to reinforce the salience of the academic’s gender. Details of the study, including highlighted manipulations for high on warmth and low on warmth lectures, follow at Appendix A.

**Pilot study:** I conducted a pilot study to test whether the teaching context (astronomy lecture) was perceived as male-dominated and whether the high warmth version of the lecture
was perceived as warmer than the low warmth version. Twenty-one individuals (M<sub>age</sub> = 24.33, SD<sub>age</sub> = 3.02, 16 males) participated in this pilot study for a payment of £1.40. For the first test, the academic was described in gender-neutral terms (surname only without a silhouette) and participants rated how likely it was that the academic was male on a 5-point Likert-type scale anchored by 1 (definitely male) to 5 (definitely female). The result, compared to the middle of the scale, confirmed that the astronomy lecture was perceived as male-dominated (χ²(12) = 45.75, p < .01). For the second test, participants rated the academic’s warmth on a 5-point Likert-type scale anchored by 1 (not at all warm) to 5 (very warm) and the high warmth version was rated higher (χ²(3) = 18.43, p < .01) confirming the manipulation of teaching style.

**Procedure**

Participants were randomly assigned to each of the four experimental conditions, and proceeded to read the astronomy lecture. Following the lecture, they assessed the academic candidate in terms of warmth and competence, and provided a hiring recommendation. The survey finished with socio-demographic questions about the participants.

**Measures**

**Warmth.** Participants were asked to assess the academic’s warmth using the items “warm” and “accessible” (Fiske, Cuddy, Glick, & Xu, 2002), as detailed at Appendix A. Participants had to consider the above adjectives and indicate the extent to which they believed the candidate to be each of these things on a 5-point Likert-type scales anchored by 1 (not at all) to 5 (very). The items were averaged together to form a single composite score, where higher scores indicated greater warmth (Cronbach’s α = .76).

**Competence.** Participants were asked to assess the academic’s competence using the items “professional” and “knowledgeable” (Fiske et al., 2002), as detailed at Appendix A. Participants had to consider the above adjectives and indicate the extent to which they
believed the candidate to be each of these things on a 5-point Likert-type scales anchored by 1 (not at all) to 5 (very). The items were averaged together to form a single composite score, where higher scores indicated greater competence (Cronbach’s $\alpha = .70$).

**Hiring recommendation (Overall Approval Rating).** Participants were asked whether the candidate who had given the lecture should be hired on a 5-point Likert-type scales anchored by 1 (definitely reject) to 5 (definitely hire). Given that participants were restricted to between 18 and 30 years old, due to the teaching evaluation context of the study, and that students do not normally contribute to hiring decisions, the wording of the study made it clear that participants were making a recommendation only to a university hiring committee (see materials in Appendix A). I considered a hiring recommendation to be the strongest and most salient measure for the overall evaluation of academic candidates rather than an overall approval rating.

**Control variables.** I controlled for age, gender, level of education, student status, and cross-cultural differences operationalized as the World Economic Forum’s Global Gender Gap Index for 2016 for the country of birth of each participant (World Economic Forum, 2016).

**Manipulation and attention checks.** At the conclusion of the study, participants were asked to indicate the gender of the academic that they had evaluated. They were also presented with an attention check question telling them to complete an answer with a Likert value of 1. A further check was made on outlying survey completion time of less than one standard deviation from the mean (3 minutes 46 seconds) where the fixed pay out was claimed without paying due attention. As a result of these checks, a total of 7 participants (1.5%) were excluded from all subsequent analysis. Exclusion criteria were not applied to slow completion times as participants were still only eligible for a fixed pay out.
Results

Descriptive statistics for all study variables are given in Table 1.

---------------------------------------------
Insert Table 1 about here
---------------------------------------------

Our manipulation of the teaching style worked as expected. The high warmth lecture was rated more highly on warmth than the low warmth lecture ($M = 3.97, SD = 0.68$ versus $M = 3.53, SD = 0.80$, $t(477) = 6.51$, $p < .001$). The manipulation of the candidate’s gender was also successful. 94% of participants in the male condition remembered the academic delivering the lecture as male $\chi^2(1) = 13.68$, $p < .01$ compared to 50%, and 90% in the female condition remembered the academic as female $\chi^2(1) = 12.09$, $p < .01$ compared to 50%.

To test Hypothesis 1, I conducted a multiple regression analysis with the hiring recommendation as the dependent variable. The independent variables were dummies for the gender of the candidate (gender: 1=male, 0=female), the teaching style (warmth: 1=high warmth, 0=low warmth), and their interaction (gender $\times$ warmth). All control variables were included (see Table 2, column 1). Consistent with Hypothesis 1, I found a gender bias: male academics were more likely to be recommended for hiring than their female peers ($\beta = .21$, $p < .05$). The bias held under both high and low warmth teaching style. The effect size for the gender bias was small, and it did not differ substantially across the low warmth and high warmth scenarios (Cohen’s D of -.27 and -.19 respectively).

---------------------------------------------
Insert Table 2 about here
---------------------------------------------
To test Hypothesis 2 (mediation), I analysed whether warmth and competence acted as mediators between the gender of the academic and the hiring recommendation (D. Kenny, 2016). Without controlling for warmth and competence, the hiring recommendation was correlated with gender (Table 2, column 1). For a low warmth teaching style, the indirect (mediated) effects of gender through warmth and competence were significant (Warmth: standardized path coefficient = .10, \( p < .05 \), Competence: standardized path coefficient = .00, \( ns \), Total indirect effect: standardized path coefficient = .10, \( p < .05 \)) whilst the direct effect of gender lost significance (standardized path coefficient = .04, \( ns \)) (see Table 3, row 1). For a high warmth teaching style, the indirect (mediated) effects of gender through warmth and competence were significant (Warmth: standardized path coefficient = .01, \( ns \), Competence: standardized path coefficient = .07, \( p < .05 \), Total indirect effect: standardized path coefficient = .08, \( p < .05 \)) whilst the direct effect of gender lost significance (standardized path coefficient = .00, \( ns \)) (see Table 3, row 2). The indirect effects of gender on the hiring recommendation remained significant when bootstrapping standard errors to allow for kurtosis (95% unstandardized bias corrected confidence interval for low warmth teaching style= 0.02, 0.30 & for high warmth teaching style = 0.00, 0.27, see Table 4). (Preacher & Hayes, 2008). Thus, I found support for Hypothesis 2 that gender bias in hiring recommendations was mediated by warmth and competence perceptions.

Insert Table 3 about here

Insert Table 4 about here
Contrary to Hypothesis 3 but consistent with the competing Hypothesis 3A (effect of style on warmth and competence), I found that the high warmth style had different effects on the evaluations of warmth and competence of academics depending on their gender. For warmth, the high warmth style led to more positive evaluations of warmth for female academics, and the effect was larger than the same effect for male academics. I conducted the regression analysis with warmth as the dependent variable and the academic’s gender, teaching style, and the interaction between the two as independent variables (see Table 2, column 2). The main effect of male gender was positive and significant ($\beta = .34, p < .01$) qualified by a negative and significant interaction term ($\beta = -.27, p < .05$). As for competence, the high warmth style led to somewhat more negative evaluations of competence for female versus male academics. I conducted the regression analysis with competence as the dependent variable and the academic’s gender, teaching style, and the interaction between the two as independent variables (see Table 2, column 3). The interaction term was correctly signed but failed to reach statistical significance ($\beta = -.18, ns$). I further examined the simple slopes for the effect of teaching style on competence evaluations depending on gender (Aiken, West, & Reno, 1991). The results showed there was a statistically significant decrease in the evaluations of competence for women when they taught in a high warmth style ($\beta = -.19, p < .05$), but not for men ($\beta = -.01, ns$) (see Figure 1).

Overall, the results show that gender bias persisted in the hiring recommendation in the high warmth scenario because of lower competence evaluations for female academics.
Discussion

Study 1 showed that in a male-dominated discipline the delivery of the same teaching content led to greater hiring recommendations for male rather than female academics, irrespective of whether the style of delivery was low or high on warmth. Female academics benefited more than male academics from teaching in a style that was high on warmth (as opposed to low on warmth) in terms of evaluations of their warmth. However, they also suffered a somewhat greater penalty in terms of evaluations of their competence, which led to lower hiring recommendations.

Study 2

In Study 2 I tested the de-biasing role of seniority. To date, many empirical studies of gender bias in teaching evaluations examine junior academics (Boring, 2017; MacNell et al., 2015), and a recent field study finds stronger effects of gender for junior as opposed to more senior academics (Mengel et al., 2017). In Study 2, I distinguished deliberately between junior (post-PhD) and senior (Professor Level) academics to test the debiasing role of seniority.

Participants and Design

I recruited a further 478 participants, aged between 18 and 30, on Prolific.com (M_{age} = 24.40, SD_{age} = 3.24, 49.9% female, 64.6% with undergraduate or postgraduate degrees) for a study that asked them to assess a lecture by a candidate in the academic job market, and provide a hiring recommendation to the university. Participants were paid £1.40 for completing a 10 minute study (average completion time was 9 minutes). Data were gathered during November 2017.

The study consisted of a 2 (gender: male vs. female) x 2 (seniority: Professor vs. junior) between subjects design. Participants were randomly assigned to one of the four
conditions. The number of participants required for the study was determined as in Study 1 to be powered at 90% with small effect sizes.

**Materials**

Participants read the low warmth version of the astronomy lecture used in Study 1. The academic was described as a post-PhD male/female candidate or as a Professor male/female candidate. A silhouette of either a male or a female head was shown on each of the five screens of text to reinforce the gender manipulation. In addition, depending on the experimental condition, each screen showed the post-PhD candidate’s name without the use of any titles, or the senior academic’s name used next to the “Professor” title (e.g., Sue Smith versus Professor Sue Smith). Details of the study, including highlighted manipulations for high on warmth and low on warmth lectures, follow at Appendix A.

**Procedure**

Participants were randomly assigned to each of the four experimental conditions, and proceeded to read the astronomy lecture. Following the lecture, they assessed the academic candidate in terms of warmth and competence, and provided a hiring recommendation. The survey finished with questions about gender bias and socio-demographic questions.

**Measures**

**Warmth.** Participants were asked to assess the academic’s warmth using the items “warm” and “accessible” on a 5-point Likert-type scale as in Study 1, as detailed at Appendix A. The items were averaged together to form a single composite score, where higher scores indicated greater warmth (Cronbach’s $\alpha = .77$).

**Competence.** Participants were asked to assess the academic’s competence using the items “professional” and “knowledgeable” on a 5-point Likert-type scale as in Study 1, as detailed at Appendix A. The items were averaged together to form a single composite score, where higher scores indicated greater competence (Cronbach’s $\alpha = .67$).
Hiring recommendation (Overall approval rating). Participants were asked whether the candidate who had given the lecture should be hired on a 5-point Likert-type scale as in Study 1, as detailed in Appendix A. Given that participants were restricted to between 18 and 30 years old, due to the teaching evaluation context of the study, and that students do not normally contribute to hiring decisions, the wording of the study made it clear that participants were making a recommendation only to a university hiring committee (see materials in Appendix A). I considered a hiring recommendation to be the most salient measure for the overall evaluation of academic candidates.

Bias awareness. Following the survey questions, participants were asked whether they thought there is a male, female or no bias in evaluations of warmth, competence and the hiring recommendation generally. The order of the questions was randomized and I coded for bias awareness as 1 if participants believed in a male bias and 0 if participants did not believe in any bias or believed in a female bias (1 = Bias Aware, 0 = Not Aware).

Control variables. I controlled for age, gender, level of education, student status, and cross-cultural differences in the gender gap as in Study 1.

Manipulation and attention checks. Participants were asked the gender of the academic that they had evaluated. A total of 6 participants (1.25%) failed the manipulation check and further checks on outlying survey completion times less than one standard deviation from the mean (3 minutes 30 seconds) where the fixed pay out was claimed without paying due attention. They were excluded from all subsequent analysis. Exclusion criteria were not applied to slow completion times as participants were still only eligible for a fixed pay out.

Results

Descriptive statistics for all study variables are given in Table 5.
To test Hypothesis 5, I used multiple linear regressions of the hiring recommendation on dummies for the gender of the candidate (gender: 1 = male, 0 = female), the seniority of the candidate (seniority: 1 = Professor, 0 = Junior), and their interaction (gender × seniority) (see Table 6, column 1).

There was no significant gender bias at Professor Level for the hiring recommendation ($\beta = -0.03$, $ns$, effect size Cohen’s $D = 0.05$) supporting Hypothesis 5. Simple slopes analysis showed a significant improvement in the hiring recommendation at Professor Level, compared to junior levels, for female academics with little change for in the hiring recommendation for male academics ($\beta = 0.23$, $p < 0.05$ for female academics, versus $\beta = 0.01$, $ns$ for male academics, see Figure 2). Gender bias for junior levels was comparable in size to the bias in Study 1 (Cohen’s $D = 0.21$) but only weakly significant ($\beta = 0.19$, $p < 0.1$). To clarify what the findings implied for testing the existence of a gender bias at junior academic levels, I compared the gender bias in the hiring recommendation between the two studies. This does not yield a significant difference between the sizes of the effects found (($\beta = -0.03$, $ns$, see Table 7).
As in Study 1, gender bias in the evaluations of the academic’s warmth was statistically significant ($\beta = .22, p < .05$) (see Table 6, column 2). The regression of warmth evaluations on dummies for the gender of the candidate (gender: 1 = male, 0 = female), the seniority of the candidate (seniority: 1 = Professor, 0 = Junior), and their interaction (gender $\times$ seniority), included a significant interaction effect ($\beta = -.39, p < .05$). At first sight, this seemed to be consistent with Hypothesis 6. However, simple slopes analysis showed that rather than significantly improving evaluations of warmth for female academics, seniority diminished the evaluations of warmth for male academics ($\beta = .15, ns$ for female academics, versus $\beta = -.24, p < .05$ for male academics, see Figure 2). This was unexpected, and contrary to the rationale of Hypothesis 6 which predicted higher warmth evaluations for senior female academics (due to the fact that they overcame double standards) without any drop in the warmth evaluations of male academics.

For competence, seniority affected the evaluation of female and male academics the same (see Table 6, column 3).

As for bias awareness, I found that it was generally low. Irrespective of whether respondents considered warmth, competence or hiring recommendations, roughly 70% believed no gender bias existed. Participants who believed in a female advantage were most numerous when it came to warmth evaluations (24% versus 2% when competence was considered, and 5% when hiring recommendation was considered). To analyse whether awareness of gender bias helped participants correct their evaluations, I performed regressions of warmth, competence and the hiring recommendation on all independent and control variables from the previous analyses, adding the variable “bias aware” and the
interaction between “bias aware” and the dummy for the academic’s gender to the analysis (see Table 8). The coefficients for bias awareness and the interaction of bias awareness and the gender dummy were not significant in any of the regressions. Contrary to Hypothesis 7, I did not find that being aware of a male bias de-biases teaching evaluations.

Insert Table 8 about here

Discussion

In Study 2, I found that the gender bias against junior female academics, in warmth and in hiring recommendations, disappeared with seniority. The finding of a significant gender bias for junior academics proved wrong roughly 70% of respondents who considered that gender bias was not a factor in teaching evaluations and hiring.

Moreover, there was an unexpected bias against senior male academics such that their warmth evaluations diminished and became inferior to those of senior female academics while the latter did not improve in comparison to junior female academics. This pattern of results suggested a mechanism for the elimination of gender bias that I did not initially foresee. At senior levels, male academics seem to have lost the advantage that drove their hiring recommendations when academics were portrayed as juniors. The mechanism for this effect should be tested in future research.

Importantly, those who reported being aware of the bias did not show more accurate evaluations of the candidates on either warmth, competence, or hiring recommendations.

General Discussion

In two controlled experiments, I showed evidence of gender bias in teaching evaluations and hiring recommendations for junior academics. These results are consistent
with the predictions of the gender-role congruity theory (Cejka & Eagly, 1999; Eagly & Karau, 2002) and previous empirical findings (MacNell et al., 2015; Mengel et al., 2017; Ottoboni et al., 2016). My work provides a more fine-grained understanding of the workings of gender bias and points to the settings in which the bias disappears, representing a novel contribution to the existing literature. I discuss below the important practical and theoretical implications of my work.

**Practical Implications**

I submitted to an empirical test the idea that the academic’s teaching style may matter for the magnitude of the bias (Mitchell & Martin, 2018). In particular, the hope was that a female-stereotypic high warmth style may shield female academics from lower evaluations of the same teaching content. Yet, the results provided support to a competing hypothesis whereby a high warmth teaching style brought a competence penalty which led to lower hiring recommendations for female academics despite the fact that evaluations of their warmth improved to the level of the evaluations for their male peers. These results may be specific to the higher education setting because there are important information asymmetries between students and academics in understanding the subject matter and hence, evaluating the competence of the academic. Because student evaluations of the academic’s competence are therefore uncertain, they are particularly likely to be vulnerable to bias. Hence, whereas in other settings, a competent performance by a woman would be perceived more positively if the woman adopted a high warmth style (Carli, 2001), in an academic setting, her high warmth style triggered a greater reliance on the gender stereotype and considerations of gender-role congruity. This stereotyping exerted a downward pressure on the evaluations of her competence and the hiring recommendation.
In study 2, I showed that the gender bias is sensitive to seniority and disappears for professors as opposed to junior academics. This result supports recent calls in the literature to shield academics from decisions based on teaching evaluations, and qualifies it by the importance of doing so at least in the early stages of their careers. With seniority, the female academic’s title begins to pave the way for her to be assessed on par with her male peer for equal performance. Unexpectedly, seniority produced a negative effect on warmth evaluations of male academics. On the one hand, this suggests that male professors may stand nothing to gain from showcasing their senior status. On the other hand, it is important to understand the underlying mechanism for this effect. It may be that at junior levels, male academics experience an unfair advantage (rather than female academics experiencing a disadvantage), which is corrected at senior levels. However, it may also be that at junior levels, female academics are subjected to stereotype-driven unfair disadvantage (as argued in this paper) whereas, at senior levels, a seniority-related stereotype produces a similar disadvantage for male academics. Interestingly, a study of dynamic discrimination finds a similar reversal where biased beliefs initially led to a female disadvantage but a subsequent female advantage with an accumulation of favourable evaluations (Bohren et al., 2018). It remains to be investigated in future research which of these mechanisms applies in academia and whether my findings are specific to academia or generalize beyond the higher education context.

Finally, I examined bias awareness among the very people who evaluated a given teaching content in my experiments, and tested the idea that bias awareness leads to less biased teaching evaluations. It was informative to find that the vast majority of the study participants did not believe that gender played a role in teaching evaluations. Regrettably, those who anticipated the gender bias failed to correct for it in their own teaching evaluations. Although many organizations may rely on building awareness about the gender bias as the
bias mitigation strategy, this result suggests caution in relying on that kind of intervention alone without other forms of career support to junior female academics.

**Theoretical Implications**

My work makes theoretical contributions and opens promising avenues for future research. First, I show that gender biases may benefit from a systematic study in credence versus non-credence settings (Darby & Karni, 1973; Gruber & Frugone, 2011; Kasnakoglu, 2016). My prediction is that when violation of a misaligned (e.g. high warmth-low competence) stereotype on one of its dimensions affects overall performance evaluations, behaving in a stereotype-consistent manner will be beneficial if the other dimension is non-credence-based, and may not be beneficial if the other dimension is credence-based. In the latter case, behaving in a stereotype-consistent manner may simply reinforce the stereotype.

Second, most of the literature on gender bias in academia focuses on a female disadvantage (Carli, 2001). In contrast, I found evidence of a male disadvantage for senior academics in a male-dominated discipline. This finding merits further research attention. In fact, the field as a whole could benefit from a more thorough understanding of all the explanatory mechanisms behind gender biases that produce either male or female disadvantages.

Finally, I focused in the experimental work on the evaluations of the academic along two fundamental dimensions of social cognition (warmth and competence) which impact the career decisions of individual academics (Fiske et al., 2007). Yet, it is also instructive to shed light on the role that gender biases may play in the evaluations of learning outcomes and taught content which impact decisions regarding academic institutions. At the level of academic institutions, this may help further inform the impact of initiatives that aim to reduce
possible gender biases and provide impetus for more active research on bias-reducing interventions.

**Limitations and Future Directions**

There are some potential limitations to the studies conducted. Firstly, a manipulation check for seniority was not performed in study 2 but may be performed as an out of sample test to ensure that the participants attended to differences in the seniority of academics as intended. However, manipulation checks were conducted for gender in both studies. Secondly, the external (ecological) validity of the assessment of warmth for participants reading, rather than being physically present and observing, a lecture can be questioned. Thirdly, the questions to participants on the bias awareness of others, eliciting judgements of other participant’s bias may be problematic as research has shown that people in general have a bias blind spot, tending to think of others as more prone to bias than themselves (Page, 2009; Pronin et al., 2002; Scopelliti et al., 2015). Given the methodology I have used, my finding that mere awareness of bias is insufficient to debias teaching evaluations may be erroneous. In this context, a large field study has recently evaluated the effect of differing interventions to debias student evaluations (Boring & Arnaud, 2017). The study found that a normative statement that students should not discriminate in teaching evaluations, making students aware of the existence of bias, did not reduce gender bias. However, when the normative statement was supplemented by precise information on how male students, like the participants, had discriminated against female lecturers in the past, gender bias was reduced. Other research indicates that the effect of interventions is stronger for participants with a lower susceptibility to bias blind spot (Scopelliti et al., 2015). These results clarify my experimental finding, confirming that mere awareness of bias is insufficient to debias teaching evaluations.
Looking forwards, a number of studies have found that gender bias occurs in male-dominated roles, but may not be present in gender balanced or stereotypically female roles (Carli, 2001; Koch et al., 2015; Mengel et al., 2017). Whether the effects I find generalize to non-male dominated academic disciplines will be the subject of further research.
Tables and Figures

**TABLE 1**
*Means (and standard deviations in brackets) for Study 1 variables*

<table>
<thead>
<tr>
<th></th>
<th>Low Warmth Scenario</th>
<th>High Warmth Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Warmth</td>
<td>3.36</td>
<td>3.70</td>
</tr>
<tr>
<td></td>
<td>(0.78)</td>
<td>(0.79)</td>
</tr>
<tr>
<td>Competence</td>
<td>4.37</td>
<td>4.36</td>
</tr>
<tr>
<td></td>
<td>(0.55)</td>
<td>(0.56)</td>
</tr>
<tr>
<td>Hiring</td>
<td>3.81</td>
<td>4.02</td>
</tr>
<tr>
<td></td>
<td>(0.85)</td>
<td>(0.75)</td>
</tr>
<tr>
<td>Control Variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male Respondent</td>
<td>0.51</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(0.50)</td>
</tr>
<tr>
<td>Graduate Degree</td>
<td>0.46</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(0.50)</td>
</tr>
<tr>
<td>Postgraduate Degree</td>
<td>0.16</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(0.37)</td>
<td>(0.39)</td>
</tr>
<tr>
<td>Over 24 years old</td>
<td>0.46</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(0.50)</td>
</tr>
<tr>
<td>Gender Index</td>
<td>0.54</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(0.50)</td>
</tr>
<tr>
<td>Observations</td>
<td>125</td>
<td>127</td>
</tr>
</tbody>
</table>

*Note*: the control variables are for the gender of the respondent, whether the respondent has a graduate or postgraduate degree, whether the respondent is over 24 years old, and a (global) gender index measuring the degree of gender inequality in the respondent’s country of origin.
TABLE 2
Regressions for Hiring Recommendation, Warmth and Competence in Study 1

<table>
<thead>
<tr>
<th>Variables</th>
<th>Hiring</th>
<th>Warmth</th>
<th>Competence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Academic</td>
<td>0.21**</td>
<td>0.34***</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.09)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>High Warmth Scenario</td>
<td>0.17</td>
<td>0.58***</td>
<td>-0.19**</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.09)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Male Academic x High Warmth Scenario</td>
<td>-0.06</td>
<td>-0.27**</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.14)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>Male Respondent</td>
<td>0.03</td>
<td>0.06</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Graduate Degree</td>
<td>-0.02</td>
<td>0.00</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Postgraduate Degree</td>
<td>0.04</td>
<td>0.20*</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.10)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Over 24 years old</td>
<td>0.04</td>
<td>-0.06</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Gender Index</td>
<td>0.06</td>
<td>0.09</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Constant</td>
<td>3.74***</td>
<td>3.27***</td>
<td>4.39***</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.10)</td>
<td>(0.08)</td>
</tr>
</tbody>
</table>

Observations: 479
R-squared: 0.02

Note: standard errors in parenthesis, * = p < .10, ** = p < .05, *** = p < .01
TABLE 3
*Standardized Mediation Effects for Warmth and Competence on the Hiring Recommendation in Study 1.*

<table>
<thead>
<tr>
<th>Teaching Style</th>
<th>DV = Hiring Recommendation</th>
<th>Indirect effect of IV on DV</th>
<th>Total Effect IV on DV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direct Effect of IV on DV</td>
<td>Via Warmth</td>
<td>Via Competence</td>
</tr>
<tr>
<td>Gender - Low Warmth</td>
<td>.04</td>
<td>.10</td>
<td>.00</td>
</tr>
<tr>
<td>Gender - High Warmth</td>
<td>.00</td>
<td>.01</td>
<td>.07</td>
</tr>
</tbody>
</table>

*Note:* Low Warmth N=252, High Warmth N=227, DV = dependent variable, IV = independent variable (Gender).

Values in **bold** = significant at p < .05.
<table>
<thead>
<tr>
<th>Teaching Style</th>
<th>Low Warmth</th>
<th>High Warmth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average bootstrap estimate</td>
<td>0.1587</td>
<td>0.1319</td>
</tr>
<tr>
<td>Bootstrap standard error</td>
<td>0.0715</td>
<td>0.0714</td>
</tr>
<tr>
<td>Asymptotic standard error</td>
<td>0.0700</td>
<td>0.0682</td>
</tr>
<tr>
<td>Normal-based interval</td>
<td>(0.0215, 0.2960)</td>
<td>(0.0046, 0.2656)</td>
</tr>
<tr>
<td>Percentile interval</td>
<td>(0.0216, 0.2992)</td>
<td>(0.0018, 0.2747)</td>
</tr>
<tr>
<td>Adjusted percentile interval</td>
<td>(0.0232, 0.3004)</td>
<td>(0.0018, 0.2740)</td>
</tr>
</tbody>
</table>

Notes: Three bootstrapping confidence intervals are shown for each coefficient. Asymptotic standard errors are also shown for comparison.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Junior Level</th>
<th>Professor Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Warmth</td>
<td>3.38</td>
<td>3.62</td>
</tr>
<tr>
<td></td>
<td>(0.78)</td>
<td>(0.78)</td>
</tr>
<tr>
<td>Competence</td>
<td>4.30</td>
<td>4.25</td>
</tr>
<tr>
<td></td>
<td>(0.59)</td>
<td>(0.57)</td>
</tr>
<tr>
<td>Hiring</td>
<td>3.83</td>
<td>4.02</td>
</tr>
<tr>
<td></td>
<td>(0.86)</td>
<td>(0.89)</td>
</tr>
<tr>
<td>Control Variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male Respondent</td>
<td>0.40</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>(0.49)</td>
<td>(0.50)</td>
</tr>
<tr>
<td>Graduate Degree</td>
<td>0.50</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(0.50)</td>
</tr>
<tr>
<td>Postgraduate Degree</td>
<td>0.13</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>(0.33)</td>
<td>(0.37)</td>
</tr>
<tr>
<td>Over 24 years old</td>
<td>0.48</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(0.50)</td>
</tr>
<tr>
<td>Gender Index</td>
<td>0.57</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(0.50)</td>
</tr>
<tr>
<td>Observations</td>
<td>118</td>
<td>126</td>
</tr>
</tbody>
</table>
TABLE 6
Regressions for Warmth, Competence and Hiring Recommendation in Study 2

<table>
<thead>
<tr>
<th>Variables</th>
<th>Hiring</th>
<th>Warmth</th>
<th>Competence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Academic</td>
<td>0.19*</td>
<td>0.22**</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.11)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Professor Level</td>
<td>0.23**</td>
<td>0.15</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.11)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Male Academic x Professor Level</td>
<td>-0.22</td>
<td>-0.39**</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.15)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>Male Respondent</td>
<td>-0.00</td>
<td>0.07</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Graduate Degree</td>
<td>-0.12</td>
<td>0.00</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.09)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Postgraduate Degree</td>
<td>-0.13</td>
<td>0.09</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Over 24 years old</td>
<td>-0.05</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Gender Index</td>
<td>-0.04</td>
<td>-0.06</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Constant</td>
<td>3.95***</td>
<td>3.36***</td>
<td>4.30***</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.11)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Observations</td>
<td>478</td>
<td>478</td>
<td>478</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses, * = p < .10, ** = p < .05, *** = p < .01. Low warmth scenario only.
TABLE 7
Comparison of Male Academic Coefficients in Study 1 and Study 2

<table>
<thead>
<tr>
<th>Variables</th>
<th>Hiring</th>
<th>Warmth</th>
<th>Competence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Academic</td>
<td>0.22**</td>
<td>0.33***</td>
<td>0.01</td>
</tr>
<tr>
<td>(0.10)</td>
<td>(0.10)</td>
<td>(0.07)</td>
<td></td>
</tr>
<tr>
<td>Study 2</td>
<td>0.03</td>
<td>0.04</td>
<td>-0.08</td>
</tr>
<tr>
<td>(0.11)</td>
<td>(0.10)</td>
<td>(0.07)</td>
<td></td>
</tr>
<tr>
<td>Male Academic x Study 2</td>
<td>-0.04</td>
<td>-0.11</td>
<td>-0.02</td>
</tr>
<tr>
<td>(0.15)</td>
<td>(0.14)</td>
<td>(0.10)</td>
<td></td>
</tr>
<tr>
<td>Male Respondent</td>
<td>0.07</td>
<td>0.08</td>
<td>-0.08</td>
</tr>
<tr>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.06)</td>
<td></td>
</tr>
<tr>
<td>Graduate Degree</td>
<td>-0.03</td>
<td>-0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.06)</td>
<td></td>
</tr>
<tr>
<td>Postgraduate Degree</td>
<td>0.02</td>
<td>0.15</td>
<td>0.04</td>
</tr>
<tr>
<td>(0.12)</td>
<td>(0.11)</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>Over 24 years old</td>
<td>-0.10</td>
<td>-0.11</td>
<td>-0.03</td>
</tr>
<tr>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.06)</td>
<td></td>
</tr>
<tr>
<td>Gender Index</td>
<td>0.09</td>
<td>0.08</td>
<td>0.09</td>
</tr>
<tr>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.06)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>3.79***</td>
<td>3.31***</td>
<td>4.35***</td>
</tr>
<tr>
<td>(0.11)</td>
<td>(0.10)</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>494</td>
<td>494</td>
<td>494</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.02</td>
<td>0.04</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses, * = p < .10, ** = p < .05, *** = p < .01. Low warmth scenario only.
# TABLE 8
Regression Analysis of the Role of Bias Awareness on Gender Bias in Study 2

<table>
<thead>
<tr>
<th>Variables</th>
<th>Hiring</th>
<th>Warmth</th>
<th>Competence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male Academic</td>
<td>0.21*</td>
<td>0.20*</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.11)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Professor Level</td>
<td>0.23**</td>
<td>0.15</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.11)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Male Academic x Professor Level</td>
<td>-0.21</td>
<td>-0.40**</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.15)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>Bias Aware</td>
<td>-0.01</td>
<td>-0.31</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.22)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Male Academic x Bias Aware</td>
<td>-0.10</td>
<td>0.40</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.32)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>Male Respondent</td>
<td>-0.01</td>
<td>0.07</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Graduate Degree</td>
<td>-0.12</td>
<td>0.00</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.09)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Postgraduate degree</td>
<td>-0.13</td>
<td>0.09</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Over 24 years old</td>
<td>-0.04</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Gender Index</td>
<td>-0.04</td>
<td>-0.07</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Constant</td>
<td>3.96***</td>
<td>3.38***</td>
<td>4.32***</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.11)</td>
<td>(0.08)</td>
</tr>
</tbody>
</table>

Observations: 478 478 478
R-squared: 0.02 0.03 0.02

Note: Standard errors in parentheses, * = p < .10, ** = p < .05, *** = p < .01
FIGURE 1
Study 1 – Simple Slopes Analysis for the Hiring Recommendation, Warmth and Competence.

Note: 95% confidence interval shown as a dotted line.
Note: * = p < .10, ** = p < .05, *** = p < .01
FIGURE 2
Study 2 – Simple Slopes Analysis for the Hiring Recommendation, Warmth and Competence.

Note: 95% confidence interval shown as a dotted line.
Note: * = p < .10, ** = p < .05, *** = p < .01
Appendix A

**Materials: Studies 1 and 2**

**Introductory Screens**

Welcome to this study. The purpose of the study is to assist in evaluating two candidates' job applications. Both are graduate students who are applying for lecturer positions at a university. You will be asked to rate the work prepared by each candidate and to recommend to a recruitment committee whether you think each candidate should be hired or rejected.

Before commencing the study you will be presented with a consent form.

When you are ready to proceed, please press NEXT PAGE

***

Please select I CONSENT AND WISH TO PROCEED and NEXT PAGE if you agree with the terms below to continue to the survey.

- I have read and understood the Participant Information Sheet provided (version 2.0, dated 24/07/2017).
- I agree for my anonymised data to be used for this study.
- I understand that data will be collected using Qualtrics software. This software is used by academic institutions worldwide and Qualtrics give assurances that data collected in the EU will remain in the EU. I understand that all the data collected from my responses will be anonymised and the collected data points will not be linked to my identity.
- I understand that all project data will be held for at least 6 years and all research data for at least 10 years in accordance with University policy and that my personal data is held and processed in the strictest confidence, and in accordance with the Data Protection Act (1998).

163
I understand that I can request for my data to be withdrawn up to one month after completing the survey and that following my request all data already collected from me will be deleted. I confirm that I have read and understood the above and freely consent to participating in this study. I have been given adequate time to consider my participation.

I CONSENT AND WISH TO PROCEED

I DO NOT CONSENT AND WISH TO EXIT FROM THIS SURVEY

***

Lecture Introduction: junior level [professor level]

You are about to see the text of an astronomy lecture (of about 900 words) presented by a graduate student who is a candidate applying for a [senior] science lecturer position at a university. [The candidate you are about to assess is a Professor of Astronomy.] The candidate's [professor’s] performance in this lecture will be crucial to the recruitment committee's decision. Your role, once you have seen the lecture, is to rate the candidate's [professor’s] performance and indicate whether you would either hire or reject them.

When you are ready to proceed, please press NEXT PAGE.

***
Welcome to today’s introductory lecture. My name is <Professor> Sue {Steve} Smith and I am a graduate student [Professor of Astronomy]. I hope you will find my talk interesting. I will be very happy to discuss complex areas with individual students afterwards. [I would appreciate quiet throughout the lecture so please turn mobile phones off. I will take questions after the lecture].

My talk is on the history of black holes. It is said that fact is sometimes stranger than fiction, and nowhere is that more true than in the case of black holes.

Black holes are stranger than anything dreamed up by science fiction writers, but they are firmly matters of science fact. The scientific community was slow to realize that massive stars could collapse in on themselves, under their own gravity, and how the object left behind would behave.

In fact Albert Einstein even wrote a paper in 1939, claiming stars could not collapse under gravity, because matter could not be compressed beyond a certain point. Many scientists shared Einstein’s gut feeling.

The principal exception was the American scientist John Wheeler, who in many ways is the hero of the black hole story. In his work in the 1950s and '60s, he emphasized that many stars would eventually collapse, and the problems that posed for theoretical physics.
He also foresaw many of the properties of the objects which collapsed stars become, that is, black holes. The phrase ‘black hole’ is simple enough but it’s hard to imagine one out there in space. Think of a giant drain with water spiralling down into it. Once anything slips over the edge or ‘event horizon’, there is no return.

NASA describes stars as rather like pressure-cookers. The explosive force of nuclear fusion inside them creates outward pressure which is constrained by gravity pulling everything inwards. Eventually, however, the star will exhaust its nuclear fuel. The star will contract. In some cases, it may be able to support itself as a white dwarf star. However, it was shown before the war, that the maximum mass of a white dwarf star exceeds that of the sun. A similar maximum mass was calculated by a Soviet physicist for a star made entirely of neutrons. [However, Subrahmanyan Chandrasekhar showed in 1930, that the maximum mass of a white dwarf star is about 1.4 times that of the sun. A similar maximum mass was calculated by a Soviet physicist, Lev Landau, for a star made entirely of neutrons.]

In 1939 Robert Oppenheimer, of later atomic bomb fame, showed that when a massive star exhausted its nuclear fuel it could not be supported by pressure. [In 1939 Robert Oppenheimer, with George Volkoff and Hartland Snyder, showed that when a massive star exhausted its nuclear fuel it could not be supported by pressure.] And that if one neglected pressure, a uniform spherically systematic symmetric star would contract to a single point of infinite density. Such a point is called a singularity.

Take my word for it. [Evidentially.] all our theories of space are formulated on the assumption that spacetime is smooth and nearly flat, so they break down at the singularity,
where the curvature of space-time is infinite. In fact, it marks the end of time itself. That is what Einstein found so objectionable.

Then the war intervened.

Most scientists, including Robert Oppenheimer, switched their attention to nuclear physics, and the issue of gravitational collapse was largely forgotten. Interest in the subject revived with the discovery of distant but bright objects, called quasars. The first of these, 3C273, was discovered in 1963.

Nuclear processes could not account for their energy output, because they release only a percent fraction of their rest mass as pure energy. Of course the only alternative was gravitational energy, released by gravitational collapse.

Gravitational collapses of stars were re-discovered. It was clear that a uniform spherical star would contract to a point of infinite density, a singularity.

When John Wheeler introduced the term black hole in 1967, it replaced the earlier name, frozen star. Wheeler's coinage emphasized that the remnants of collapsed stars are of interest in their own right, independently of how they were formed.

So, what do we need to know about black holes? [So, what can anyone tell me about the properties of black holes?]
From the outside, you can't tell what is inside a black hole. You can throw television sets, diamond rings, or even your worst enemies into a black hole, and all the black hole will remember is the total mass, and the state of rotation.

A black hole has a boundary, called the event horizon. It is where gravity is just strong enough to drag light back, and prevent it escaping.

Because nothing can travel faster than light, everything else will get dragged back also. Falling through the event horizon is a bit like going over Niagara Falls in a canoe. If you are above the falls, you can get away if you paddle fast enough, but once you are over the edge, you are lost. There's no way back.

It appears that the number of configurations that could form a black hole of a given mass, although very large, may be finite. You won’t need to know the math for this. Jacob Bekenstein suggested that from this finite number, you could interpret what we call the entropy of a black hole. This would be a measure of the amount of information that was irretrievably lost during the collapse when a black hole was created.

The apparently fatal flaw in Bekenstein's suggestion was that if information is lost, which is apparently what is happening in a black hole, there should be some release of energy - but that flies in the face of the theory that nothing comes out of black holes.

This is a paradox. And it's one which I am going to return to in my next lecture, when I'll be exploring how black holes challenge the most basic principle about the predictability of the
universe, and the certainty of history, and asking what would happen if you ever got sucked into one.

My name is <Professor> Sue {Steve} Smith and I'd like to thank you for attending today.

*I am available to discuss this lecture now and during office hours on the complex topics that we have covered today. [For anyone who was not able to follow this introductory lecture there is an extensive reading list on my website.]*

**Evaluation Questions (Likert scale: 1 lowest, 5 highest)**

This is the first lecture in an optional science course at a college. From the text that the candidate has prepared did they appear professional?

From the text that the candidate has prepared did they appear knowledgeable?

From the text that the candidate has prepared, did they appear to be a warm person?

From the text that the candidate has prepared, did they appear to be an accessible person?

Should the university hire the candidate as a science lecturer?

**Gender Questions**
What was the gender of the candidate that you assessed for the astronomy lecture? (M/F)

**Study 2 only**

Which of the following will be true for other people assessing the astronomy lecture?

(Female warmer/male warmer/neither)

(Female more competent/male more competent/neither)

(Female more likely to be hired/Male more likely to be hired/neither)

**Demographic Questions**

Finally, please provide some information about yourself. What is your gender? (M/F)

Please enter your age in years (Numeric)

For how many years have you spoken English? (Numeric)

What level is your highest qualification? (High School/Degree/Postgraduate degree)

Where were you born? (Table)

Where do you currently live? (Table)

Please enter your Prolific.com id. (THIS IS NEEDED FOR PAYMENT)
Chapter 5
Conclusions and implications for policy and further research

Summary of Findings

This thesis examined how academics handle multiple demands in the UK higher education sector. It addressed the questions of why some academics progress faster through their careers than others and what regularities may help sustain gender imbalances in career success in academia. The value of considering these questions lies in the economic importance of, and the increasing demands placed on, the UK higher education sector.

An argument made in this thesis is that academics have some degree of discretion in their time allocation decisions between research, teaching and administration/service with consequences for career outcomes. The marketization of UK public sector higher education has reduced academic autonomy by establishing principal-agent line management chains leading to control through workload accountability (Canaan & Shumar, 2008; J. Kenny & Fluck, 2014; J. Kenny, 2017; Olssen & Peters, 2005; Olssen, 2016; Vardi, 2009). With universities largely funded on a per student basis, both high quality teaching and prestige-enhancing research are desirable for universities. However, for academics, teaching, research and administration/service may be viewed as substitutes (Gautier & Wauthy, 2007; Hattie & Marsh, 1996; Marsh & Hattie, 2002). Indeed, research and teaching may have been differentiated by an increasing focus on teaching for employability, associated with neoliberalism in higher education (Canaan & Shumar, 2008). Neoliberalism has also, arguably, increased the incentive to devote time to research as a portable asset in an increasingly globalised job market (Gautier & Wauthy, 2007; Murphy & Sage, 2014). So, the incentives facing academics, including their intrinsic motivations for each role, may influence choices between research and teaching, subject to institutional constraints on choices, producing very different allocations of time to research, teaching and administration/service (Allgood & Walstad, 2013; Boyd & Smith, 2016; Gautier & Wauthy, 2007; Kossi et al., 2013; Olssen, 2008; Vardi, 2009).
On this basis, chapters two and three considered how individual preferences, and responsiveness to student expectations and engagement, affect discretionary time allocations to research, teaching and service. Chapter four moved on to the assessment of outputs, considering differences in the evaluation of teaching by gender.

A common theme across the three chapters is differences in effects by gender. Gendered differences in academic career outcomes have been attributed to systemic cultural sexism in a male-centered system, especially in respect of working mothers (Heijstra et al., 2015; O'Hagan, 2018; Savigny, 2014; Savigny, 2017). Certainly female academics face different societal expectations for care giving which may lead to gendered differences in time usage and power over time (Beddoes, 2015; Heijstra et al., 2015; Rafnsdóttir & Heijstra, 2013). Further, gender differences in volunteering for non-promotable administration/service tasks and gender differences in recognizing contributions may influence time allocations (Babcock et al., 2017; Sarsons, 2017; Vesterlund et al., 2014). Conceptually, whilst controlling for domestic responsibilities (number of children, young children, job mobility), I assume that both male and female academics have some power over their allocation of time. On this basis, I find gender differences working in favour of male academics in terms of allowing for greater emphasis on research in chapter two (due to preferences for research) and chapter four (due to more favourable evaluations of teaching) but not in chapter three where male and female academics were equally responsive to student expectations and engagement.

I contribute to existing literature in each chapter. Chapter two’s focus on academic’s preferences for research and teaching has received limited attention and my work represents a step in exploring the link between research, teaching and time preferences and career outcomes (Callaghan & Coldwell, 2014; Matthews et al., 2014). I shed light on the relationship between the roles of research and teaching, by finding a conflict between
research and teaching preferences and that greater preferences for research predict more time allocated to research (Hattie & Marsh, 1996; Marsh & Hattie, 2002). I also find a mediating role of a preference for research in devoting time to research activities based on gender. Thus I contribute to the study of gender differences in role preferences (Gino et al., 2015).

In chapter three I find that high student expectations of support and student engagement give rise to extra-curricular teaching effort. I test whether guilt aversion could be a possible mechanism leading to extra-curricular teaching efforts but obtain conflicting findings. In particular, guilt-proneness is found to be positively associated with extra-curricular teaching effort. However, guilt aversion might not be the reason why academics respond to greater student expectations and engagement. I thus contribute to the literature on social exchange theory (Blau, 1964; Emerson, 1976) and to the study of guilt aversion (Battigalli & Dufwenberg, 2007; Charness & Dufwenberg, 2006).

The subject of chapter four is teaching evaluation bias. I find a bias in favour of male academics at junior levels, and I explore the role of warmth and competence perceptions about academics as driving the academic’s overall evaluation, which is a novel approach in this setting (MacNell et al., 2015; Ottoboni et al., 2016). In addition, very few studies have considered what may help eliminate the gender bias in teaching evaluations. I examined the role of academic seniority (Mengel et al., 2017) and bias awareness (Boring & Arnaud, 2017). I thus extend current work on the teaching evaluations bias, and derive new insights, in a controlled experimental setting.

The following sections consider the policy implications arising from the thesis, the external (ecological) validity of the research instruments used in each chapter, and promising directions for future research.
Policy Implications

A number of policy implications arise from this thesis. I discuss below these implications referencing each of the chapters in turn.

Chapter two has generated preliminary evidence that research preferences may drive allocations of time to research, which, as shown in the literature (Barbezat, 2006; Hattie & Marsh, 1996; Rey, 2001), contributes to greater research productivity and career success. It was also found that research preferences are not aligned to teaching preferences but rather greater research preferences are typically associated with lower teaching preferences. One policy implication therefore is the importance of encouraging research-based teaching so that academics do not regard teaching and research as conflicting activities but find ways of leveraging the time devoted to one of these for making progress on the other front. Recent studies that find that greater research productivity is associated with better teaching (Healey et al., 2010; Shin, 2011; Zhang & Shin, 2015) suggest that this is possible. Practices such as the analysis of data from classroom simulations and assessments, and hands-on projects whereby the students are asked to reach out to specific target populations for data, may help in making better teachers become more research-productive, further strengthening the relationship between research and teaching. I also find a gender difference in preferences for research which helps explain why female academics may be devoting less time to research. Such a difference could be predicted from a previously hypothesised stronger social orientation of women (Croson & Gneezy, 2009) which may apply to female academics. In this sense, policies that celebrate female role models that devote themselves to research and showcase forms of research that benefit others and the broader society may help women develop stronger preferences for research, and help, therefore, the career success of female academics.
In chapter three, I showed that both male and female academics are more likely to exert extra-curricular teaching effort in response to greater student expectations of support and greater student engagement. It is important to note, therefore, that when current policies in higher education raise student expectations, this may have repercussions for how much academics work in addition to what their reward systems compensate them for. Whereas the bulk of regulatory concerns centre around student experience and value for money (Department for Business Innovation and Skills, 2016; Kandiko & Mawer, 2013), they should attend to the experiences of academics as well, and allow for greater flexibility in the reward systems so that academics do not begin to regard the system as unfair and demotivating. As my work shows, the risk of devoting more time to teaching away from research time, which is highly valuable to the careers of academics, in response to high student expectations and high engagement is faced by both male and female academics. Again, teaching that is more research-based and where the interactions with students would centre more around research-relevant assignments may be valuable in addressing this problem as well.

Finally, the policy implications from chapter four have to do with the experiences of those academics whose evaluations of teaching are likely to distort the effort invested into this activity. I have seen that the delivery of equal teaching content was assessed more negatively (as reflected in lower recommendations to hire the academic for a university position) for women, and junior women, in particular. Some authors found evidence of a similar phenomenon in field studies of teaching evaluations (Mengel et al., 2017; Ottoboni et al., 2016). Collectively, what these findings suggest is that a policy of shielding academics from the impact that teaching evaluations would normally have on their careers for a period of time would be useful. The period of time needs to be chosen closer to the beginning of the academic career. In this time, teaching evaluations may be collected for their role of
informational feedback. The policy may be terminated once both male and female academics have had the time to experiment and learn, as well as grow in their careers to a level at which their credentials may protect them from biases in teaching evaluations. Chapter four also contributes to the debate on the importance of bias awareness for bias reduction. In particular, I showed that those more aware of gender bias in teaching evaluations did not correct their evaluations to compensate for the bias. Hence, mere awareness might not be the most effective tool for overcoming biases in teaching evaluations, a finding clarified by a recent field study (Boring & Arnaud, 2017).

**External Validity of the Research Instruments**

I will consider the external (ecological) validity of the research instruments used in each chapter. In chapter two, my findings for time preferences are subject to the question of whether academics consider the extrinsic rewards from research outputs as a motivation to engage in research in the real world. I argued that monetary rates of time preferences may act as a mediator of time allocated to research given that the monetary benefits of research, in terms of career advancement, may be deferred far into the future. However, some activities have been argued to provide their own intrinsic reward (Deci, 1971; Kahneman & Thaler, 2006), so an academic may allocate time to research because they enjoy it without regard to the longer term career benefit. Indeed, I did not find a significant relationship between rates of time preference and time allocated to research in my study.

In chapter three, the vignettes used to determine extra-curricular teaching effort in response to the manipulations of student expectations and engagement were designed to mirror a real world conflict. Further, unlike student population samples, the target audience faced similar research and teaching responsibilities in the real world. The online format avoided cues from an invigilator in a laboratory setting. However, to guard against possible
experimenter demand effects, prompted by cues about what constitutes acceptable behaviour, I embedded the experiment in a lengthy set of survey questions (Zizzo, 2010). The ordering of vignettes was also randomized to avoid spillover effects (Transue, Lee, & Aldrich, 2009).

In chapter four, I have highlighted that the methodology used eliciting judgements of other participant’s bias may be problematic as research has shown that people in general have a bias blind spot, tending to think of others as more prone to bias than themselves (Page, 2009; Pronin et al., 2002; Scopelliti et al., 2015). A recent field study (Boring & Arnaud, 2017) clarified my experimental finding that mere awareness of bias is insufficient to debias teaching evaluations. Also in chapter four, the external (ecological) validity of the assessment of warmth for participants reading, rather than being physically present and observing, a lecture can be questioned. Further, the sample population was drawn from Prolific.com and so, unlike in a field experiment, was not directly invested in the teaching that they were evaluating (MacNell et al., 2015; Ottoboni et al., 2016). However, unlike in a field experiment, in the controlled experimental setting I use, “teaching quality and style can literally be held constant by deceiving students about the instructor’s true gender identity” (Mengel et al., 2017, p. 4). So, experiments can provide vital insights into real-world behaviours and complement field experiments.

The internal validity of the analytic methods employed is strong. The use of randomized economic experiments in chapters three and four, represent a benchmark, in terms of their empirical credibility, against which non-experimental studies can be evaluated (Angrist & Pischke, 2010). Further, the analysis of mediating variables used in chapters two and four is an established tool in the psychology literature to understand the mechanism through which the causal variable affects the outcome (D. Kenny, 2016).

In summary, whilst concerns can be raised about the external validity of specific results they do not necessarily undermine the main findings of this thesis. However, the
concerns raised provide a basis for adopting better methodology to further explore my findings as detailed in the next section.

**Future Directions**

An assumption made in this thesis is that academics, of both genders, have at least some discretion over their allocation of time. Capturing additional data, comparing actual against contractual allocation of time and exploring restrictions on the allocation of time, could enhance the strength of this argument. Further, all the findings in this thesis open promising avenues for future research both in terms of methodological improvement, and in terms of testing novel theoretical insights.

**Better Methodology.** Chapter two, on the importance of the academic’s preferences for time dedicated to research and subsequent career growth through higher research productivity, would be best addressed using a longitudinal study of preferences, time allocation and career outcomes. This would entail careful consideration of incentives for study participants to minimize rates of attrition, and a choice of the appropriate data-analytic strategy (Diggle et al., 2002).

Chapter three was a 3x3 experiment on the role of student expectations and engagement. Additionally, I was interested to test for the gender effects in the sensitivity of academics to the manipulated levels of expectations and engagement. Perhaps the design could be simplified (e.g. only engagement being manipulated) and gender differences could be tested not across all academic disciplines but with the consideration of the specificities of work within different academic disciplines. So, a possible design could be 3 (engagement) x 2 (gender) x 2 (hard sciences versus not), and this could provide for a more thorough investigation of the context in which the effects of engagement arise.
In chapter four, I examined the bias in teaching evaluations by means of a controlled experiment. In respect of the perceptions of bias awareness, the approach adopted in a recent field study on interventions to debias student evaluations (Boring & Arnaud, 2017), should be adopted to avoid the issue of participants tending to think of others as more prone to bias than themselves (Page, 2009; Pronin et al., 2002; Scopelliti et al., 2015). In respect of the main results, the people evaluating a given teaching content were not actual students (although I sampled only respondents who were aged between 18 and 30) and not particularly invested in my chosen domain of study (astronomy). I could improve on this study by better tailoring the context of teaching evaluations to the actual experience of the respondents, and run similar studies in different disciplines of study with actual students in that discipline. Such a design would produce additional insight into the differences in bias between various study disciplines.

**Theoretical insight.** Importantly, everything I learned in the thesis may be equally relevant to the study of public services other than academia. First my findings in chapter two open an interesting agenda for exploring the effect of role preferences in shaping career progress in a range of occupations in public services as many occupations may require choices concerning the time to be devoted to more technical versus more social aspects of the job. Just as academics may have a stronger preference for research versus teaching, healthcare professionals may have a stronger preference for surgery versus primary care, and experience dramatically different career outcomes as a result. As in education, it would be important to test then whether such impactful differences in preferences align with gender differences, and whether it is in the interest of the broader society to sustain these differences culturally and by means of job design.

As for the focus of my investigation in chapter three, in primary care the General Practitioners (GPs) face real conflicts between engaging in extra time with a particular patient
and fulfilling local authority data requirements. This is the setting I examined in chapter three where academics were shown to respond with extra-curricular teaching effort to their students’ high expectations and high engagement. The design of the study reported in chapter three may be employed to test whether patient expectations matter for how GPs handle 10 minute consultations with their patients, and what the repercussions are for the quality of patient outcomes.

In chapter four, I note that education is an example of a credence service whereby students lack the knowledge necessary to confidently judge the academic, especially concerning competence (Darby & Karni, 1973; Kasnakoglu, 2016). Similarly healthcare is a credence service, whereby patients lack the knowledge necessary to confidently judge the competence of the doctor. This suggests that the “bedside manner” of the doctor may have similar effects to those of teaching style in the studies reported in chapter four. It would be interesting to examine to what extent the effect I find in teaching style (a warm teaching style improves perceptions of the female academic’s warmth at the expense of somewhat diminishing her competence perceptions) generalizes beyond credence settings such as education and healthcare, and what are the factors (e.g. a teacher’s or a doctor’s visible competence credentials) that may protect the person delivering the service from the perverse effects of approaching their job in a gender-stereotypical manner.

I would like to conclude by noting that the theoretical insights in this thesis lead the way to further research into productivity within the higher education sector specifically, and potentially other areas of the public sector. The economic importance of the UK higher education sector, and the effects of the demands currently being placed on academics by government and institutional policies, suggest that the questions raised here merit further research attention.
References


Anderson, J., Burks, S. V., Carpenter, J., Götte, L., Maurer, K., Nosenzo, D., . . . Rustichini, A. (2013). Self-selection and variations in the laboratory measurement of other-


doi: [http://dx.doi.org/10.1348/096317907X174349](http://dx.doi.org/10.1348/096317907X174349)


doi: [http://dx.doi.org/10.1111/0022-4537.00238](http://dx.doi.org/10.1111/0022-4537.00238)


doi: [http://dx.doi.org/10.2139/ssrn.2465594](http://dx.doi.org/10.2139/ssrn.2465594)


doi:http://dx.doi.org/10.1348/096317905X68709


doi:[http://dx.doi.org/10.1002/ase.1609](http://dx.doi.org/10.1002/ase.1609)


doi: [http://dx.doi.org/10.1016/j.tics.2006.11.005](http://dx.doi.org/10.1016/j.tics.2006.11.005)

doi: [http://dx.doi.org/10.1037/0022-3514.82.6.878](http://dx.doi.org/10.1037/0022-3514.82.6.878)


Fumasoli, T., Gornitzka, Å, & Maassen, P. A. (2014). *University autonomy and organizational change dynamics*. Oslo: ARENA.

doi: [10.1016/j.euroecorev.2006.02.006](http://dx.doi.org/10.1016/j.euroecorev.2006.02.006)


Kenny, D. (2016, ). Power analysis app MedPower. learn how you can do a mediation analysis and output a text description of your results: Go to mediational analysis using


