NEW DIRECTIONS IN BEHAVIOURAL ECONOMICS

ESSAYS ON PERSONALITY AND WELL-BEING

NEEL OCEAN

Thesis towards the degree of Doctor of Philosophy
Department of Economics
University of Warwick

September 2016

This thesis was typeset using the typographical look-and-feel CLASSICTHESIS, developed by André Miede.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AWB Affective Well-being

A Agreeableness

APS Annual Population Survey

BHPS British Household Panel Survey

BIC Bayesian Information Criterion

C Conscientiousness

CEM Coarsened Exact Matching

CWB Cognitive Well-being

E Extraversion

EIV Errors-in-Variables

GPA Grade Point Average

IPIP International Personality Item Pool

IV Instrumental Variables

LR Likelihood Ratio

MTURK Amazon Mechanical Turk

N Neuroticism

NEO-FFI Neuroticism-Extraversion-Openness Five-Factor-Inventory

O Openness

OLS Ordinary Least Squares

ONS Office for National Statistics

RF Range-Frequency model

SOEP Socio-Economic Panel

SWLS Satisfaction With Life Scale

WERS Workplace Employment Relations Study

WTP Willingness to Pay

ACKNOWLEDGEMENTS

I thank my supervisors Andrew Oswald and Daniel Sgroi for their support during my studies.

The University of Warwick benefits from a large group of interdisciplinary researchers interested in behaviour and decision making. I thank all the participants and members of DR@W and the Behavioural Sciences Group at Warwick Business School; especially Gordon Brown, Neil Stewart, Ganna Pogrebna, Eugenio Proto, Nick Chater, and Alexander Mushore.

Joe Miele from MTurkData has been invaluable in helping me to administer surveys online.

I thank the Economic and Social Research Council for funding me with their studentship for the first 3 of the 4 years of my degree.

Thanks to my parents for helping me to solve practical life problems (which I doubt I will ever be very good at). Lastly, thanks to Casey for being my best friend throughout my journey. All the squishes.

DECLARATION

All of the work in this thesis is my own, and was completed during the period of my registration. Data sources will be indicated where appropriate. No material in this thesis has been used before, or published. This thesis has not been submitted at any institution other than the University of Warwick.

Neel Ocean
University of Warwick, September 2016

THESIS SUMMARY

This thesis is divided into two parts, each consisting of two self-contained chapters. The first part provides new findings in the economics of personality, and well-being.

Chapter 1 studies the implications of 'personality mismatch'. Mismatch in labour economics has generally been treated as a 'black box'. Therefore, the well-being impact on a poorly matched worker is not well understood. I find that workers whose personalities are more poorly matched to the requirements of their job have substantially lower levels of life satisfaction, and a lower wage.

Chapter 2 is the first study that attempts to uncover the determinants of well-being prioritisation. There is no consistent evidence of variation in priorities over the life cycle. Life satisfaction is the most valued aspect of well-being throughout life, yet people overestimate the relative value placed by others on happiness. Well-being prioritisation is strongly influenced by well-being levels and by individual fixed effects such as personality, health level, and smoking frequency.

The second part of this thesis explores two novel ideas previously unconsidered. It represents a first attempt at providing some insight to these issues.

Chapter 3 develops a model describing how consumers might adjust for a potential bias in extreme online review scores. A randomised experiment finds that individuals do not seem to be making such adjustments. Hence, there are negative implications for consumer welfare from false or biased extreme reviews.

Finally, Chapter 4 is an ambitious investigation into how personality characteristics of workers within an economy may influence the composition of its industrial output. Big Five personality factors are predictive of future industry change, but further work needs to be done to verify this. This work highlights the relevance of personality data in the analysis of long-standing economic issues.

FOREWORD

In the heyday of mathematical economics, Leontief (1971) made a bold and honest statement of the fact that empirical evidence in support of theoretical assumptions in economics was severely lacking. Whilst more and more complex models were being employed, their correspondence to the real world became less and less clear. This, he said, was largely responsible for the relative isolation of the discipline when compared with other social sciences.

With this thesis, I wish to emphasise the fact that human behaviour and cognition is the fundamental building block of any economic system – a fact acknowledged nearly a century ago by J. M. Clark (1918). By applying insights from psychology (as well as from other modern sciences, such as computing), I believe we can address many problems in economics that have been exhaustively conceptualised, yet poorly understood in practice. One hopes that in the coming years, this unified approach will be appreciated as a necessity. As Thaler (2016) argues, this would take economics back to the way it first began – an intuitive and open discipline, based upon observation and evidence.

Part I

PERSONALITY AND WELL-BEING

1

PERSONALITY MISMATCH AND WORKER WELL-BEING

Mismatch in labour economics has generally been treated as a 'black box'. Therefore, the well-being impact on a poorly matched worker is not well understood. This chapter is one of the first to study the implications of 'personality mismatch'. Workers whose personalities are more poorly matched to the requirements of their job have substantially lower levels of life satisfaction, and a lower wage. This relationship holds even when job satisfaction is accounted for, suggesting that a personality mismatch at work has welfare implications outside the work environment. A mismatch in Conscientiousness has a relationship with earnings that is twice as strong as that of experience. These findings imply that achieving a good worker-occupation personality match is important in maximising the well-being of an economy.

1.1 INTRODUCTION

Understanding unemployment and labour market flow is a classical line of research in economics. There have been numerous models in economics exploring job search and matching (e.g. Mortensen and Pissarides, 1994; Shimer, 2007). The traditional research focus in this area has been to uncover the form of a mismatch function that abstractly explains the matching of unemployed workers to vacancies. However, the mismatch function is a 'black box'. It is designed to explain the high-level Beveridge Curve relationship that describes the empirical observed negative correlation between unemployment and vacancies.

The ingredients of the mismatch function have not been explicitly tested (Petrongolo and Pissarides, 2001). Its determinants are commonly thought of as being related to factors such as education, ability, and geographic mobility. However, insights from behavioural economics about individual characteristics have proved to be important in the analysis of labour market issues (Dohmen, 2014). A key aspect of individual heterogeneity from psychology that has only recently been explored in economics is personality (e.g. Almlund et al., 2011; Borghans et al., 2008; Heckman, Stixrud, and Urzua, 2006). Employers spend considerable time and effort on the recruitment process in order to find a suitably matched worker. Workers also spend considerable time and effort making sure they choose the 'right' job for them.

In the past few decades, it has come to our attention that unemployment has important implications for an individual's subjective well-being (e.g. Blanchflower and Oswald, 2004b; A. E. Clark, 1996; A. E. Clark and Oswald, 1994; Oswald, 1997). However, poorly matched workers who are still employed may also suffer. Therefore, in order to understand the welfare impact of personality mismatch, I provide evidence to show that a mismatch in terms of Big Five personality traits between a worker and their job is associated with lower levels of subjective well-being, even when controlling for key determinants of life satisfaction. I also find that a mismatch in Conscientiousness leads to lower earnings.

Well-being is important to economists for the simple reason that, above all else, all of us want to maximise it.[1] Happiness is particularly relevant to the labour market, as there are clear links between unemployment and unhappiness after controlling for more standard economic variables such as income (A. E. Clark and Oswald, 1994).[2] This research supports the view that unemployment is largely involuntary (Frey and Stutzer, 2002), and therefore it is important to identify causes of attrition for preventative purposes. Easterlin (2005) explains that *work & personality* is one of the three main factors that affect happiness, alongside *material standard of living*, and *family & health*.

Subjective well-being (life satisfaction in particular) has been found to correspond extremely well to economic choice. Benjamin et al. (2012) find that choices correspond to the option that provides higher well-being 83% of the time. Whilst other factors (such as money) also contribute to choice, it is clear that any notion of a utility function has subjective well-being as its core argument. Hence, any finding in relation to well-being levels is highly relevant to economic decision making.

1 Or *experienced utility*, as in Kahneman, Wakker, and Sarin (1997).
2 The magnitude of the effect is dependent on unemployment duration, among other factors.

The Big Five personality factors (e.g. Goldberg, 1990, 1992) are widely studied and implemented in personality psychology. Measures based upon the Big Five are the most prominent to penetrate the economics literature, due to the ability to easily test for predictive power of particular traits on outcomes. The Big Five originated through adjective analysis of the English language, the idea being that if someone has an observed personality characteristic that is consistent across individuals, then there must be a word to describe it (Allport and Odbert, 1936). Words were grouped in order to identify common factors. These factors were reduced over time until the following five were obtained: Agreeableness (A), Conscientiousness (C), Extraversion (E), Neuroticism (N), and Openness (O). Whilst these factors are not independent of each other, it is generally accepted that they cannot be reduced further without loss of information (McCrae and John, 1992).

Extraversion and Neuroticism are linked to responsiveness to positive and negative affect, respectively. As such, these two factors have the strongest relationship with subjective well-being (e.g. DeNeve, 1999; DeNeve and Cooper, 1998; Diener and Lucas, 1999; Diener, Oishi, and Lucas, 2003), with Neuroticism being the most predictive of the Big Five factors. Openness is related to intellectualism and creativity. Conscientiousness, having roots in self-discipline and orderliness, is predictive of achievement and success related outcomes (e.g. Nyhus and Pons, 2005). Agreeableness captures the tendency to be warm, compassionate, altruistic, and so on. It has been linked with social-cognitive theory of mind (Nettle and Liddle, 2008), which has implications for behaviour in strategic settings.

Many studies have addressed the relationship between personality and labour market outcomes, though work in this area has been relatively recent. For example, Judge, C. A. Higgins, et al. (1999) show using panel data that Conscientiousness has a positive predictive effect on career success, both in subjective satisfaction terms and objective income terms. Neuroticism has a negative effect, but only in objective terms. High Conscientiousness is linked with gaining more satisfaction from having higher income (Boyce and Wood, 2011). However, highly Conscientious people also suffer the largest drop in life satisfaction when they become unemployed (Boyce, Wood, and Brown, 2010).

Uysal and Pohlmeier (2011) study the effects of personality on the probabilities of entering and leaving employment. They find in German Socio-Economic Panel (SOEP) data that instantaneous employment probability is significantly affected by Conscientiousness (positively) and Neuroticism (negatively). They also find that Big Five traits are powerful in explaining unemployment duration. Bowles, Gintis, and Osborne (2001) stress the importance of noncognitive abilities in explaining wage determination. Nyhus and Pons (2005, 2012) explore these links empirically to find strong negative effects of Neuroticism on wages as well as the fact that adding personality traits to an econometric model explains 11.5% of the observed gender wage gap.

The majority of elicitation methods for Big Five factors involve self-reporting. Depending on incentives and context, one might expect an individual to be biased (or even outright lie) in their responses. However, these effects have little impact on predictive power (Borghans et al., 2008). Economists may find the derivation of the Big Five measure through factor analysis to be less appealing than tests that are specifically tailored to predict real-world outcomes (Borghans et al., 2008). Although both approaches have their merits, this may be part of the reason economics as a whole has taken longer to recognise the potential of personality measures.

Personality traits are relatively stable across the working lifespan, though they are unsettled in both early age and post-retirement (Lucas and Donnellan, 2011). Although there are theorists who would claim that personality is 'set like plaster', personality is a combination of both genetics and early environment (Almlund et al., 2011; Polderman et al., 2015). In practical terms, Cobb-Clark and Schurer (2012) show using Australian HILDA panel data that the Big Five are robust over time to all but repeated extreme life shocks. Therefore, they are suitable for use as economic explanatory variables.[3] More recently, Boyce, Wood, Daly, et al. (2015) show that an extended period of unemployment is an extreme enough shock to change mean levels of Agreeableness, Conscientiousness, and Openness. However, more work (and better large scale longitudinal personality data) is required in order to understand when and why personality changes.

3 The data they analysed was only for a 4-year period, however, so it would seem a longer panel is required to validate this finding.

Whilst generalised theories of mismatch and labour market flows have been explored in economics for many years, assessing the impact of mismatch empirically has been rare. However, there has been considerable interest in business and management to try to match the right worker to the right job. For example, Larson, Rottinghaus, and Borgen (2002) explain that the RIASEC model (Holland, 1997) has been in existence for around 20 years longer than the Big Five. Rather than looking at personality characteristics of the individual, the RIASEC model is focused specifically on assigning people to the correct work category, by way of the 'Self-Directed Search' questionnaire.[4]

Since the RIASEC asks what people *want* to do rather than where they fit, the model is likely eliciting preferences rather than underlying traits. As Almlund et al. (2011) explain, preferences and personality are different concepts. Personality helps to shape one's preferences, and likely acts like a set of constraints on behaviour, or as a functional mapping of genetic traits to preferences and actions. On a related note, empirical work comparing economic preference measures (such as attitudes to risk) with personality traits highlights that the two concepts are complementary, and not substitutes for one another (Becker et al., 2012).[5]

De Fruyt and Mervielde (1999) find a clear disparity between Big Five personality factors and RIASEC, despite some overlap. Whilst RIASEC is more predictive of employment status, the Big Five tend to be better in finding the best fit employees. This is intuitive in the following way. Taking on a job that fits with preferences may bring short term utility. However, a poor fit in terms of personality is likely to cause dissatisfaction and a reduction in efficiency.

The 'person-fit' literature in management supports the view that individuals who fit their jobs better are more satisfied with them, and therefore are less likely to become voluntarily unemployed. Most notably, Chatman (1991) found this effect in the public accounting industry by defining 'fit' in terms of shared *values* between individuals and firms. However, Judge and Cable (1997) explain that values are more similar to preferences than personality traits. Therefore, the finding from this line of literature is that choosing a less valued job leads to an

4 The RIASEC categories are: Realistic, Investigative, Artistic, Social, Enterprising and Conventional.

5 It should be emphasised, however, that the economic preferences considered in Becker et al. (2012) are not the same as the aspirational preferences captured by the RIASEC categories.

increased likelihood of dissatisfaction and job attrition. This is somewhat tautological - people choosing a *preferred* option are, by definition, attaining higher levels of utility. The management literature does not address the effects of personality matching on job satisfaction, or indeed overall life satisfaction.

Personality seems intuitively more likely to be stable than preferences. Whilst personality does change across the life cycle (Lucas and Donnellan, 2011; Specht, Egloff, and Schmukle, 2011), most psychologists today accept that individuals have some form of stable personality (see Almlund et al., 2011, for further references and a brief discussion). In contrast, preferences and tastes are strongly affected by environment and context (e.g. Simonson and Tversky, 1992), and often develop based on experience. For example, a new drinker may prefer light beer, but after drinking for some time, may eventually prefer stronger beer (Hoeffler and Ariely, 1999).

Traits act more like constraints to preferences than defining preferences themselves (Almlund et al., 2011). This being the case, finding a good fit in terms of personality should be much more relevant to longer term well-being. Gardner et al. (2012) describe that the closeness of personality between an individual and the modal personality of an organisation is a key ingredient for a good fit, since organisations are relatively homogeneous. They find various links between Big Five trait combinations and goodness of fit to certain organisational cultures. For example, less Agreeable people perceive themselves as a better fit for a market culture (Gardner et al., 2012). To the extent that Big Five factors are predictive of fit, this suggests that personality traits are closer to the root of the labour market matching problem than self reported values or preferences.

In addition to this, the Big Five measure of personality is more established and validated than many employer-employee matching indicators, such as the Myers-Briggs Type Indicator (McCrae and Costa, 1989). These studies also seem to be aimed at specific cases or industries, rather than assessing more global effects on individual well-being. In their meta-analysis of the person-fit literature, Kristof-Brown, Zimmerman, and Johnson (2005) suggest that a personality trait approach to matching would be most appropriate:

> "Future studies of personality-based fit are advised to use measures that are capable of assessing various conceptualizations of fit at the trait level, rather than overall personality profiles."

The closest work to the present study appears to be by L. Winkelmann and R. Winkelmann (2008). They identify mean personality traits for each occupation from the German SOEP, and relate these to personality traits of workers in order to find implications for life satisfaction. Whilst they use a much larger sample size than the present study, they do not have the data to compute personality mismatch directly. By predicting the job satisfaction for workers, had they been employed in different occupational areas, and linking this to other results, they conclude that mismatch leads to lower life satisfaction. Their finding matches the overall conclusion of this chapter. However, it may be argued that the present study shows this relationship more directly (at the expense of sample size). I also analyse the impact of personality mismatch on happiness, and wage. Neither of these are addressed in L. Winkelmann and R. Winkelmann (2008).

The rest of this chapter is organised as follows. Section 1.2 presents a simple framework to explain how a personality trait mismatch can cause a reduction in well-being. Section 1.3 describes details of survey design and data collection. Section 1.4 presents and discusses the empirical results. Section 1.5 briefly describes the limitations of the findings. Finally, Section 1.6 summarises and concludes.

1.2 LINKING PERSONALITY MISMATCH TO WELL-BEING

I consider a mechanism that directly maps personality mismatch to some welfare or well-being level, holding other factors constant. Building upon an idea by Brown (2013), suppose that a worker has a vector of K personality traits: $T = \{T_1, T_2, T_3, ..., T_K\}$. For convenience, assume each personality trait k has a value in the trait space $t = [\underline{t}, \bar{t}]$, so that $T_k \in t$. Each worker undertakes a job, which has its own trait vector $J = \{J_1, J_2, J_3, ..., J_K\}$. A *personality mismatch* occurs when $||T - J|| \neq 0$, with $||T - J||$ representing the degree of mismatch.

The significance of matching T_k with J_k is that optimum productive efficiency is achieved by the individual for that job. This can be modelled as follows. For trait k, the worker has a productive efficiency function $P(t)$.[6] $P(t)$ resem-

6 I use the term *productive efficiency* as opposed to just *productivity* to emphasise the fact that a better personality match allows an individual to be more productive for a given level of effort. $P(.)$ can also depend on other parameters, but these are left exogenous.

bles a beta density function, and is maximised when $t = T_k$. However, the realised efficiency of the worker is $P(J_k)$. This means that the worker is producing at peak efficiency (in terms of that trait) when the trait value of the job is the same as the trait value of the worker. If there is a mismatch between a worker's trait and that of the job, then $P(T_k) > P(J_k)$, since the worker has to adjust their behaviour to the requirements of the job. $P(T_k) - P(J_k)$ represents the productive efficiency that is foregone due to the mismatch.

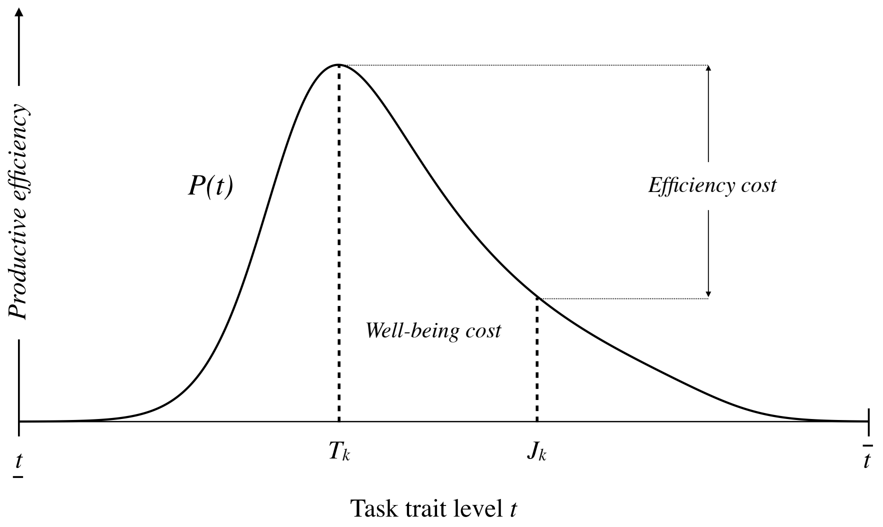


FIGURE 1: Diagram showing how a personality trait mismatch can generate a well-being cost.

A worker with T_k further from J_k must make a greater compromise to their natural tendencies in order to successfully complete their work tasks. This is effortful, and therefore, the worker incurs a psychological adjustment cost. This can be thought of as an experienced utility cost, or a reduction in well-being. The cost is represented by the area under P, bounded by T_k and J_k. More formally, if $M = max\{T_k, J_k\}$ and $m = min\{T_k, J_k\}$, the well-being cost to the worker is given by:

$$Wellbeing\ cost = \int_{m}^{M} P(t)\, dt \tag{1}$$

The cost will, therefore, be larger as M and m are further apart i.e. when there is greater mismatch. The marginal effect of increased mismatch depends on the shape of $P(t)$. Since this cost will be incurred over the duration of employment, one would expect to see this reflected in lower subjective well-being scores reported by the individual. A visual representation is shown in Figure 1. To verify whether this holds empirically, we can test the following hypothesis:

Hypothesis 1.1 *Individuals with a larger personality trait mismatch have lower subjective well-being.*

Further support for this hypothesis comes from self-discrepancy theory (E. T. Higgins, 1987). This theory distinguishes between three different domains of 'self': the *actual* self; the *ideal* self; and the *ought* self. Furthermore, each of these selves may be recognised differently by the individual, in comparison to another person's evaluation of that individual. According to the theory, the nature of the repercussions experienced as a result of a discrepancy between different versions of the self depends upon the types of selves being compared. My definition of mismatch in this chapter compares the perception of an individual's actual self to their perception of the ideal self that another person (i.e. their employer) would want them to be. E. T. Higgins (1987) posits that this type of mismatch would result in emotions pertaining to *dejection*, such as shame, as a result of an expected loss in social esteem. Hence, this should be reflected in measures of well-being, particularly those that measure happiness and life satisfaction. Anxiety and Neuroticism are related to feelings of agitation, and are less likely to be affected by this form of mismatch.

An employer can only reward an individual based on observed productivity, and not on the difference between observed and theoretical efficiency. However, where there are other workers performing the same job, an employer may reward those in the same position who are relatively more efficient. Whilst some of this added efficiency may be attained through ability and skill level, it is possible that firms are also implicitly rewarding workers that have a better personality match with the job with a higher wage. Therefore, we can test a second hypothesis:

Hypothesis 1.2 *Individuals with a larger personality trait mismatch receive a lower wage.*

The most similar work in existing empirical literature is on the relationship between educational mismatch and wages. The generalised finding is that those who are overeducated relative to the requirements of their job earn less than those with identical levels of education, but working in a job that correctly matches their education level (e.g. Bauer, 2002; Budría and Moro-Egido, 2008). Bauer (2002) also finds that including unobserved heterogeneity can eliminate much of this wage difference. For the present study, this means both that educational mismatch is likely to have an impact on wage, and that personality mismatch may account for some of the unobserved heterogeneity that clearly has an impact on wage differentials.

The theoretical ideas in this section are conceptually similar to those surrounding identity (Akerlof and Kranton, 2000). In terms of their framework, we might think of the trait associated with a job as an identity *prescription* for worker behaviour. Unlike their more generalised notion of identity, however, an individual cannot choose their personality. Although some aspects of outward behaviour can be adjusted to fit a given situation, underlying traits are generally stable. This means that any disparity between the identities of the job and worker leads to a loss of utility.[7]

Sackett and Walmsley (2014) make an important distinction between personality as behaviour (I *act*...) and as identity (I *am*...). The former is more malleable, and potentially what an employer is most interested in. However, the latter is more rigid and defines how a person sees themselves. The usage of personality in this chapter is more closely related to this second idea. Whilst an individual with trait level T_k may be able to compensate their behaviour somewhat in order to fit better with the job trait level J_k, they cannot change their core sense of self. It is this friction, I posit, that is likely to result in the individual incurring a cost to their well-being.

7 This loss is referred to I_s in Akerlof and Kranton (2000).

1.3 SURVEY DESIGN

To test these hypotheses, I designed a survey to be administered online, using participants recruited from Amazon Mechanical Turk (MTurk). MTurk is one of the largest crowdsourcing websites on the internet and is widely used to recruit online subjects for academic research. The subject pool is much larger and more representative than standard undergraduate recruitment systems within universities, and results have a high degree of validity (see Mason and Suri, 2012, for a detailed discussion).

Bertrand and Mullainathan (2001) objectively test the use of subjective survey questions for validity. Although they find that data tends to be noisy due to measurement error, they conclude that subjective variables are useful in comparing across individuals (but not within individuals). They also raise concerns about using subjective questions as dependent variables. Since I compare across individuals, a subjective survey should be valid. Furthermore, there is a large body of literature that validates the use of subjective well-being measures in economic research (e.g. Blanchflower and Oswald, 2004b).

I include a combination of scales to measure various aspects of well-being. First, I use the Satisfaction With Life Scale (SWLS), which is a widely used measure of the cognitive aspects of subjective well-being (Diener, Emmons, et al., 1985). These are separate from shorter-term affective aspects of well-being, such as mood (Pavot and Diener, 1993). Each individual component question is summed to give a life satisfaction score between 5 and 35.

Second, to measure aspects of well-being specifically related to job satisfaction, I include two items from the British Workplace Employment Relations Study (WERS), found in their survey of employees.[8] The first of these breaks down job satisfaction into more specific components. The second asks how one's job affects general emotional states. Individuals rate themselves on a five-point scale for each component, ranging from 'very dissatisfied' to 'very satisfied'.

Third, I use two subjective well-being items from the UK Annual Population Survey (APS). The first asks the participant to rate their happiness yesterday; the second asks them to rate their overall life satisfaction (both on a scale from 0-10). The questions have been reframed so that participants rank themselves in

8 Specifically, I use items A8 and A9 from the survey.

relation to 10 other people by clicking on an image which contains a visual representation of 10 people in a row. The results are, therefore, still captured on a 0-10 scale. However, since people find it easier to rank their relative position than make absolute judgements about their condition (Frey and Stutzer, 2005), it should make responses more accurate and consistent. To control for any bias arising from colour, 50% of participants received a similar image, but with the background colours reversed (see Figure 2 for illustration).

(A) Regular colour scale

(B) Reversed colour scale

FIGURE 2: Images used in survey for happiness and life satisfaction relative rank responses. Clicking on the right end of the scale is equivalent to a score of 10/10.

In order to measure personality mismatch between individuals and jobs, I used a 50-item Big Five measure from the International Personality Item Pool (IPIP). This measure is from Goldberg (1992), and has a high degree of correlation with each of the Big Five factors.[9] Participants were asked to rate the accuracy of each item on a 1-7 scale for both themselves and for their perception of an 'ideal' worker for their job. This allows us to directly assess the degree of mismatch, according to the individual's perception. Whilst this is a less powerful measure than the L. Winkelmann and R. Winkelmann (2008) approach of finding mean trait scores for various occupations in a large scale dataset, it is more effective in capturing mismatch at the individual level. Additionally, the approach used in this study does not assume that mean personality in an occupation represents the 'ideal' personality. The order of statements was randomised on a per-subject basis to eliminate any bias.

9 The weakest is Conscientiousness, with an α score of 0.79.

The measure of mismatch used in this study is more sophisticated than subjective fit measures used in job-fit studies, such as Judge and Cable (1997). I measure mismatch on individual personality factors, rather than solely asking people whether they fit well with their jobs.[10] This means that, although individual ratings are subjective, the framework for computing the mismatch (the Big Five) is more rigorous in terms of eliciting traits than a purely self-reported declaration of fit.

I also collect standard demographic information, employment information (such as experience, job title, working hours, and employment status), salary, and information regarding health and marital status, since these are important determinants of well-being (e.g. Frey and Stutzer, 2005; Pavot and Diener, 1993). Artés, Mar Salinas-Jiménez, and Salinas-Jiménez (2013) find that individuals more over-qualified for a specific job will suffer more in terms of subjective well-being, both in terms of life satisfaction and happiness, after controlling for other factors. However, they find that within a given job, those that are under-educated relative to their peers will suffer in terms of well-being - the 'small fish in a big pond' effect. Due to this result, I include the following question: *"What is the minimum level of education required for someone doing your job?"*. This will allow me to control for the impact of educational mismatch on well-being.

As order effects can have a large impact on survey responses, and since personality alone is predictive of well-being, the well-being items are placed at the beginning of the survey. This rules out the possibility of personality questions priming subjects' perceptions about their happiness and life satisfaction.

1.4 RESULTS

1.4.1 *Summary Statistics*

The survey was administered in 3 batches, two in mid-2014, and one in early 2015.[11] 282 responses were obtained overall. The first batch consisted of 97 subjects, each of whom were paid $3 for survey completion. The second batch con-

10 Though I do also ask individuals separately whether they feel they are a good fit for their job.
11 The gap between the first two and the third is due to additional research funding becoming available after the 2014 sample was collected.

sisted of 25 subjects, each paid $2.88. The final batch contained 160 subjects, each paid $4.

I first present some statistics on the data collected to provide the reader with background information about the sample. 220 of the 282 respondents were in full time employment at the time of the survey. Of the remaining 62, only 24 had not been in full time employment within the previous 12 month period. Subjects were employed in a wide variety of occupations and industries. A question at the end of the survey asked respondents how seriously they took the survey, on a scale from 0 to 10. Only three responses reported a seriousness score of less than 7. These three observations were omitted from subsequent regression analyses.

Mean age is 34.5, with a minimum of 20 and maximum of 68.[12] 54.3% of the sample are male. 47.5% are married or living with a partner. Respondents are predominantly U.S. nationals, and 74.8% are white. They are also well educated: approximately 67.7% have at least an undergraduate degree. A breakdown of respondent salaries is shown in Figure 3. The distribution of salaries in the sample is positively skewed, reflecting the pattern observed in U.S. household income distribution data.

Exactly half of the subjects were randomly shown the standard well-being rank question image, whilst the other half were shown the reversed colour version. A t-test comparing mean responses between these two groups for both the happiness and life satisfaction ranking questions yielded p-values of 0.4069 and 0.1778 respectively. Therefore, no significant difference was found between people's subjective well-being ranking when colour was reversed.

Table 1 shows that the three primary well-being measures being used are highly correlated. The weakest correlation is between the SWLS and happiness yesterday rank. This is likely due to the fact that happiness has a stronger correspondence with mood and affect. Since life satisfaction obtained via the ranking task had a stronger correlation with happiness than the SWLS did, it suggests that the elicitation method was important.

12 One respondent did not provide their age, and so was excluded from the majority of the regression analyses.

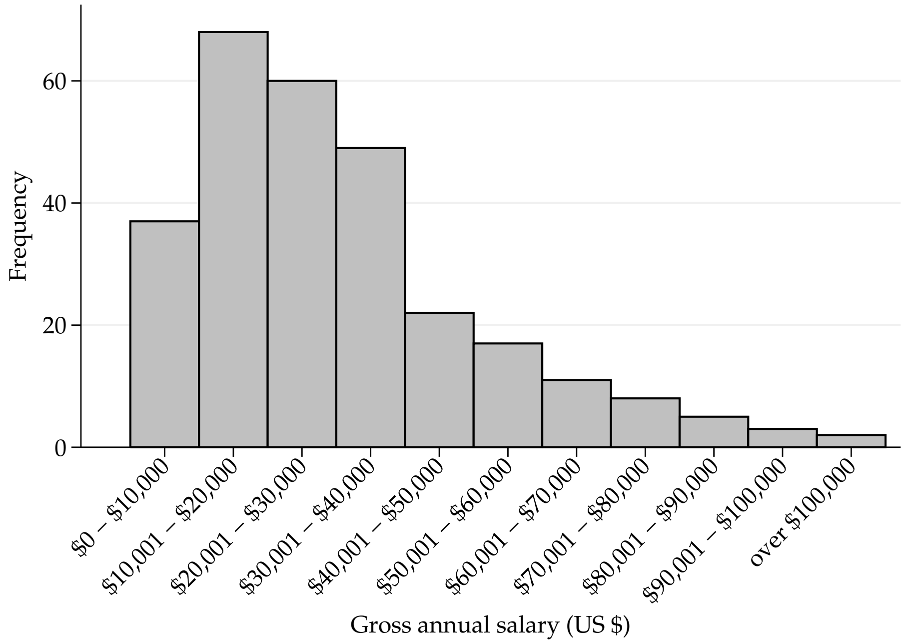


FIGURE 3: Annual salary of survey respondents, in U.S. dollars.

TABLE 1: Correlations between measures of well-being.

	Happy	Life Sat	SWLS
Happiness yesterday (relative rank)	1		
Life satisfaction (relative rank)	0.8667	1	
SWLS	0.6898	0.7727	1

1.4.2 *Personality Mismatch and Well-being*

1.4.2.1 *Measuring Mismatch*

To compute Big Five mismatch, I calculate mean trait scores from the 50 items in the survey for both the individual and their perception of an 'ideal' worker in their position. I create a personality mismatch vector, $|\hat{M}|$, that is the absolute value of the difference between an individual's actual trait score (denoted by subscript a), and their perceived ideal personality for the job (denoted by subscript i):

$$|\hat{M}| = \left| \begin{pmatrix} A_a \\ C_a \\ E_a \\ N_a \\ O_a \end{pmatrix} - \begin{pmatrix} A_i \\ C_i \\ E_i \\ N_i \\ O_i \end{pmatrix} \right| \tag{2}$$

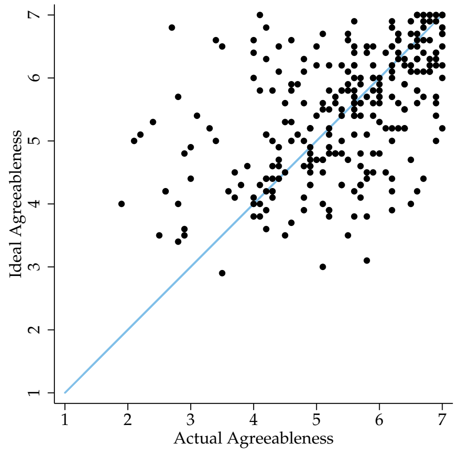
To assess whether the direction of mismatch is relevant for specific traits, I also use the raw difference, $\hat{M}$, for some regressions.

Second, I create a scalar mismatch measure, $\hat{m}$, that is obtained by computing the Euclidean distance between actual and ideal personality vectors:

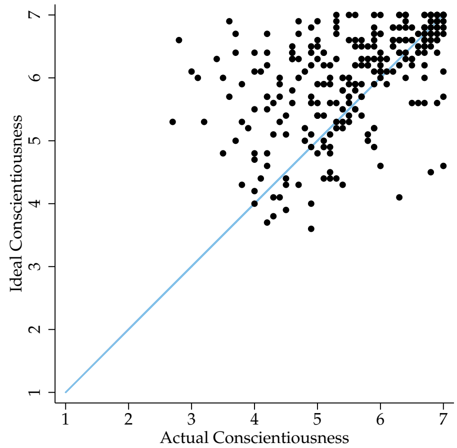
$$\hat{m} = \sqrt{(A_a - A_i)^2 + ... + (O_a - O_i)^2} \tag{3}$$

This provides us with a single, holistic measure of personality mismatch.

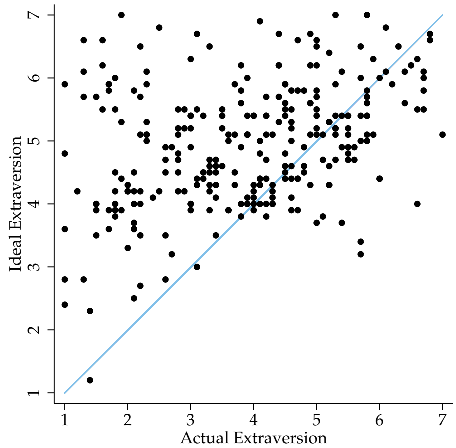
Figure 4 shows plots of Big Five factors for an individual plotted against the ideal personality trait value of their job. The 45-degree line represents a trait mismatch of zero. Visually, one can see that the mismatch is skewed to one direction in the cases of Extraversion and Neuroticism. More people are lacking in Extraversion than being too Extraverted. Furthermore, more people seem to be excessively Neurotic for their jobs than those who are less Neurotic than required.



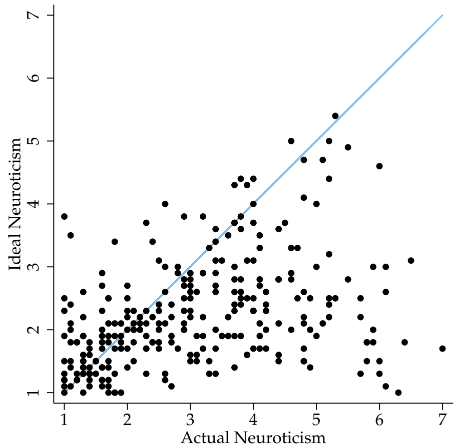
(A) Agreeableness



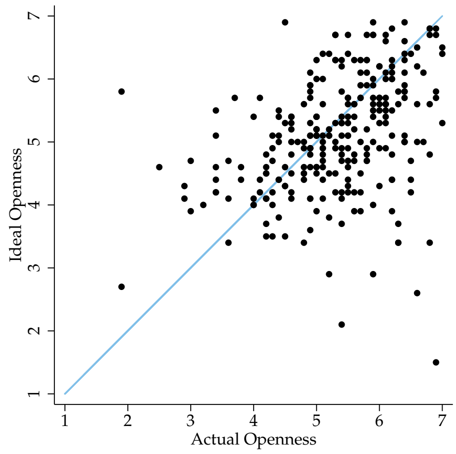
(B) Conscientiousness



(C) Extraversion



(D) Neuroticism



(E) Openness

FIGURE 4: Personality mismatch: actual personality traits plotted against 'ideal' personality traits for each of the Big Five.

We can also use these ideal personality traits to form a picture about relative differences in trait values. Table 2 shows the mean ideal trait reported by individuals for each industrial employment sector. Health Care & Social Assistance jobs require a level of 6 in Agreeableness on a 1-7 scale, whereas jobs in Manufacturing only require 4.67. Jobs in Management require a 1 point higher level of Openness than those in Accommodation and Food Services, and just below 0.9 points higher than jobs in Construction, Manufacturing, Transportation & Warehousing, and Utilities. Similarly, jobs in Arts, Entertainment, and Recreation require at least 1 point more Extraversion than jobs in Administrative Support & Waste Management, Construction, Information, and Manufacturing. These paint a picture of jobs in various industries that would match our intuition somewhat regarding worker stereotypes: the outgoing and charismatic people required in the leisure industry, versus the relatively reserved bricklayer, for example.

Table 3 shows ideal traits for job titles containing a specific keyword. Even though the number of observations is too small to make any conclusive assertions about ideal trait values for each job, the results show quite a surprising amount of precision (evidenced by low standard deviations) given the sample size. Sales jobs unsurprisingly require over 1 point more Extraversion than jobs with "admin" or "research" in their title. In contrast, research jobs require around half a point more Openness than sales jobs. Finally, teaching jobs appear to require particularly high levels of Agreeableness, with a mean requirement of 6.29 out of a possible 7. As with sectoral ideal traits, these seem to correspond well with intuitions and stereotypes about the traits required for particular occupations.

L. Winkelmann and R. Winkelmann (2008) also measure mean personality traits for occupations. Two of their occupations, 'manager' and 'teacher', correspond to job titles measured in Table 3. Both in their SOEP data, and in Table 3: teachers are more Open, less Conscientious, less Extraverted, and more Agreeable. Only the ordering of Neuroticism is different between studies. This is encouraging, as though the sample in this chapter is much smaller, there appears to be a degree of agreement in measured personality.

This evidence points to a potential problem in the labour market. As we see in Table 3, each job appears to require trait values above the midpoint of the

TABLE 2: 'Ideal' Big Five traits for each employment sector

		Ideal Personality Trait for Job									
		Agreeableness		Conscientious		Extraversion		Neuroticism		Openness	
Sector	n	Mean	sd	Mean	sd	Mean	sd	Mean	sd	Mean	sd
Accommodation and Food Services	16	5.26	0.97	6.34	0.70	4.74	1.06	2.23	0.81	4.59	0.68
Administrative and Support and Waste Management	6	4.68	1.06	5.88	0.82	4.28	0.26	2.48	1.14	4.78	0.73
Arts, Entertainment, and Recreation	21	5.79	0.99	6.29	0.67	5.54	0.90	1.91	0.75	5.87	0.88
Construction	5	5.58	0.89	5.88	1.25	4.54	0.74	2.26	1.38	4.72	0.48
Educational Services	23	5.81	1.08	5.95	0.76	4.88	1.23	2.19	0.75	5.27	0.96
Finance and Insurance	22	5.26	0.94	6.05	0.59	4.85	0.93	2.20	0.79	4.95	0.52
Health Care and Social Assistance	27	5.99	0.85	6.30	0.69	5.01	0.91	1.92	0.71	5.14	0.95
Information	28	4.89	0.77	5.56	0.92	4.46	0.99	2.77	0.62	5.11	0.88
Management of Companies and Enterprises	7	5.63	0.82	6.29	0.95	5.10	0.93	2.21	0.98	5.60	1.21
Manufacturing	7	4.67	1.04	5.90	1.24	4.41	1.33	2.67	1.11	4.73	1.26
Other Services (except Public Administration)	26	5.42	1.11	5.85	0.83	4.73	1.07	2.32	0.99	4.87	1.08
Professional, Scientific, and Technical	31	5.24	0.88	5.83	0.97	4.83	0.61	2.64	1.14	5.33	0.77
Public Administration	6	5.50	0.86	6.32	0.78	5.22	0.77	2.08	0.58	5.72	0.75
Real Estate and Rental and Leasing	2	5.75	1.77	6.60	0.57	5.80	0.71	2.05	0.78	5.80	1.27
Retail Trade	36	5.56	0.92	5.99	0.80	5.45	0.85	2.13	0.80	4.83	1.00
Transportation and Warehousing	12	4.92	1.02	6.48	0.43	4.63	1.02	1.70	0.58	4.73	0.73
Utilities	3	5.00	0.53	5.17	0.93	4.60	0.56	2.90	1.25	4.73	0.93
Wholesale Trade	4	5.13	0.51	6.05	1.07	5.15	0.78	2.53	1.66	5.38	0.40
	282	5.34	0.95	6.04	0.83	4.90	0.87	2.29	0.93	5.12	0.86

scale (the reverse being true for Neuroticism). Whilst this corresponds approximately to modal trait values amongst workers in a given population, there is still a significant proportion of workers who have low trait values for Agreeableness, Conscientiousness, Extraversion, and Openness (or high Neuroticism). If very few jobs demand these personalities, then there may be a subset of the working population that will be perennially unemployed (or at least consistently poorly matched). I do not explore this issue further. However, this may help to explain theories that propose the existence of a 'natural' rate of unemployment.

TABLE 3: 'Ideal' Big Five trait scores for popular job titles in the sample

		Ideal Personality Trait Score for Job									
		A		C		E		N		O	
Job title contains:	n	Mean	sd	Mean	sd	Mean	sd	Mean	sd	Mean	sd
Sales	21	5.84	0.81	6.27	0.65	5.73	0.92	1.90	0.56	5.08	0.79
Admin	14	5.39	1.08	6.06	0.65	4.65	0.63	2.27	0.91	5.11	0.71
Research	8	5.45	0.63	5.91	0.63	4.68	0.46	2.44	0.53	5.59	0.78
Manage (excl. sales)	31	5.45	0.96	6.18	0.83	5.25	0.74	2.18	0.90	5.25	0.73
Teach	9	6.29	0.34	6.03	0.58	5.01	0.92	2.10	0.65	5.50	0.78

Note: Each personality trait is scored on a scale from 1 to 7. 4 represents a score at the mid-point.

To compare my mismatch measure with the kind of subjective fit response used in Judge and Cable (1997), I asked respondents to answer the question "Do you feel you are/were a good fit for this job?" on a five-point scale. Mismatch scores for all Big Five traits are negatively correlated with this measure of job fit (Table 4). In other words, people who consider themselves to be a good fit for their job have lower personality trait mismatch. This is in line with what we would expect.

However, the strongest correlation for subjective job fit and individual trait mismatch, is $\rho = -0.39$ for Neuroticism. The scalar mismatch measure $\hat{m}$ is similarly correlated with subjective job fit. These correlations are not close to 1 in absolute value. This suggests that self-reported fit is a broader concept than personality mismatch, though measurement error in both the personality mismatch and fit measures may affect the extent to which this is true. If an individual's perception of fit includes other latent factors, then measuring Big Five mismatch allows us to isolate attention to personality effects.

TABLE 4: Correlations between Big Five mismatch and subjective self-reported job fit.

	SR fit	$\lvert\hat{M}_A\rvert$	$\lvert\hat{M}_C\rvert$	$\lvert\hat{M}_E\rvert$	$\lvert\hat{M}_N\rvert$	$\lvert\hat{M}_O\rvert$	$\hat{m}$
SR job fit	1						
Mismatch in A ($\lvert\hat{M}_A\rvert$)	-0.1811	1					
Mismatch in C ($\lvert\hat{M}_C\rvert$)	-0.2384	0.2904	1				
Mismatch in E ($\lvert\hat{M}_E\rvert$)	-0.2221	0.2785	0.2280	1			
Mismatch in N ($\lvert\hat{M}_N\rvert$)	-0.3893	0.3428	0.3103	0.5026	1		
Mismatch in O ($\lvert\hat{M}_O\rvert$)	-0.3033	0.2602	0.1251	0.1410	0.2499	1	
$\hat{m}$	-0.3991	0.5557	0.4784	0.7956	0.7945	0.4566	1

1.4.2.2 *Job Satisfaction*

First, I consider the effects of personality mismatch on individual aspects of job satisfaction, as measured by 8 items from the WERS. If a personality mismatch between worker and job translates to overall well-being, then at least part of this relationship should act through job satisfaction.

Eight separate ordered logit regressions were run, with each WERS satisfaction measure as the dependent variable. A full set of controls are used. The general specification is as follows:

$$JS_i = \alpha P_i + \beta WAGE_i + \gamma E_i + \theta X_i + \delta YEAR_i + \epsilon_i \tag{4}$$

where P_i represents a vector of raw Big Five, and Big Five mismatch measures; E_i represents a vector of education and educational mismatch; and X_i represents other demographic variables, such as gender. $YEAR_i$ is a dummy variable that has the value 1 if the survey year was 2015 (i.e. the third batch) and 0 otherwise.

I estimate two versions of this regression. The first has a smaller X_i vector - it includes only age, gender, and information about wage, employment status, and working hours. The second can be described as a 'kitchen sink' regression. It includes WERS measures of job affect, as well as both the happiness yesterday and life satisfaction rank measures, in order to determine whether general well-being has a reverse effect on job satisfaction. I also include additional demographic information (such as race) and the general health level of each individual. Dependent variable descriptions can be found in Table 5.

Table 5: Job satisfaction variables key.

	How satisfied are you with:
WERS 1	The sense of achievement you get from your work?
WERS 2	The scope for using your own initiative?
WERS 3	The amount of influence you have over your job?
WERS 4	The training you receive?
WERS 5	The opportunity to develop your skills in your job?
WERS 6	The amount of pay you receive?
WERS 7	Your job security?
WERS 8	The work itself?

Table 6 shows ordered logit estimates for both scalar and vector mismatch under the two different specifications (only one measure of mismatch was included in each group of regressions). We see that coefficients for scalar mismatch are all negative, suggesting a higher level of mismatch is associated with reduced job satisfaction. Furthermore, the estimates are strongly significant, save for the regressions that measure satisfaction with training (WERS 4) and job security (WERS 7). This finding is the first stage in the validation of Hypothesis 1.1. Of the remaining control variables, education level and education mismatch had very strong relationships, more than double the magnitude of personality mismatch in some of the regressions. Surprisingly, salary generally had an insignificant impact on job satisfaction.

All parameter estimates for mismatches in Neuroticism and Openness are negative (and most are significant), whilst estimates for mismatches in Agreeableness, Conscientiousness, and Extraversion have mixed signs. This suggests that mismatches in Neuroticism and Openness account for most of the scalar mismatch relationship with job satisfaction. Whilst significance drops in the second (deliberately overfitted) specification, the signs and many of the estimated magnitudes do not differ substantially from the first specification. Surprisingly, point estimates for Neuroticism mismatch are relatively stable between specifications, even though the second specification includes measures of mood and affect that are likely to overlap with Neuroticism.

In particular, a mismatch in Neuroticism (in both specifications) is associated with significantly lower satisfaction with pay (WERS 6), even after controlling for

Table 6: Job satisfaction is inversely related to personality mismatch.

	Dependent variables are measures of job satisfaction (see Table 5 for variable details).							
	WERS 1	WERS 2	WERS 3	WERS 4	WERS 5	WERS 6	WERS 7	WERS 8
				Specification 1				
Regression group 1								
$\hat{m}$	-0.319***	-0.359***	-0.238**	-0.171*	-0.340***	-0.289***	-0.185*	-0.319***
	(0.103)	(0.0980)	(0.0958)	(0.0966)	(0.1000)	(0.0989)	(0.0998)	(0.101)
Regression group 2								
$\lvert\hat{M}_A\rvert$	-0.13	-0.0833	0.0632	0.0233	-0.313	-0.138	0.0429	-0.0706
	(0.198)	(0.193)	(0.196)	(0.200)	(0.191)	(0.189)	(0.202)	(0.192)
$\lvert\hat{M}_C\rvert$	0.021	0.152	0.161	0.114	0.101	0.293	0.145	0.131
	(0.228)	(0.221)	(0.218)	(0.211)	(0.213)	(0.214)	(0.215)	(0.231)
$\lvert\hat{M}_E\rvert$	0.0156	-0.0253	-0.0397	0.0962	0.166	-0.086	-0.12	-0.0789
	(0.156)	(0.153)	(0.153)	(0.150)	(0.153)	(0.142)	(0.148)	(0.156)
$\lvert\hat{M}_N\rvert$	-0.237	-0.326*	-0.292	-0.28	-0.400**	-0.469***†	-0.223	-0.258
	(0.188)	(0.181)	(0.184)	(0.179)	(0.184)	(0.176)	(0.174)	(0.188)
$\lvert\hat{M}_O\rvert$	-0.444**	-0.476***†	-0.424**	-0.469***†	-0.424**	-0.168	-0.201	-0.507***†
	(0.181)	(0.175)	(0.172)	(0.178)	(0.174)	(0.166)	(0.165)	(0.175)
Regression group 3								
Job fit	1.196***	1.446***	1.264***	0.630***	1.032***	0.480***	0.826***	1.681***
	(0.167)	(0.170)	(0.166)	(0.149)	(0.158)	(0.139)	(0.150)	(0.186)
				Specification 2 ('kitchen sink')				
Regression group 4								
$\hat{m}$	-0.235**	-0.320***	-0.207**	-0.0999	-0.288***	-0.235**	-0.0964	-0.263**
	(0.108)	(0.104)	(0.101)	(0.101)	(0.104)	(0.102)	(0.103)	(0.108)
Regression group 5								
$\lvert\hat{M}_A\rvert$	-0.151	-0.00272	0.0425	0.0802	-0.263	-0.0816	-0.0265	-0.0169
	(0.208)	(0.204)	(0.205)	(0.205)	(0.203)	(0.200)	(0.210)	(0.206)
$\lvert\hat{M}_C\rvert$	-0.259	-0.0684	-0.032	0.0201	-0.114	0.156	0.00644	-0.0597
	(0.249)	(0.237)	(0.234)	(0.233)	(0.233)	(0.231)	(0.235)	(0.248)
$\lvert\hat{M}_E\rvert$	0.107	0.00109	0.0356	0.185	0.21	-0.0414	-0.00911	-0.00918
	(0.166)	(0.164)	(0.159)	(0.157)	(0.161)	(0.149)	(0.154)	(0.167)
$\lvert\hat{M}_N\rvert$	-0.118	-0.336*	-0.313	-0.297	-0.350*	-0.498***†	-0.143	-0.327
	(0.199)	(0.194)	(0.192)	(0.188)	(0.193)	(0.187)	(0.190)	(0.201)
$\lvert\hat{M}_O\rvert$	-0.227	-0.340*	-0.276	-0.342*	-0.326*	-0.0298	-0.0914	-0.265
	(0.188)	(0.182)	(0.179)	(0.183)	(0.181)	(0.176)	(0.180)	(0.187)

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1, † p<0.05 after Bonferroni correction. n=278.

Notes: These are mismatch estimates from 5 groups of ordered logit regressions. Specification 1 includes controls for raw personality, age, age-squared, gender, employment status, education, education mismatch, working hours, and salary. Specification 2 adds further controls for mood, well-being, health, marital status, and race. Bonferroni corrections are applied to regression groups 2 and 5 to control for multiple hypotheses. These represent conservative lower bounds for significance.

actual salary. The coefficient for salary was equally significant, but with magnitudes of 0.27 and 0.257 in the first and second specifications, respectively. This is only approximately 55% of the estimates for Neuroticism mismatch. Therefore, satisfaction with pay is more strongly associated with personality trait mismatch than with pay itself, by nearly a factor of 2. This suggests that relative incomes, personal characteristics, or personal expectations about one's earning potential (Rutledge et al., 2014), are more important than absolute earnings in determining satisfaction with pay.

For regression groups 2 and 5 in Table 6, I apply Bonferroni corrections to adjust for multiple comparisons. There are five hypotheses corresponding to each of the Big Five mismatch variables if we are testing for the impact of personality mismatch as a whole on job satisfaction. For joint significance at the $\alpha = 0.05$ level, we require a corrected p-value that is below 0.01. Therefore, any mismatch coefficient that is significant at the $\alpha = 0.01$ level is also jointly significant at least at the $\alpha = 0.05$ level. The Bonferroni correction is conservative as a correction for multiple comparisons. It is likely to be even more conservative in this case, because the five tests are not independent (Abdi, 2007). Therefore, these tests of joint significance represent 'worst case' lower bounds. Despite this, the strong association between Neuroticism mismatch and satisfaction with pay is significant in both specifications after correcting for multiple hypotheses.

Taking the mean of scalar mismatch estimates from Specification 1 gives us an average coefficient for mismatch on job satisfaction of -0.278. All other things constant, a one point increase in scalar mismatch means that the predicted odds of being 'very satisfied', relative to the set of all four lower satisfaction categories, changes by a factor of 0.758. In other words, one is only about 76% as likely to report being 'very satisfied' if personality mismatch increases by one point.

Point estimates for personality mismatch are, in general, much greater in magnitude than those for salary. However, education level and education mismatch were much more strongly related to job satisfaction than personality or personality mismatch.

Finally, it should be noted that when self-reported job fit was substituted for scalar personality mismatch in specification 1, it dominated all other estimates. When including both self-reported fit and scalar mismatch, the coefficients for personality mismatch were smaller in general, but still significant. Therefore, personality mismatch appears to be a key determinant of job satisfaction levels. However, there are other aspects of job fit that are clearly not personality related.

1.4.2.3 *Subjective Well-being*

We have established that a mismatch in personality is related to lower job satisfaction. To answer the main question of this chapter, we need to determine whether this relationship holds for subjective well-being.

The general specification of the regression to be estimated is given by (5), with W representing a well-being measure. Again, P represents raw personality and personality mismatch, E represents education and educational mismatch, and X captures a series of additional controls. This equation captures the three main channels that determine happiness and well-being, as observed by Easterlin (2005) - personality, health, and standard of living (represented by wage).

$$
\begin{aligned}
W_i = & \beta_0 + \beta_1 P_i + \beta_2 HEALTH_i + \beta_3 WAGE_i + \gamma E_i + \delta YEAR_i \\
& + \theta X_i + \epsilon_i
\end{aligned}
\tag{5}
$$

Whilst Ordinary Least Squares (OLS) is often used to estimate well-being equations, coefficient estimates will be inconsistent since one cannot assume that the distances between response categories are perceived as being the same. In the present case, one would expect this to be more of a problem when using the SWLS than with the rank-based measures, which are designed to make it easy for an individual to rank themselves on an evenly-spaced scale. To estimate the regression equations, therefore, I used both ordered logit and OLS approaches.[13] However, since both approaches gave very similar results, I report only the estimates from OLS in order to allow for more intuitive interpretation. Regular OLS standard errors were computed for all regressions, as Breusch-Pagan and Cook-Weisberg tests did not find evidence of heteroskedasticity.

---

13 An ordered probit model gives qualitatively very similar results, and is often preferred in economics research. However, there is no theoretical reason for us to prefer a probit specification i.e. we would not expect well-being *a priori* to be dependent on a latent normally distributed random variable.

Table 7: Relationship between Big Five mismatch and subjective well-being, with no controls.

	Dependent variable:								
	Happiness	Life Satis	SWLS	Happiness	Life Satis	SWLS	Happiness	Life Satis	SWLS
$\hat{m}$	-0.556***	-0.619***	-2.443***	-	-	-	-	-	-
$\lvert\hat{M}_A\rvert$	-	-	-	0.0681	0.11	0.155	-	-	-
$\lvert\hat{M}_C\rvert$	-	-	-	-0.179	-0.251	-1.533**	-	-	-
$\lvert\hat{M}_E\rvert$	-	-	-	-0.0815	-0.168	-0.968**	-	-	-
$\lvert\hat{M}_N\rvert$	-	-	-	-0.813***	-0.836***	-2.374***	-	-	-
$\lvert\hat{M}_O\rvert$	-	-	-	-0.0774	0.102	-0.471	-	-	-
$\hat{M}_A$	-	-	-	-	-	-	0.135	-0.0528	0.435
$\hat{M}_C$	-	-	-	-	-	-	-0.135	0.185	0.727
$\hat{M}_E$	-	-	-	-	-	-	-0.0154	0.138	0.817**
$\hat{M}_N$	-	-	-	-	-	-	-0.792***	-0.733***	-2.025***
$\hat{M}_O$	-	-	-	-	-	-	-0.144	-0.152	-1.207***
Constant	7.442***	7.399***	28.09***	7.151***	6.972***	27.07***	6.605***	6.655***	24.88***
Observations	279	279	279	279	279	279	279	279	279
R^2	0.139	0.152	0.235	0.191	0.192	0.253	0.213	0.211	0.248

Standard errors omitted for brevity. *** p<0.01, ** p<0.05, * p<0.1.

Table 7 shows the raw relationship between personality mismatch and three different measures of subjective well-being. Without including control variables, we see that an increase in scalar mismatch by one point corresponds with just over a half-point reduction in life satisfaction and happiness yesterday rankings. When we separate mismatch for each Big Five factor, only mismatch in Neuroticism has a consistently significant relationship with all three measures of well-being. Mismatches in Extraversion and Conscientiousness are only significant in the SWLS regression, but they have consistently negative associations with all three well-being measures. The SWLS measure contains more items than the single life satisfaction ranking task, and so it may be that this measure is able to capture a wider gamut of life satisfaction determinants. The negative relationship between personality mismatch and life satisfaction supports the finding obtained by L. Winkelmann and R. Winkelmann (2008).

The third group of regressions uses the $\hat{M}$ measures of mismatch (i.e. without taking the absolute value). These show that being too Extraverted for your job is associated with significantly *higher* life satisfaction.[14] The same relationship holds for being too Conscientious, however estimates were not statistically significant in this case. Being too Neurotic or too Open is associated with lower levels of *all* well-being measures.

Adding control variables, Table 8 shows the results of the regressions with each of the three measures of well-being as dependent variables. The strongest covariate with happiness yesterday and both life satisfaction measures is the self-assessed health level of the individual. Healthier people, unsurprisingly, report higher subjective well-being. Neither salary, gender, race, education, nor employment status are strongly associated with well-being. This seems to support the theory (and previous evidence) that suggests relative income is a stronger determinant of well-being than absolute income (e.g. Blanchflower and Oswald, 2004b).[15] We can also see that a U-shape in life satisfaction is evident over the life cycle, in accordance with a number of studies.

---

14 This is somewhat surprising, considering the theoretical discussion in Section 1.2 suggests that being mismatched in either direction should result in a well-being loss.

15 Results from the ordered logit versions of these regressions give largely the same conclusion, though OLS may have underestimated the positive effects of marriage/cohabitation on happiness.

TABLE 8: The relationship between Big Five mismatch and subjective well-being, with full set of controls.

	Dependent Variable								
	Happiness	Life Satis	SWLS	Happiness	Life Satis	SWLS	Happiness	Life Satis	SWLS
$\hat{m}$	-0.152	-0.218*	-1.131***	-	-	-	-	-	-
$\lvert\hat{M}_A\rvert$	-	-	-	-0.0423	-0.0134	0.112	-	-	-
$\lvert\hat{M}_C\rvert$	-	-	-	-0.0444	0.191	-1.132	-	-	-
$\lvert\hat{M}_E\rvert$	-	-	-	-0.0872	-0.184	-0.777	-	-	-
$\lvert\hat{M}_N\rvert$	-	-	-	-0.109	-0.277	-0.616	-	-	-
$\lvert\hat{M}_O\rvert$	-	-	-	-0.0878	0.0733	-0.186	-	-	-
$\hat{M}_A$	-	-	-	-	-	-	0.196	-0.0718	0.267
$\hat{M}_C$	-	-	-	-	-	-	-0.0279	-0.0311	0.928
$\hat{M}_E$	-	-	-	-	-	-	-0.204	0.0695	0.58
$\hat{M}_N$	-	-	-	-	-	-	-0.336	-0.477*	-0.307
$\hat{M}_O$	-	-	-	-	-	-	-0.221	-0.125	-1.109*
Agreeableness	0.235*	0.0541	0.206	0.233	0.0896	0.354	0.0732	0.131	0.0958
Conscientiousness	-0.0605	0.202	0.46	-0.0498	0.358*	0.166	0.0361	0.317	0.0148
Extraversion	-0.0148	0.0095	0.0646	-0.0106	-0.0211	0.0773	0.202	0.0166	0.0665
Neuroticism	-0.599***	-0.550***	-1.499***	-0.580***	-0.460***	-1.472***	-0.387*	-0.238	-1.638**
Openness	0.0217	-0.0849	-0.449	0.016	-0.128	-0.587	0.143	-0.0185	0.191
Health level	0.742***	0.766***	1.764***	0.747***	0.769***	1.741***	0.743***	0.770***	1.716***
Salary	-0.00965	0.0522	0.400*	-0.00719	0.0673	0.415*	-0.00527	0.0758	0.461**
Age	0.00193	-0.0921	-0.690***	-0.000858	-0.103	-0.730***	-0.0132	-0.117	-0.713***
Age^2	-0.000229	0.000719	0.00618**	-0.000197	0.000856	0.00666**	-2.85e-05	0.00102	0.00645**
Male	0.11	-0.221	-0.117	0.101	-0.191	0.00721	0.0636	-0.219	0.13
Married/cohabiting	0.511	0.422	2.698***	0.504	0.421	2.940***	0.53*	0.457	2.854***
Have children?	-0.729**	-0.194	1.523	-0.727**	-0.234	1.374	-0.849**	-0.296	1.1
White	-0.0519	-0.0812	1.218	-0.05	-0.0358	1.222	0.039	0.0516	1.468
Is religious?	0.378	0.12	0.00156	0.385	0.12	0.0135	0.331	0.133	0.0799
Education level	-0.119	-0.106	1.072	-0.122	-0.0872	1.158*	-0.167	-0.158	0.972
Education mismatch	0.0996	0.255	-0.457	0.104	0.187	-0.635	0.204	0.347	-0.275
Work hours /wk	-0.11	-0.0438	-0.0935	-0.109	-0.0325	-0.0575	-0.114	-0.0466	-0.101
Is in FTE?	0.106	-0.0292	0.612	0.111	-0.0865	0.323	0.171	-0.0672	0.542
Year Dummy	-0.231	-0.281	-1.01	-0.235	-0.285	-1.076	-0.238	-0.314	-1.427*
Constant	5.770***	6.878***	30.57***	5.676**	5.714**	32.15***	3.997*	4.849*	29.97***
Observations	278	278	278	278	278	278	278	278	278
R^2	0.378	0.342	0.46	0.378	0.344	0.46	0.388	0.349	0.453

Standard errors omitted for brevity. *** p<0.01, ** p<0.05, * p<0.1.

Now that raw personality trait scores have been controlled for, we see that the strongly significant coefficient for a mismatch in Neuroticism in Table 7 was likely picking up the fact that more Neurotic people in general rate themselves lower on all three well-being measures. This is consistent with the finding that personality traits themselves can explain a large portion of the variance in subjective well-being (Steel, Schmidt, and Shultz, 2008). These results also support previous research that identifies Neuroticism as the strongest personality predictor for life satisfaction and happiness, as well as Conscientiousness having the strongest *positive* correlation with life satisfaction (DeNeve and Cooper, 1998). Extraversion and Neuroticism are the two Big Five factors for which the largest body of underlying theory exists. Extraversion is closely related to the tendency to experience positive affect, and Neuroticism to negative affect (e.g. McCrae and John, 1992). The finding that raw Neuroticism is related to lower well-being is therefore not surprising, and mirrors early findings from Costa and McCrae (1980). However, according to this research, we would also expect raw Extraversion to be positively related with well-being, which is not consistently the case in the present analysis.

Scalar personality mismatch is significantly related to well-being, but more so with life satisfaction than happiness. The coefficient is now smaller in magnitude, owing to variance being captured by Neuroticism and health level. However, a one-point mismatch is still associated with a 0.15 position reduction in happiness ranking, and a 0.22 position reduction in life satisfaction ranking. To put this into perspective, the mean scalar mismatch score in the sample is 2.58. For an individual with this mean level of mismatch, they would rate themselves 0.56 of a ranking position lower on life satisfaction than someone with no mismatch.

When the trait-separated measures of mismatch are used, we see that although Neuroticism mismatch has the strongest relationship across all three measures of well-being of the five factors, coefficient estimates are of low statistical significance. Using the non-absolute mismatch measures does suggest that individuals suffer from almost half a position reduction in life satisfaction for each excess point in Neuroticism. Furthermore, there is weak evidence suggesting that, as we saw with the raw correlations in Table 7, excess Extraversion may be beneficial for life satisfaction, but the opposite is true for excess Openness.

Since we established that personality mismatch is related to job satisfaction in Section 1.4.2.2, I now test whether it can explain any portion of the variance in well-being through a channel other than job satisfaction. Intuitively, a personality mismatch at work may have implications for general physical and mental well-being, which are not realised exclusively in the workplace.

Table 9 shows that job satisfaction is indeed important for general well-being. In particular, one's sense of achievement from work, and skill development opportunities are strongly related to higher well-being. However, there is still a significant coefficient for scalar mismatch on life satisfaction. For the SWLS regression, adding job satisfaction variables has only lowered the mismatch coefficient by 0.264 in absolute terms, and still $p < 0.01$. This suggests that personality mismatch has negative implications for well-being outside the workplace.

Finally, Table 10 repeats the regressions from Tables 8 and 9, but using self-reported job-fit instead of personality mismatch. For the regressions exclusive of job satisfaction variables, job-fit has a smaller relationship (in absolute terms) with both measures of life satisfaction. Furthermore, the coefficients for job-fit in the life satisfaction regressions are not strongly significant. However, the absolute relationship between happiness and job-fit is nearly twice as strong as that of happiness and scalar personality mismatch. A similar conclusion can be drawn when job satisfaction variables are included.

This finding suggests that although self-reported fit is more useful than personality mismatch in accounting for an individual's level of job satisfaction, it is less useful in accounting for general life satisfaction. This again implies that a personality mismatch at work may have more long-term implications for well-being.

1.4.3 *Personality Mismatch and Wage*

Section 1.2 describes how a worker who has a larger personality mismatch is likely to be less productive. Although the employer may not realise this at the time of employment, over a longer period, we might expect the wage of these mismatched individuals to be lower than those better matched to their jobs.

Classical earnings regression specifications, such as Mincer (1974), measure earnings by taking into account the effects of schooling and work experience. I

TABLE 9: Personality mismatch is related to life satisfaction not only through job satisfaction.

	Dependent variable:					
	Happiness	Life Satis	SWLS	Happiness	Life Satis	SWLS
$\hat{m}$	-0.124	-0.182	-0.867***	-	-	-
$\lvert\hat{M}_A\rvert$	-	-	-	-0.00435	0.0505	0.426
$\lvert\hat{M}_C\rvert$	-	-	-	-0.0496	0.171	-1.247*
$\lvert\hat{M}_E\rvert$	-	-	-	-0.106	-0.224	-0.896*
$\lvert\hat{M}_N\rvert$	-	-	-	-0.0689	-0.202	-0.181
$\lvert\hat{M}_O\rvert$	-	-	-	-0.08	0.0877	-0.0291
Agreeableness	0.203	0.00181	-0.0295	0.215	0.0546	0.2
Conscientiousness	-0.0722	0.191	0.489	-0.068	0.328*	0.0723
Extraversion	-0.0299	-0.0192	0.0652	-0.0439	-0.0792	-0.0585
Neuroticism	-0.580***	-0.553***	-1.494***	-0.574***	-0.482***	-1.592***
Openness	0.0666	-0.021	-0.42	0.0582	-0.0709	-0.588
Health level	0.771***	0.831***	1.659***	0.769***	0.821***	1.557***
Salary	-0.0298	0.026	0.259	-0.0285	0.0424	0.258
Age	0.0161	-0.0795	-0.641**	0.0125	-0.0917	-0.680**
Age^2	-0.000341	0.000643	0.00568*	-0.0003	0.000799	0.00615**
Male	0.0306	-0.408	-0.641	0.0344	-0.351	-0.401
Married/cohabiting	0.438	0.281	2.139**	0.437	0.289	2.421**
Have children?	-0.723**	-0.119	1.509	-0.724**	-0.158	1.345
White	0.00234	0.0669	1.27	8.03e-05	0.0993	1.22
Is religious?	0.411	0.206	-0.000801	0.409	0.194	-0.0834
Education level	-0.146	-0.128	0.791	-0.143	-0.115	0.888
Education mismatch	0.102	0.217	-0.104	0.101	0.153	-0.296
Working hours per wk	-0.111	-0.0441	-0.171	-0.109	-0.033	-0.14
Is in FTE?	0.153	0.0521	1.499	0.145	-0.0239	1.113
Year Dummy	-0.194	-0.249	-0.891	-0.204	-0.262	-1.013
WERS 1	0.398*	0.495**	1.398**	0.404*	0.523**	1.421**
WERS 2	-0.163	-0.362*	-0.579	-0.158	-0.341	-0.493
WERS 3	-0.0231	0.102	0.299	-0.0275	0.0909	0.291
WERS 4	0.0559	0.114	-0.787	0.052	0.127	-0.804
WERS 5	0.0976	0.339*	1.303**	0.103	0.352*	1.435***
WERS 6	0.104	0.177	0.904**	0.105	0.167	0.938**
WERS 7	0.147	0.189	0.366	0.148	0.179	0.386
WERS 8	-0.362*	-0.718***	-1.325**	-0.370*	-0.719***	-1.340**
Constant	4.575**	5.452**	26.70***	4.621**	4.477*	29.24***
Observations	278	278	278	278	278	278
R^2	0.401	0.404	0.508	0.401	0.407	0.514

Standard errors omitted for brevity. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 10: Job-fit is less significantly related to life satisfaction than personality mismatch

	Dependent Variable:					
	Happiness	Life Satis	SWLS	Happiness	Life Satis	SWLS
Self reported job-fit	0.319**	0.207	0.801*	0.468**	0.347*	0.328
Agreeableness	0.229	0.0451	0.158	0.217	0.00235	-0.103
Conscientiousness	-0.0939	0.191	0.446	-0.105	0.173	0.52
Extraversion	0.0577	0.115	0.614*	0.0353	0.0712	0.476
Neuroticism	-0.586***	-0.583***	-1.723***	-0.576***	-0.575***	-1.706***
Openness	-0.0443	-0.167	-0.862*	0.00114	-0.0975	-0.709
Health level	0.721***	0.765***	1.790***	0.760***	0.826***	1.678***
Salary	-0.0094	0.0583	0.438**	-0.033	0.0268	0.284
Age	-0.00551	-0.101	-0.733***	0.00376	-0.0904	-0.666***
Age^2	-0.000154	0.000813	0.00665**	-0.000206	0.000764	0.00596*
Male	0.157	-0.17	0.127	0.0606	-0.373	-0.505
Married/cohabiting	0.543*	0.466	2.927***	0.48	0.32	2.240**
Have children?	-0.760**	-0.218	1.418	-0.760**	-0.145	1.49
White	-0.108	-0.104	1.161	-0.0325	0.0475	1.301
Is religious?	0.32	0.114	0.0502	0.363	0.191	0.142
Education level	-0.131	-0.107	1.084	-0.106	-0.101	0.796
Education mismatch	0.159	0.288	-0.342	0.124	0.24	-0.0328
Working hours per wk	-0.108	-0.0372	-0.0544	-0.11	-0.0412	-0.15
Is in FTE?	0.156	0.0142	0.806	0.229	0.13	1.74
Year Dummy	-0.28	-0.344	-1.324	-0.244	-0.306	-1.099
WERS 1	-	-	-	0.413**	0.511**	1.450**
WERS 2	-	-	-	-0.226	-0.393*	-0.488
WERS 3	-	-	-	-0.0505	0.0772	0.24
WERS 4	-	-	-	0.0524	0.111	-0.794
WERS 5	-	-	-	0.0823	0.340*	1.403***
WERS 6	-	-	-	0.132	0.205	0.984**
WERS 7	-	-	-	0.109	0.161	0.344
WERS 8	-	-	-	-0.513**	-0.824***	-1.385**
Constant	4.579**	5.783**	25.54***	3.652*	4.438*	23.19***
Observations	278	278	278	278	278	278
R^2	0.383	0.336	0.441	0.411	0.404	0.495

Standard errors omitted for brevity. *** p<0.01, ** p<0.05, * p<0.1.

use this as a starting point from which to add additional explanatory variables. Since data on total years of work experience were only collected in the final sample, duration of employment at current/most recent job is used as a suitable alternative to capture the variation in wage accounted for by experience. This allows for the full sample to be used.[16]

As is often the case in earnings regressions, ability is unobservable and there is no true measure available to us to control for this. Heckman, Lochner, and Todd (2003) explain that the importance of ability bias is still a point of contention in economics. The Big Five factor Openness is closely tied to intellect and creativity, and so its inclusion may account for at least some of the individual difference in ability.

Table 11 shows the results of the wage regressions, estimated using ordered logit.[17] In regressions (1) and (2), only a minimal set of control variables is included. Additional controls are added for regressions (3) and (4).

There are a number of interesting findings from Table 11. First, we see that raw personality traits are related to wages, as we might expect from previous work by Nyhus and Pons (2005). I find, as they do, that Neuroticism is negatively associated with earnings. However, in contrast to their findings, Conscientiousness appears to be associated with a *lower* wage. The reason for this is not clear, however, it is likely to be a feature of the specific sample obtained for this study.

Second, a mismatch in Conscientiousness is significantly negatively related to wage in regressions (2) and (4). The odds of being in a higher salary band are significantly reduced with a one-point increase in Conscientiousness mismatch. This reduction outweighs the positive impact on wages that on-the-job experience brings by more than a factor of two. An OLS estimate of this coefficient tells us that a one point increase in Conscientiousness mismatch reduces an individual's salary by over a third of a band (each band being approximately $10,000).[18]

Finally, education has the strongest observable relationship with earnings, as we would expect. However, an educational mismatch is also very strongly related with earnings. In the sample, only 1.43% are undereducated for their

16 It was found that duration of employment at the most recent job was more relevant in determining current wage than work experience. These results are available upon request.

17 The reason for using ordered logit in this case is that the top salary band (over $100,000) is disproportionate in size in comparison to the other bands, which have a fixed width of $10,000.

18 The direction of Conscientiousness mismatch was not important (data available upon request).

Table 11: Big Five mismatch and its relation to wages (ordered logit).

	Dependent Variable: Gross annual salary level			
	(1)	(2)	(3)	(4)
$\hat{m}$	-0.152	-	-0.105	-
$\lvert\hat{M}_A\rvert$	-	-0.0952	-	-0.0983
$\lvert\hat{M}_C\rvert$	-	-0.574***	-	-0.488**
$\lvert\hat{M}_E\rvert$	-	-0.0115	-	0.108
$\lvert\hat{M}_N\rvert$	-	0.21	-	0.0537
$\lvert\hat{M}_O\rvert$	-	-0.175	-	-0.0268
Agreeableness	-0.0652	-0.12	-0.0344	-0.0819
Conscientiousness	-0.107	-0.380**	-0.152	-0.387**
Extraversion	0.0878	0.177	0.046	0.163
Neuroticism	-0.313***	-0.464***	-0.286***	-0.347**
Openness	0.19	0.216	0.269*	0.286*
Health Level	-	-	0.0886	0.104
Age	0.106	0.114	0.0227	0.0207
Age^2	-0.00145	-0.00157*	-0.000372	-0.000362
Male	0.405*	0.379	0.204	0.161
Married/cohabiting	-	-	0.46	0.49
Have children?	-	-	-0.0666	-0.0907
White	-0.00226	-0.0447	0.0449	0.035
Is religious?	-	-	-0.897***	-0.858***
Education level	0.650***	0.675***	1.176***	1.184***
Education mismatch	-	-	-0.955***	-0.932***
Working hours per wk	-	-	0.221***	0.212***
Is in FTE?	-	-	0.973***	0.944**
Time in job	0.194***	0.201***	0.200***	0.209***
Year Dummy	0.332	0.302	0.429*	0.454*
Industry Dummies?	No	No	Yes	Yes
Observations	278	278	278	278

Cut constants and standard errors are omitted for brevity.

*** p<0.01, ** p<0.05, * p<0.1.

job, as opposed to 37.63% who are overeducated. Whilst this is an issue outside the scope of this chapter, we may be concerned by this result, as it suggests either a shortage of jobs for high-skilled people, or simply that workers are over-investing in education.

In regressions (1) and (3), (where scalar mismatch was included, rather than trait-level mismatch) although mismatch had a negative relationship with earnings, the magnitudes of this relationship were not large. Examining the trait mismatch estimates in (2) and (4), we see that this is due to a *positive* association between Neuroticism mismatch and earnings. Since higher Neuroticism is linked to lower earnings, we might suspect that this mismatch effect is because most people in the sample are less Neurotic than a job requires (and this could be somewhat desirable). However, Figure 4 shows that we observe the *opposite* in the raw data. More people have excess levels of Neuroticism than those who are less Neurotic than required. Although the coefficient was not significant, this is a somewhat puzzling finding.

1.5 LIMITATIONS

Whilst limitations of individual findings are discussed in previous sections, here I briefly summarise some of the general limitations with this study. First, although every effort has been made to include confounding factors, the chain of causality is difficult to establish. An instrumental variables approach is difficult to apply to the Big Five. In theory, the Big Five factors are supposed to represent primitive aspects of human personality. To find reasonable correlates for each factor, as well as having these uncorrelated to the dependent variable, is therefore difficult to achieve.

Due to the novel measures of personality mismatch used, the sample size is relatively small and cross-sectional. A longitudinal and representative version of these data would be ideal in theory in order to obtain clearer causality, although one would have to be careful of attrition and other endogeneity issues. Longitudinal data would also allow us to control for personality change, if future research suggests personality traits are less stable than currently appears to be the case.

1.6 CONCLUSION

The main findings of this chapter can be summarised as follows. First, personality mismatch between a worker and their job has a strong negative relationship with their level of job satisfaction. When traits are separated out, mismatches in Neuroticism and Openness have the strongest relationships with job satisfaction. In particular, personality mismatch is more strongly related to satisfaction with pay than actual salary.

Second, personality mismatch has a significantly negative association with life satisfaction. This relationship holds even when controlling for job satisfaction, suggesting that a mismatch in personality may be harming a worker even outside the work environment. An individual with a mean level of personality mismatch places approximately 0.5 lower on a 0-10 scale for life satisfaction. In addition, there is some weaker evidence that suggests being too Open or Neurotic lowers life satisfaction, but being too Extraverted may actually prove to be beneficial. Self-reported job fit is not as strongly related to life satisfaction as personality mismatch. Therefore, personality mismatch appears to provide us with a more holistic metric for measuring long-term well-being implications of job fit.

Third, a mismatch in Conscientiousness has a negative relationship with annual earnings. The magnitude of this relationship is more than double the impact of the time an individual had been working at their job. This suggests that personality trait mismatch can completely offset the positive effect that work experience has on salary. Hence, individuals with a higher level of personality mismatch are also likely to be less well-off.

This study highlights the need for further attention to be given to personality factors and optimal personality matching in the labour market. More generally, it stresses the importance of psychological measures, relative to classical observed variables used commonly in labour economics. Previous research has highlighted the severe impact that unemployment has on well-being. The findings in this chapter suggest that personality mismatch has a severe impact on well-being even for those in employment.[19]

19 There is also a distinct possibility that mismatched individuals are more likely to become unemployed. If this is found to be true, then addressing the problem of personality mismatch may help to mitigate against job attrition. This is left for future work.

Even though poor job fit has been shown to be related to lower job satisfaction, *fit does not correlate with job choice* (Judge and Cable, 1997). Therefore, a simple and potent policy would be to educate younger individuals about the value of personality matching when deciding upon a career path. This is likely to reduce the cost incurred as a result of investing time and effort in a pursuit that may leave them less satisfied with their lives overall, as well as reducing their lifetime income.

2

THE DETERMINANTS OF WELL-BEING PRIORITISATION OVER THE LIFE CYCLE

Recently, a novel attempt has been made to estimate priorities for the different aspects of subjective well-being, in order to understand where resources might best be allocated. However, the determinants of, and life cycle trends for prioritisation have yet to be studied. This chapter - the first to study these issues - finds no consistent (cross-sectional) evidence of variation in priorities over the life cycle, unlike the 'mid-life crisis' observed for levels. Life satisfaction is the most valued aspect of well-being throughout life. However, people overestimate the value placed by others on happiness. Well-being priorities are strongly influenced by well-being levels, and individual fixed effects such as personality, health level, and smoking frequency. The separation of aspects into cognitive and affective factors may provide additional insight into how individuals generate priorities, and hence inform the optimal targeting of policy.

2.1 INTRODUCTION

After decades of focus by economists on improving incomes and production, more recent work has highlighted that increasing income past a certain level does not necessarily translate to any marked improvement in an individual's level of subjective well-being (though this finding is still subject to debate, e.g. Stevenson and Wolfers, 2008, 2013). When examining time series data, mean happiness remains unchanged even when income increases (e.g. Easterlin, 2005). Subsequent research has uncovered important determinants of subjective well-being, such as relative comparison (A. E. Clark and Oswald, 1996; Ferrer-i-Carbonell, 2005); unemployment (A. E. Clark and Oswald, 1994); and other non-pecuniary factors (Blanchflower and Oswald, 2004b).

As a result of earlier research, the UK Office for National Statistics (ONS) now measures four aspects of subjective well-being. These are: happiness yesterday;

satisfaction with life; worthwhileness of life; and anxiety yesterday. It is clear that the collection of these data is designed to help inform policymakers of the factors that improve an individual's overall quality of life. However, limited work has been done to establish which of these well-being measures is considered important to individuals. Ultimately, if economies are to shift their attention to improving societal welfare, it is important to understand the relative significance of each aspect of well-being in order to allow for informed policy decisions.

O'Donnell and Oswald (2015) appear to be the first to obtain weightings for the four aspects of well-being, in order to estimate the linear approximation of a 'change in well-being' function. The weights they collect correspond to f_h, f_s, f_w, and f_a in the following expression:

$$\Delta W \cong K[f_h(h - h_0) + f_s(s - s_0) + f_w(w - w_0) - f_a(a - a_0)] \quad (6)$$

where h, s, w and a refer to happiness, life satisfaction, worthwhileness, and anxiety, respectively. Of the four samples they collect, three give the highest weighting to life satisfaction (f_s).

These data represent a first attempt at estimation. As such, these three samples were taken from economics students, business students, and professional economists. They are therefore likely to be unrepresentative of the wider population. Since the focus of the study was to estimate weights alone, no attempt was made to uncover how these weights might be determined, and whether they are different for different individuals.

Given that previous research on the determinants and life cycle trends of well-being has been plentiful (see Dolan, Peasgood, and White, 2008, for a review of the economic literature on well-being), a natural extension would be to link these ideas to the determination of well-being priorities. This chapter appears to be the first study of its kind to address these issues.

This study has two main aims. First, it extends the findings in O'Donnell and Oswald (2015) by uncovering which well-being aspects are given highest priority over the life cycle. Second, it provides an initial attempt to understand what determines the rank ordering of well-being aspects. Neither of these issues has previously been explored in the literature. I find that the non-linear 'mid-life crisis' dip observed in well-being levels does not reliably translate to a corresponding relationship for well-being prioritisation. However, the middle-aged focus more on their own level of well-being in their determination of happiness and anxiety. Individual characteristics have a strong influence on which aspect is given highest priority.

The trends and determinants of subjective well-being have been studied for a number of years, in both economics and psychology. A large number of general findings have emerged as a result of this research. Relative comparisons based on income have strong effects on well-being levels (e.g. A. E. Clark and Oswald, 1996; Dolan, Peasgood, and White, 2008; Ferrer-i-Carbonell, 2005), suggesting that the traditional economic focus on increasing income per capita (past some threshold) may have little impact in terms of increasing per capita well-being.

A wealth of economic studies have shown the negative impact of unemployment and poor labour market outcomes on well-being (e.g. Frey and Stutzer, 2002; Oswald, 1997; L. Winkelmann and R. Winkelmann, 2008). These seem to be linked less with a loss of income than a psychological loss. The same can be said for a reduction in levels of health. Other personal circumstances and lifestyle choices also have a significant influence on levels of subjective well-being. Most notably, this includes marriage (Blanchflower and Oswald, 2004a), exercise (Ferrer-i-Carbonell and Gowdy, 2007), and diet (Mujcic and Oswald, 2016).

In addition to these situational factors, personal characteristics also affect well-being. The Big Five personality factors Extraversion and Neuroticism are strongly linked to well-being levels; the former positively and the latter negatively (Diener and Lucas, 1999). Subsequent research has shown that finer-grained measures may have more explanatory power (Dolan, Peasgood, and White, 2008). Still, it is clear that individual characteristics shape well-being.

One of the most prominent and consistent findings is that there is a U-shape in well-being over the life cycle in cross-sectional data, and whilst controlling for factors such as health and income (e.g. Blanchflower and Oswald, 2004a; Ferrer-i-Carbonell and Gowdy, 2007). This pattern is consistent with the theory of a 'mid-life crisis' in psychology (e.g. Brim, 1976), and may be generated partly as a result of forecasting error (Schwandt, 2016). The U-shape for happiness and life satisfaction has been shown to hold across a number of different countries, and when taking into account cohort effects (Blanchflower and Oswald, 2008). This rules out the explanation that the observed mid-life dip in well-being is being caused by generational differences in the trajectory of happiness over the life cycle. The U-shape holds longitudinally within individuals (Cheng, Powdthavee, and Oswald, 2015). Evidence of a mid-life crisis has also been found in primates (Weiss et al., 2012). Some recent studies (e.g. Frijters and Beatton, 2012; Kassenboehmer and Haisken-DeNew, 2012) have highlighted issues with this pattern due to unobserved heterogeneity (such as interviewer effects), reverse causality, and fixed effects. Despite this, the majority of evidence points towards the presence of a U-shape over the life cycle.

The remainder of this chapter is organised as follows. Section 2.2 uses data from the UK APS to analyse life cycle patterns for well-being levels, in order to form expectations for prioritisation behaviour. It then discusses how prioritisation may be determined by well-being levels. Section 2.3 presents evidence obtained from the online survey data collected for O'Donnell and Oswald (2015). Section 2.4 presents evidence obtained from a new online survey designed to address shortcomings of the data from Section 2.3, and estimates the model proposed in Section 2.2. Section 2.5 provides a discussion of the results, and how the difference between cognitive and affective measures of well-being may help to explain them. Finally, Section 2.6 summarises the findings of the chapter, and concludes.

2.2 BACKGROUND AND EXPECTATIONS

2.2.1 *Well-being levels across the life cycle*

It is important to highlight at this stage that throughout this chapter, I attempt to identify patterns over the life cycle from cross-sectional data. This has the inherent problem that a comparison is made across birth cohorts, and so we may be uncovering generational differences rather than true age effects. However, as discussed briefly in Section 2.1, whilst non-linearities in well-being levels were initially shown using cross-sectional data, subsequent work has confirmed the relationships when controlling for cohort differences (Blanchflower and Oswald, 2008). Whilst this does not mean that any life cycle patterns found in the present work can be generalised to hold within the same birth cohort, it does suggest that findings from this study have the potential for wider applicability.

To understand how we might expect individuals to prioritise well-being *a priori*, and to determine whether the U-shaped pattern for levels of well-being holds in more recent data, I look at the life cycle patterns of well-being levels using data from the 2013-14 UK Annual Population survey (Office for National Statistics. Social Survey Division, 2014). The four well-being questions asked to respondents are as follows (each is scored on a scale from 0-10):

1. "Overall, how happy did you feel yesterday?"
2. "Overall, how satisfied are you with your life nowadays?"
3. "Overall, to what extent do you feel the things you do in your life are worthwhile?"
4. "Overall, how anxious did you feel yesterday?"

Figure 5 plots fitted third-order polynomial curves in age to reported levels of happiness yesterday, life satisfaction, worthwhileness of life, and anxiety. We see that happiness and life satisfaction exhibit a U-shape across the life cycle, and anxiety is hump-shaped. No additional controls were used - this is a pattern found in raw data.[1] According to these data, it does seem that the middle-aged are, in fact, suffering from lower levels of well-being than the young and old.

The pattern for worthwhileness is somewhat more ambiguous. The overall life cycle pattern could be described as a 'late wave'. However, when separating for gender, we see that males seem to experience an increasing level of worthwhileness with age, whilst females suffer a sharp decline in later life. Further investigation reveals that this pattern disappears (i.e. the female curve looks more similar to the male curve) when looking only at those with good or very good levels of general health. From this, it appears as though females may be factoring health more highly in their evaluation of worthwhileness than males.[2]

When controls are added (Table 12), we see that the U-shape is still present. Worthwhileness becomes U-shaped, in a similar fashion to happiness and life satisfaction. The estimates in Table 12 are obtained using both OLS (i.e. assuming cardinality of well-being responses) and ordered logit (i.e. assuming ordinality

1 This is in contrast to Easterlin (2006), who claims that the U-shape only arises when control variables are included in a quadratic regression of well-being on age.
2 This could be due to evolutionary reasons, such as fertility, and ability to nurture offspring.

only). Both estimation methods offer similar interpretations in terms of direction, which is consistent with Ferrer-i-Carbonell and Frijters (2004).

Gender differences over the life cycle are minimal for levels of life satisfaction and worthwhileness. They are significantly different for happiness and anxiety, though gender differences in anxiety are more pronounced. Anxiety levels drop off much more rapidly for males after middle age, whereas they peak closer to the age of 60 for females. All four of the life cycle relationships in Table 12 point to a middle-aged dip in subjective well-being (i.e. a U-shaped pattern for levels of happiness, life satisfaction, and worthwhileness; and a hump-shaped pattern for levels of anxiety).

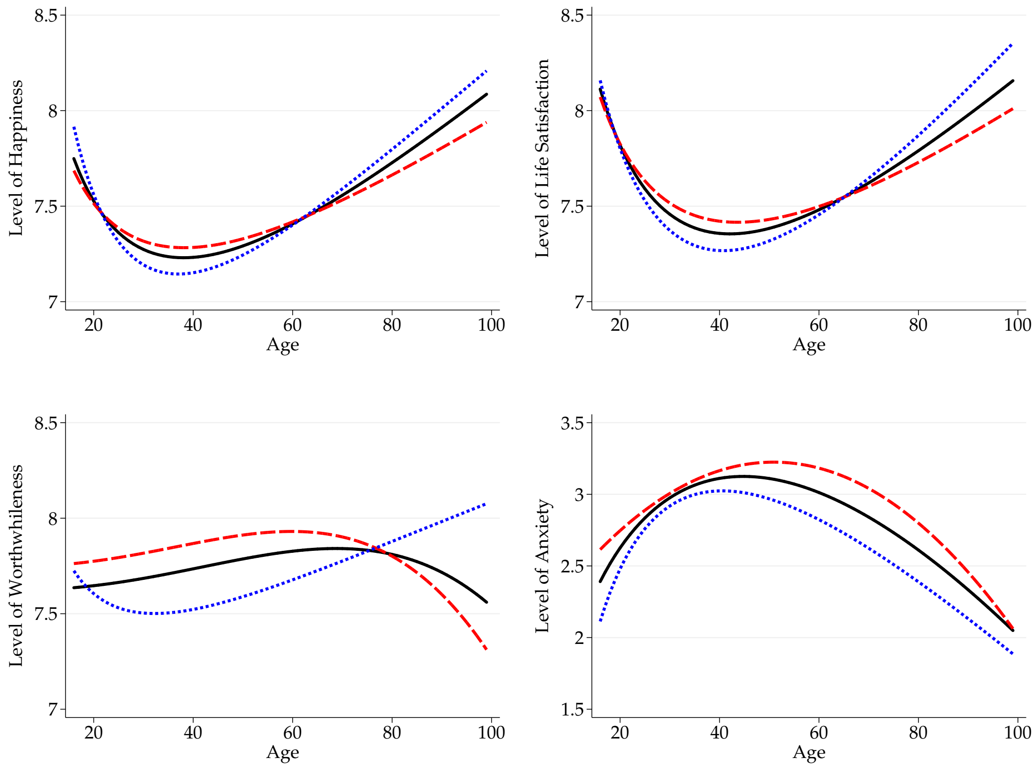


FIGURE 5: 3rd order polynomial age curves for levels of well-being, from UK APS, 2013-14. Solid black line = raw data; dashed red line = females; dotted blue line = males. n=165,122.

TABLE 12: The quadratic life cycle relationship of well-being levels, APS 2013-14

	OLS				Ordered Logit			
	Happiness	Life Satisfaction	Worthwhileness	Anxiety	Happiness	Life Satisfaction	Worthwhileness	Anxiety
Age	-0.0551***	-0.0818***	-0.0428***	0.0409***	-0.0591***	-0.103***	-0.0618***	0.0318***
Age^2	0.000659***	0.000896***	0.000530***	-0.000412***	0.000713***	0.00115***	0.000769***	-0.000338***
Male	0.165	-0.0476	-0.185	-1.011***	0.218	0.0224	-0.283*	-0.695***
Age * Male	-0.0140*	-0.00267	-0.00607	0.0449***	-0.0167**	-0.00868	-0.00592	0.0318***
Age^2 * Male	0.000186**	4.63e-05	9.76e-05	-0.000586***	0.000202**	0.000115	9.47e-05	-0.000402***
General health (reverse)	-0.538***	-0.510***	-0.382***	0.634***	-0.474***	-0.612***	-0.467***	0.413***
Net pay	9.70e-07	1.71e-06***	1.57e-06***	1.42e-06	3.98e-07	2.57e-06***	2.00e-06***	1.44e-06**
Married, living with spouse	0.340***	0.452***	0.347***	-0.108***	0.314***	0.570***	0.436***	-0.0717***
Married, separated	-0.0996**	-0.254***	0.0405	0.0497	-0.0395	-0.251***	0.0765**	0.0095
Divorced	-0.00705	-0.0598**	0.0401*	0.0295	0.0273	-0.0385	0.0941***	-0.00675
Widowed	-0.246***	-0.364***	-0.0552	0.014	-0.168***	-0.371***	-0.00151	0.00714
Were/still in civil partnership	0.0548	0.482***	0.191**	0.285	0.172	0.612***	0.240**	0.133
Mixed	-0.112	-0.209***	-0.0216	0.195	-0.0955	-0.260***	3.19e-05	0.112
Indian	0.197***	-0.0823*	-0.0159	0.112	0.190***	-0.0946*	-0.0172	0.0607
Pakistani	-0.133	-0.232***	-0.0077	0.347***	-0.00897	-0.189**	0.0555	0.166**
Bangladeshi	0.316**	-0.168	0.0963	0.125	0.353***	-0.0832	0.275**	0.0127
Chinese	-0.0249	-0.283***	-0.381***	-0.0691	-0.113	-0.409***	-0.530***	0.0264
Other Asian	0.381***	0.142*	0.115	0.220*	0.357***	0.164*	0.168*	0.134
Black	-0.0777	-0.458***	-0.0468	0.247***	3.08e-05	-0.534***	-0.0157	0.114**
Other ethnicity	-0.082	-0.127**	-0.0929	0.312***	-0.0145	-0.120*	-0.0709	0.191***
Constant	9.261***	10.07***	9.242***	1.017***	-	-	-	-
Observations	67324	67324	67324	67324	67324	67324	67324	67324
R^2	0.052	0.098	0.061	0.036	-	-	-	-

Standard errors and ordered logit cut constants omitted for brevity. Robust s.e. used for OLS. *** p<0.01, ** p<0.05, * p<0.1

2.2.2 *Determining well-being prioritisation*

Given that individuals in general do not have large amounts of empirical data about well-being in society, it seems natural to believe that they will utilise information about their own levels of well-being in determining how to prioritise a given well-being aspect. On its own, this suggests that well-being priorities may exhibit a U-shape or hump-shape similar to that shown in Figure 5. However, own well-being is likely to be combined with personal beliefs about what constitutes optimal policy when determining social priorities. These beliefs will be influenced by (limited) knowledge about the well-being levels of others, as well as by individual fixed effects.

Let P_i^A represent the well-being priority for aspect A, given an individual i.[3] The priority given to a particular aspect of well-being depends upon a function of one's own well-being level L_i, and a function of the expected level of others' well-being $L_{\neg i}$. Hence, priority can be represented in the following way:

$$P_i^A = a_i f(L_i^A) + b_i g(E_i[L_{\neg i}^A]) \tag{7}$$

where a_i and b_i are constants.

One's level of well-being at any given point in life is influenced by factors such as regret from forecasting error, and optimism (Schwandt, 2016). These factors contribute to the underlying U-shape of happiness and life satisfaction over the life cycle. This life cycle trend is implicitly contained within L_i. L_i also captures much of what constitutes preferences in terms of choice utility. Whilst there are some exceptions, what people choose in a decision scenario largely corresponds to what provides them with the highest level of subjective well-being (Benjamin et al., 2012).

We can estimate a general form of this relationship, given that we have data on levels and beliefs about others' levels. I assume there is some commonality in

3 In practice, this priority can either represent a weighting, or a simple ordinal ranking.

how priorities are formed, which allows us to take parameters across individuals as constant:

$$P_i^A = \beta_0 + \beta_1(L_i^A)^{\beta_2} + \beta_3(E_i[L_{\neg i}^A])^{\beta_4} + \epsilon_i \tag{8}$$

One hypothesis about the values of β_1 and β_2 can be formed by considering the possibility that the marginal impact of an extra unit of well-being affects overall priority. For example, if we were to assume a standard concave value function (i.e. displaying diminishing marginal returns), the value of an additional unit of well-being would diminish as well-being level increases. This would suggest that individuals with a low level of well-being for a given aspect would prioritise that aspect more highly than someone with a higher level.[4] In other words, this would imply $f(L_i^A)$ is decreasing in L_i^A, suggesting that $\beta_2 \in (0, 1)$.

Given this supposition, I hypothesise that priority will be allocated to those well-being aspects for which the corresponding level is lower (i.e. $\beta_1 < 0$). Therefore, if well-being priority was determined by own levels alone, then based on the APS data in Section 2.2, we would expect that the middle-aged place higher priority on the increase of happiness and life satisfaction; and reduction of anxiety. We may expect the opposite for worthwhileness, given the nature of the relationship for worthwhileness levels in Figure 5.

However, one expects that individuals would account for others' needs, in addition to their own. Given the arguments above about one's own level of well-being, we might expect that the same inverse relationship would hold between beliefs about others' level of well-being, and one's priority for a given aspect. Thus, β_2 and β_4 are likely to be similar, with $\beta_3 < 0$. It is not clear whether individuals would place more weight on their own well-being, or their expectations about others' well-being. However, given own levels are more available and salient, a sensible hypothesis would be that $|\beta_1| > |\beta_3|$. That is, I expect own levels of well-being to be more important in determining well-being priority for a given aspect than others' levels of well-being.

4 For example, we might expect that an increase in happiness by one point on a 0-10 scale would be more desirable to someone with a happiness level of 4 than someone with a happiness level of 8.

It is important to note the implicit assumption that all priorities are independent. However, if we ask individuals to form a rank ordering, or to provide weights that add up to a fixed sum, the priority of one of the four aspects will be determined by the other three. Therefore, it is possible that estimated parameter values for one of the four aspects will be largely different from the rest.

There are many variables representing individual differences and fixed effects that are subsumed into the error term. Easily observable determinants of well-being (such as income, employment status, and marital status) can be collected in order to eliminate at least some of the omitted variable bias that may otherwise arise from estimation of (8). These variables may also influence priorities aside from their influence on levels. The greater the impact these variables have on prioritisation, the less likely we are to see the 'mid-life crisis' pattern of well-being levels reflected in priorities.

In particular, two of the Big Five personality factors - Extraversion and Neuroticism - are strongly linked to subjective well-being levels (Diener and Lucas, 1999). The most clear link is between Neuroticism and anxiety. Since Neuroticism captures sensitivity to negative affect (McCrae and John, 1992), more Neurotic people will suffer the most from high anxiety levels. Therefore, individuals with higher Neuroticism would see greater value in addressing factors that would reduce anxiety levels (i.e. giving higher priority to anxiety). Extraversion is linked to responsiveness to positive outcomes, such as rewards. Hence, one would expect those high in Extraversion to prioritise happiness more than others.

2.3 SURVEY 1

To provide a first indication of the pattern of well-being priorities across the life cycle, I use data from a short online survey. The survey was administered in mid-2014, using participants recruited from MTurk. This sample was one of the four used in O'Donnell and Oswald (2015). Individuals were asked to prioritise the four standardised aspects of well-being by allocating points to each aspect, where the sum of points was constrained to 100. The exact text of the well-being weighting task is provided in the Appendix.

In addition, a limited set of demographic data was obtained. Demographic questions were asked at the end of the survey, in order to avoid any potential priming effects (although such effects would be unlikely in this case). 306 responses were collected in total. Each respondent was paid $1, and mean survey completion time was 7.4 minutes. The mean age of the sample was 32. 60% were male. 79% of respondents were U.S. nationals.

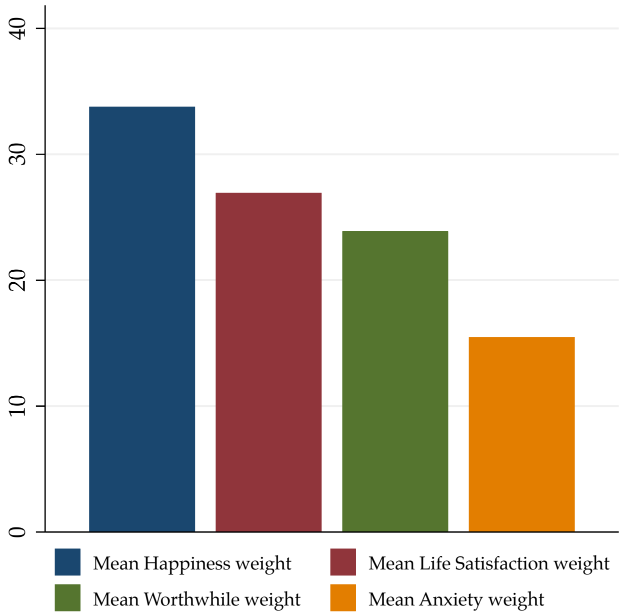


FIGURE 6: Mean weighting given to each of the four aspects of subjective well-being in survey 1. n=306.

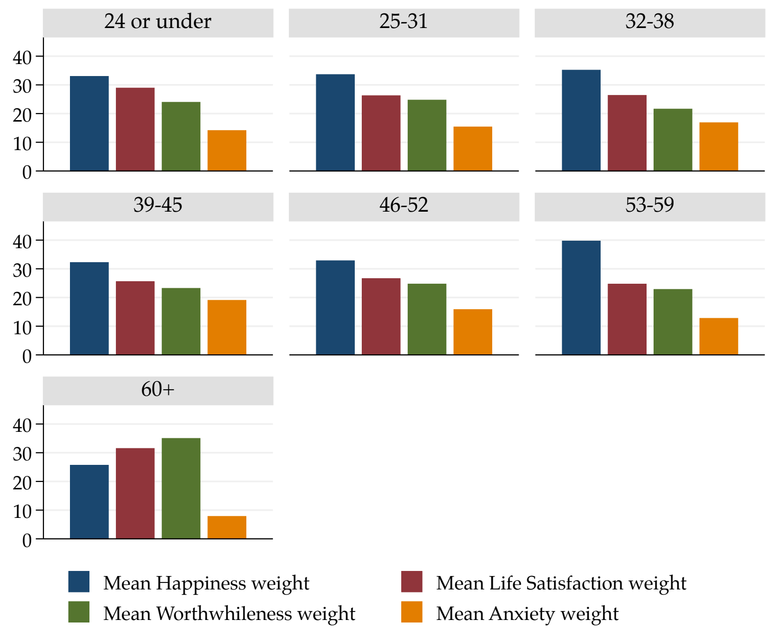


FIGURE 7: The rank ordering of well-being priorities in survey 1 is largely the same across age bands.

Figure 6 shows the mean weighting given to each aspect of well-being. Happiness is valued the highest, followed by life satisfaction, worthwhileness, and anxiety (in that order). The rank ordering of well-being aspects in this survey is preserved across all but the final age band (Figure 7). Worthwhileness is given much higher priority by those aged 60 or above, but data at this end of the age range is sparse.[5] Mean relative weightings, however, do not remain constant across bands.

Figure 8 shows a third-order polynomial fit to the weighting for each aspect of well-being over the life cycle. A cubic polynomial is used to account for non-linearity, without constraining the shape to be a parabola. Weights for happiness and anxiety follow a hump shape across the life cycle. In contrast, weights for life satisfaction and worthwhileness follow a U-shape across the life cycle. This implies that middle-aged individuals care relatively more about policies that increase happiness and reduce anxiety than the young or the elderly. The 95% confidence intervals displayed on the curves show that the trend is noisier towards

5 For example, only 6 responses are obtained from those aged 60 or above.

the upper end of the age range. This is due to the shortage of data from older individuals.

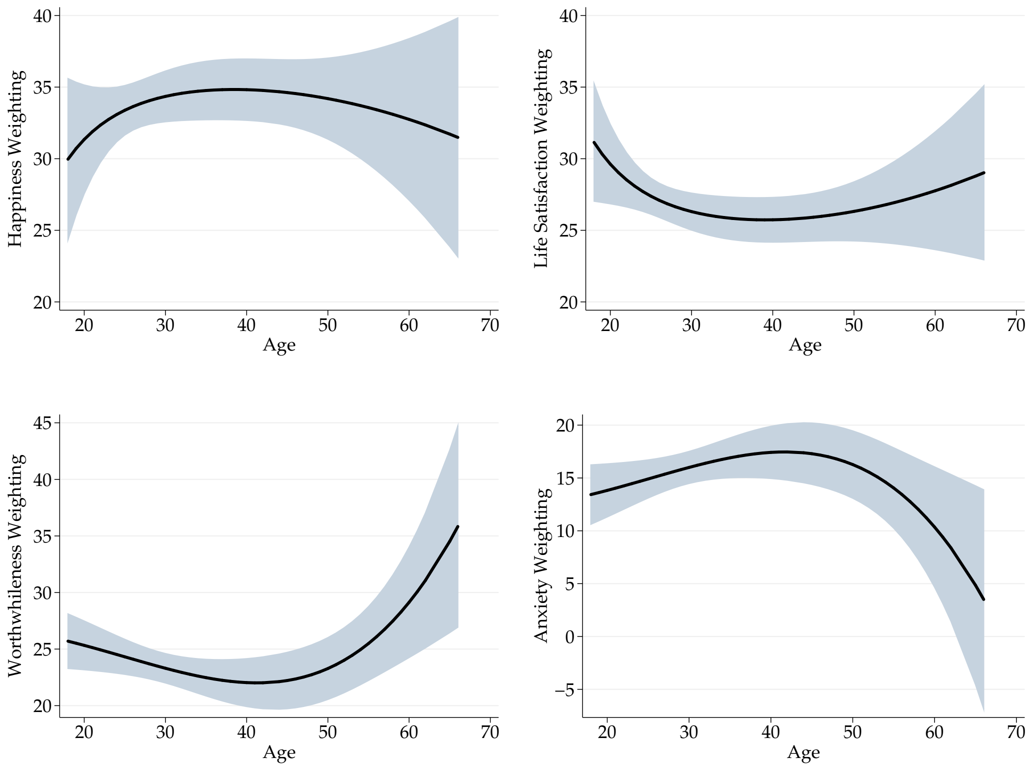


FIGURE 8: 3rd order polynomial age curves fitted to well-being weightings, with 95% confidence intervals. n=306.

In order to ascertain the strength of these parabolic relationships, I estimate a quadratic relationship in age, using OLS with robust standard errors. Specification 1 in Table 13 shows the raw quadratic relationship between age and well-being weightings. The quadratic term is not statistically significant for happiness, but is significant for life satisfaction, worthwhileness, and anxiety.

Specification 2 in Table 13 adds gender interaction terms to separate the relationship for males and females. Men place higher weight on happiness and life satisfaction, and therefore, less weight on worthwhileness and anxiety than women. The results for males in this sample suggest a flatter life cycle weighting profile than for females. This is particularly true for both happiness and worthwhileness, where the curvature is being driven primarily by females. Despite this, the polarity of the quadratic relationships remains the same, even when separating for gender.

The significance of the gender difference in worthwhileness weightings may be reflecting the unusual pattern observed in the APS data in Figure 5. Recall that the level of worthwhileness for females was somewhat hump-shaped over the life cycle, in contrast to a U-shape for males. The hypothesis that individuals are prioritising aspects that they are deficient in is consistent with a stronger U-shape for females in terms of the importance they place on worthwhileness around middle age.

2.3.1 *Limitations of survey 1*

There are a number of limitations with the findings from this simple first survey. First, and most importantly, the sample does not stratify for age. This results in highly noisy data at the upper end of the age range. 80% of respondents were below the age of 39, which means that the patterns observed after middle age are likely to be inaccurate. Whilst data acquired from MTurk have been shown to be reliable (Mason and Suri, 2012), a more age-representative sample is required to draw meaningful conclusions about life cycle patterns. Due to the novel nature of the question posed, large randomised datasets with data on well-being priorities currently do not exist. However, stratification in a smaller-scale survey is feasible.

Second, no randomisation was used in the order of well-being statements. This may result in bias due to order effects. The overall mean ordering of well-being priorities from the MTurk sample differs from the other samples used in O'Donnell and Oswald (2015), even though the same ordering of aspects was shown to each sample in that paper. In these other samples, the mean happiness weighting drops to third in the rank ordering, after life satisfaction, and worthwhileness. Whilst this may suggest that order effects are unlikely to have had a substantive impact on the results, it is important to note that these other samples came from students and professional economists. It is possible that they are less susceptible to order effects than those responding quickly to an online survey.

TABLE 13: The relationship between age and well-being weights. Weightings for the importance of happiness yesterday and anxiety are hump-shaped across the life cycle, whilst weightings for life satisfaction and worthwhileness of life are U-shaped.

	Specification 1				Specification 2			
	Happy	Satisfaction	Worthwhile	Anxiety	Happy	Satisfaction	Worthwhile	Anxiety
Age	0.577	-0.621**	-0.928**	0.972**	1.249*	-0.369	-1.695***	0.815
	(0.420)	(0.305)	(0.379)	(0.391)	(0.654)	(0.459)	(0.521)	(0.666)
Age^2	-0.00701	0.00744**	0.0124**	-0.0129***	-0.0149*	0.00504	0.0220***	-0.0121
	(0.00541)	(0.00370)	(0.00488)	(0.00478)	(0.00811)	(0.00533)	(0.00651)	(0.00785)
Male	-	-	-	-	21.68	9.312	-26.00*	-4.995
					(15.57)	(11.76)	(13.44)	(15.59)
Age * Male	-	-	-	-	-1.16	-0.355	1.433*	0.0823
					(0.844)	(0.636)	(0.729)	(0.836)
Age^2 * Male	-	-	-	-	0.0138	0.00324	-0.0185**	0.00144
					(0.0106)	(0.00779)	(0.00928)	(0.0102)
Constant	23.18***	38.43***	39.59***	-1.206	10.41	32.05***	53.97***	3.566
	(7.563)	(5.668)	(6.880)	(7.214)	(12.34)	(8.744)	(9.792)	(12.75)
Observations	306	306	306	306	306	306	306	306
R^2	0.006	0.013	0.022	0.017	0.012	0.023	0.034	0.023

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Finally, these data do not allow us to understand how priorities are formed, due to the absence of individual fixed effects and well-being levels. Although the results from survey 1 resemble the non-linear life cycle patterns for well-being levels from the APS data in Section 2.2, we do not know the extent to which levels are contributing to the prioritisation process. Collecting data on well-being levels would allow us to assess how much the observed weightings depended on individuals' own subjective well-being. In addition, individual fixed effects influence subjective well-being (Ferrer-i-Carbonell and Frijters, 2004). Therefore, their inclusion is likely to explain further variation in priorities.

2.4 SURVEY 2

2.4.1 *Design*

To address the limitations of the first survey, and allow for estimation of the model proposed in Section 2.2.2, a second online survey was designed. In order to resolve some of the noise around the extremes of the age range, the sample was stratified equally amongst seven age bands (see Table 14).[6]

TABLE 14: The difference in age distribution between survey 1 and survey 2.

	Num of observations	
	Survey 1	Survey 2
24 or under	73	40
25-31	105	40
32-38	68	40
39-45	23	40
46-52	18	41
53-59	13	40
60 or over	6	40
	306	281

6 Although the survey was designed so that 40 observations were collected in each age band, one age band received 41 responses. Rather than discarding data, I include this extra observation in the analysis.

The main task asked individuals to rank the four well-being aspects ordinally, rather than assigning weighting points. Whilst this has the disadvantage that we do not obtain information about the relative strength of a priority, it simplifies the task and prevents arbitrary weights from being assigned. To compensate for the lack of a numerical measure of weighting for each aspect, I asked respondents how sure they were of each rank position they decided upon. Certainty was recorded on a three-point scale: not at all sure; somewhat sure; and very sure.

The starting order of the four aspects was randomised for each respondent, which eliminates the problem of potential order effects. The ranking task requires that individuals drag the aspects in the ordering they desire, with a mechanism to prevent skipping ahead without moving any item.

In order to estimate a form of expression (8), participants were asked about their levels of well-being, and their beliefs about others' well-being levels and priorities. The measures used for levels are identical to those used in the APS. Information on individual fixed effects was captured by collecting data on basic demographic information; as well as employment status, marital status, children, health, education, income, and Big Five personality factors using the 20 item mini IPIP personality inventory (Goldberg et al., 2006). All of these variables are included as they have been found to be associated with levels of well-being. Statements from the mini-IPIP measure of personality were presented in a random order to each participant.

Time preference has been shown to vary across the life cycle. Empirically, the discount factor appears to be positively correlated with age and income (Green et al., 1996). Since income peaks around middle-age, the evidence would suggest that the discount factor should be relatively high at this point in the life cycle.[7]

Hence, differences in time preference over the life cycle may influence prioritisation. I use the smaller-sooner vs larger-later task developed by Coller and Williams (1999) in order to elicit time preferences. Since I do not use task-specific incentives, this measure is likely to be biased towards patience (i.e. a lower discount rate, and therefore higher discount factor). Therefore, an item on smok-

7 The majority of evidence for increasing discount factors comes from experiments that deal with monetary gains. The same effects do not hold for the discounting of emotional experiences (Löckenhoff and Rutt, 2015).

ing frequency was also included. Reimers et al. (2009) find that smoking behaviour is strongly related to time preference. Smokers prefer smaller-sooner monetary payoffs over larger-later ones, suggesting that the propensity to smoke would make a good proxy for impatience.

2.4.2 *Results*

Responses to the survey were collected in May 2016, again using participants recruited from MTurk. The mean survey completion time was 6.04 minutes. Each respondent was paid $2.50. Approximately 40% of respondents were male (112 of the 281). All but 5 respondents were U.S. nationals.

There are key differences in the results from survey 2 when compared to the those from survey 1. First, the mean rank ordering of the well-being aspects has changed. Figure 9 shows that happiness is now the lowest ranked aspect of well-being on average.[8] The relative rankings of the other three aspects remain unchanged, so that life satisfaction now receives the highest priority, followed by worthwhileness, and anxiety. Figure 10 shows that this relative ranking is preserved across age bands. The fact that life satisfaction is the highest priority well-being aspect is consistent with all of the non-MTurk samples in O'Donnell and Oswald (2015). However, in those samples, happiness is given higher priority than anxiety, which is not the case in the present data.

8 Whilst the survey had individuals rank the aspects from top to bottom, so that 1 represented the highest ranking, labelling in the analysis has been reversed to show 4 as the highest ranking.

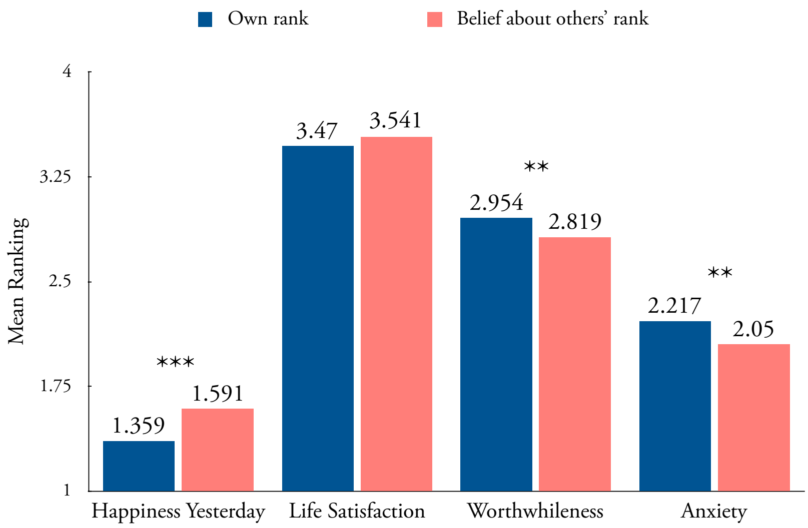


FIGURE 9: Means of well-being priority rankings from survey 2, where 4 represents the highest rank. Stars indicate significance of a t-test for a difference in means between own ranking, and beliefs about others' ranking. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

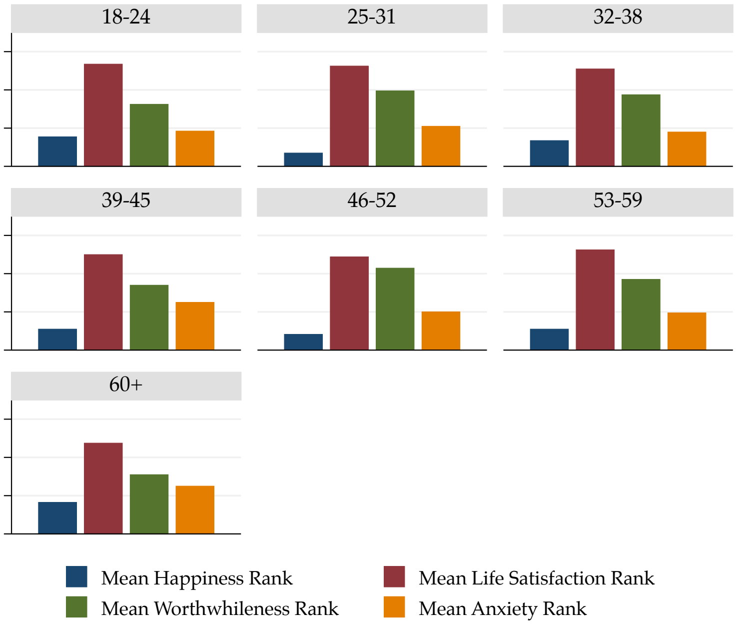


FIGURE 10: Happiness falls to the bottom of well-being priorities for all age groups in survey 2.

Figure 9 also plots beliefs about how others would rank the four aspects. The overall mean ordering of beliefs is consistent with the order generated from own mean ranking. However, t-tests of the difference between mean rankings for each aspect show that, on average, people overestimate the ranking others give to happiness. This is compensated by underestimates in the beliefs about others' worthwhileness and anxiety rankings.

Second, the non-linearity of well-being priorities across the life cycle found in survey 1 is not replicated by the data from survey 2. The third-order polynomial age curves in Figure 11 show little evidence of non-linearity. Table 15 confirms this finding by fitting a quadratic in age, with identical specifications to those in Table 13 for survey 1. An ordered logit model is used instead of OLS, since survey 2 asks for ordinal rankings, and it is not clear whether the distance between the rankings is perceived as being the same. There are no significant age trends for any of the well-being aspects.

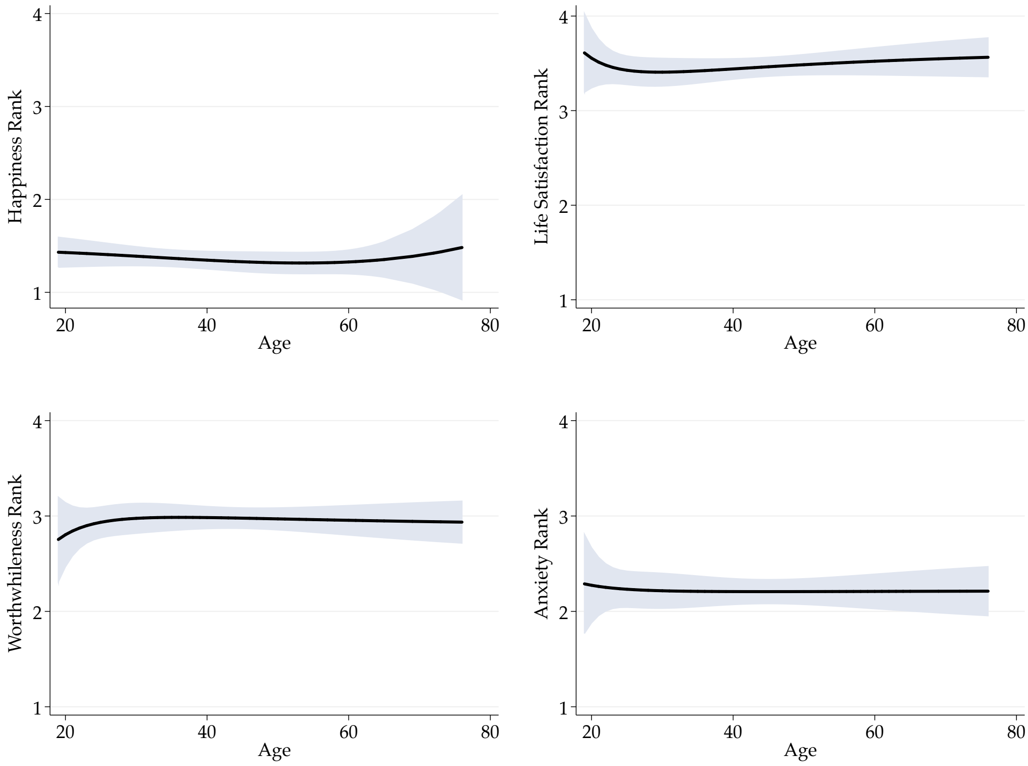


FIGURE 11: 3rd order polynomial age curves fitted to well-being ranks from survey 2, with 95% confidence intervals. n=281. Rankings appear to be constant over the life cycle.

Table 15: The relationship between age and well-being rankings from survey 2 (ordered logit). Rankings are consistent across the life cycle.

	Specification 1				Specification 2			
	Happy Rank	Satis Rank	Worth Rank	Anxiety Rank	Happy Rank	Satis Rank	Worth Rank	Anxiety Rank
Age	-0.0331	0.00534	0.00989	-0.0166	-0.0973	0.045	0.0834	-0.0901
	(0.0600)	(0.0547)	(0.0525)	(0.0507)	(0.0824)	(0.0737)	(0.0752)	(0.0683)
Age^2	0.000306	-5.39e-05	-7.35e-05	0.000181	0.000945	-0.000443	-0.000831	0.000933
	(0.000696)	(0.000630)	(0.000605)	(0.000581)	(0.000941)	(0.000831)	(0.000857)	(0.000770)
Male	-	-	-	-	-2.925	1.759	3.131	-3.028
					(2.421)	(2.254)	(2.155)	(2.082)
Age * Male	-	-	-	-	0.145	-0.0713	-0.143	0.119
					(0.121)	(0.112)	(0.107)	(0.103)
Age^2 * Male	-	-	-	-	-0.00138	0.000712	0.00146	-0.00131
					(0.00141)	(0.00130)	(0.00123)	(0.00119)
Constant 1	0.153	-3.706***	-2.840***	-1.625	-1.141	-2.723*	-1.212	-3.525**
	(1.195)	(1.168)	(1.088)	(1.032)	(1.680)	(1.584)	(1.561)	(1.437)
Constant 2	1.899	-1.969*	-0.795	0.494	0.624	-0.983	0.845	-1.354
	(1.209)	(1.111)	(1.059)	(1.024)	(1.687)	(1.544)	(1.550)	(1.421)
Constant 3	3.456***	-0.293	1.356	1.546	2.19	0.696	3.009*	-0.283
	(1.286)	(1.101)	(1.062)	(1.029)	(1.743)	(1.540)	(1.560)	(1.417)
Observations	281	281	281	281	281	281	281	281

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

Figure 12 plots the gender-separated fitted quadratic curves from the second specification in Table 15. Whilst gender differences were not statistically significant in the regression, the graphs do indicate a difference in ranking patterns over the life cycle between males and females. The dotted blue and red curves in Figure 12 represent fitted curves for all rankings that respondents rated at least *somewhat sure*, in order to exclude uncertain responses. Doing this appears to suggest that the priority given to worthwhileness increases over the life cycle. It also shifts the priority for anxiety upwards. For females, we see that happiness ranking decreases over the life cycle. This suggests that a more naive or uncertain individual might undervalue anxiety and worthwhileness, in favour of happiness.[9]

2.4.3 *Model Fitting*

In order to investigate how priorities are formed, I estimate a version of equation (8), as discussed in Section 2.2.2. Since the dependent variables will be ordinal rankings for each of the four well-being aspects, I use an ordered logit framework for estimation. First, as we do not know the nature of the functions $f(.)$ and $g(.)$ from equation (7), I compare a non-linear specification to a linear specification to establish which is a better fit for the data.

The non-linear specification is based on equation (8):

$$Rank_i^A = \beta_1 (L_i^A)^{\beta_2} + \beta_3 (E_i[L_{\neg i}^A])^{\beta_4} \tag{9}$$

The linear specification takes $\beta_2 = 1$ and $\beta_4 = 1$, i.e:

$$Rank_i^A = \beta_1 (L_i^A) + \beta_3 (E_i[L_{\neg i}^A]) \tag{10}$$

9 It also suggests that beliefs of these uncertain individuals may be driving the differences between own and others' ranking in Figure 9.

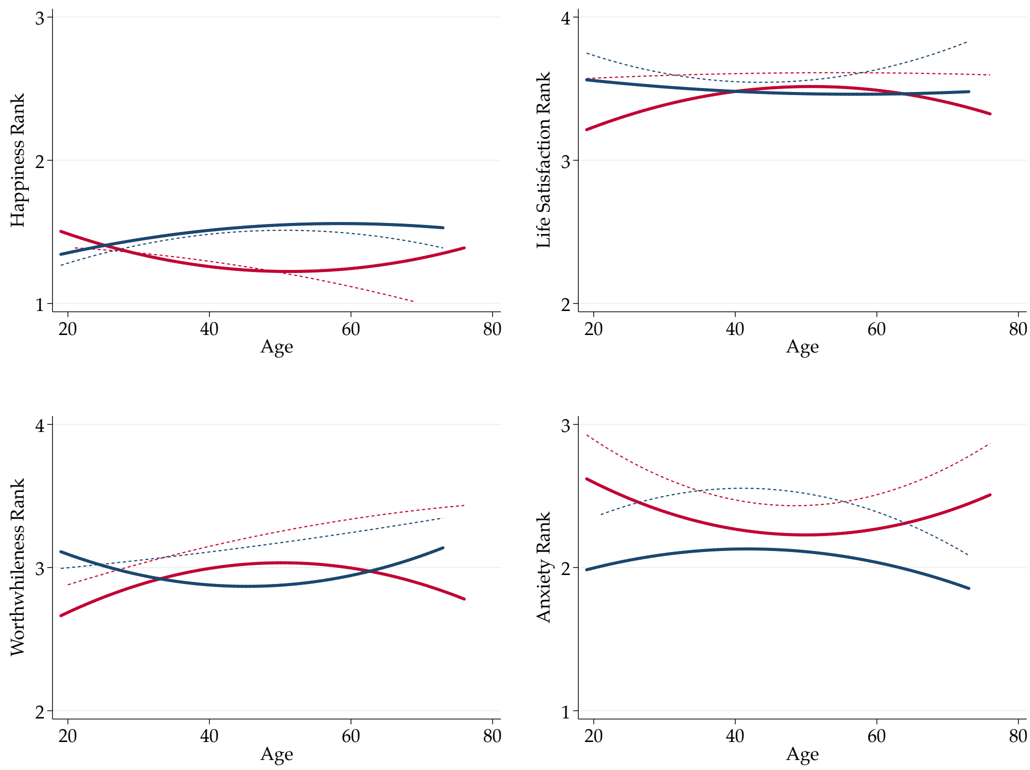


FIGURE 12: Quadratic age curves to show gender differences in well-being rank (survey 2). Solid blue and red curves are for males and females, respectively. Dashed blue and red curves are for those males and females that were at least somewhat sure of their ranking.

These specifications are used in the following ordered logistic log-likelihood function for maximum likelihood estimation (for a given aspect of well-being, A):

$$lnL^A = \sum_{i=1}^{281}\sum_{j=1}^{4} m_{ij}\, ln[F(k_j - Rank_i^A) - F(k_{j-1} - Rank_i^A)] \qquad (11)$$

where $k_0 = -\infty$ and $k_4 = \infty$; $m_{ij} = 1$ if the observed value of $Rank_i^A = j$ and 0 otherwise; and $F(.)$ is the c.d.f. of the logistic function:

$$F(x) = \frac{1}{1+e^{-x}} \qquad (12)$$

Table 16 shows maximum likelihood estimates for this model, using both linear and non-linear specifications. In order to determine which model is more appropriate, I make use of the Likelihood Ratio (LR), and the Bayesian Information Criterion (BIC).[10] The LR test is suitable for testing the validity of nested models. It can be used here, since (10) is a restricted version of (9). The BIC is more general, allowing for comparison of non-nested models. Both tests use information about the maximised likelihood function from which to draw inferences. I use both for greater clarity.

The LR test statistic is given by:

$$D = 2(ln\, L_{nonlinear} - ln\, L_{linear}) \sim \chi^2(2) \qquad (13)$$

The BIC test statistic is given by:

$$BIC = k\, ln(281) - 2\, ln\, L \qquad (14)$$

where k is the number of parameters estimated, and the log-likelihoods are the maximised values after estimation.

A lower BIC score indicates a better model, with any difference greater than approximately 6 indicating a strong preference for the specification with the lower score. In Table 16, we see from the BIC score differences that the non-linear specification is only close to the linear specification for anxiety ranking. Rankings

10 A discussion of the BIC method can be found in Burnham and Anderson (2004).

for the other three aspects are better explained by the linear specification. Even for anxiety, the BIC score for the linear specification is lower, suggesting that we should prefer this specification.

The LR statistic is only sufficiently high for anxiety ranking, since

$$\chi^2_{0.05}(2) = 5.991 \tag{15}$$

This leads to the rejection of the null hypothesis that $\beta_2, \beta_4 = 1$. The null is not rejected for each of the other three aspects, suggesting the linear specification is at least as valid as the non-linear specification for these.

TABLE 16: Maximum likelihood estimates of linear and non-linear ordered logit specifications for well-being rank.

	Linear specification ($\beta_2, \beta_4 = 1$)				Non-linear specification			
	Happy Rank	Satis Rank	Worth Rank	Anx Rank	Happy Rank	Satis Rank	Worth Rank	Anx Rank
$\hat{\beta}_1$	0.152	0.128	0.028	0.221	-9.714	0.152	0.612	1.204
$\hat{\beta}_2$	-	-	-	-	-13.449	0.920	0.005	0.306
$\hat{\beta}_3$	-0.054	0.102	-0.080	-0.119	-10.500	0.002	-0.415	-0.049
$\hat{\beta}_4$	-	-	-	-	-6.961	2.892	0.424	1.466
Constant 1	1.627	-2.520	-3.427	-1.220	0.892	-2.715	-3.408	-0.498
Constant 2	3.391	-0.765	-1.380	1.063	2.645	-0.964	-1.358	1.871
Constant 3	4.944	0.959	0.784	2.185	4.203	0.766	0.807	2.995
Observations	281	281	281	281	281	281	281	281
BIC	459.766	557.555	679.205	702.384	471.115	567.191	690.089	703.025
BIC linear - BIC nonlinear	-11.349	-9.636	-10.884	-0.641	-	-	-	-
LR statistic	-0.0724	1.641	0.392689	10.635	-	-	-	-

The estimates for β_1 are all positive when using the linear specification, which rejects the hypothesis in Section 2.2.2 that posits people will prioritise well-being aspects in which they are personally deficient. Instead, we see that having a higher level of well-being in a given aspect increases the priority given to that aspect. This relationship is weakest for worthwhileness, and strongest for anxiety. With the exception of the model fitted for life satisfaction ranking, the estimates for β_3 are negative. Therefore, for these aspects, it appears that whilst priority is increasing in own levels, it is decreasing in the beliefs about others' levels.

With the exception of the model fitted for worthwhileness ranking, I find that $|\hat{\beta}_1| > |\hat{\beta}_3|$. This suggests that an individual's priorities are more dependent on their own levels than their beliefs about others. This is consistent with the hypothesis put forward in Section 2.2.2. For the non-linear specifications, this also holds for three aspects, though this time the exception is happiness (for which the non-linear model is particularly weak, relative to the linear model).

As the linear specification provides a better fit, I estimate this model with separate regressions for each age band. Table 17 shows the estimates for β_1 and β_3 from these regressions. These estimates are plotted in Figure 13. When we look at the relationship between well-being levels and rankings, we see that happiness and anxiety exhibit stronger positive associations around middle age. In contrast, life satisfaction levels have the strongest influence on life satisfaction rank for the young and old. The trend for worthwhileness rank is less clear. These estimates suggest that there may be a common underlying factor that links how individuals think about happiness and anxiety. This corresponds to the same link observed from the priority weighting patterns in survey 1 (see Figure 8). I discuss this link further in Section 2.5.

Finally, Table 18 shows estimates from ordered logit regressions, with the inclusion of individual characteristics. As before, we see no substantial evidence of a non-linear life cycle trend in priorities. There is relatively little difference in the estimates for the influence of own and others' levels of well-being between this regression and the estimates from Table 16. Again, own levels are positively related with priorities for all aspects, save for worthwhileness. Worthwhileness level has no significant influence on worthwhileness rank.

Individuals with higher Neuroticism, as expected, give a higher priority to anxiety, at the expense of happiness in particular. The sizes of these estimates

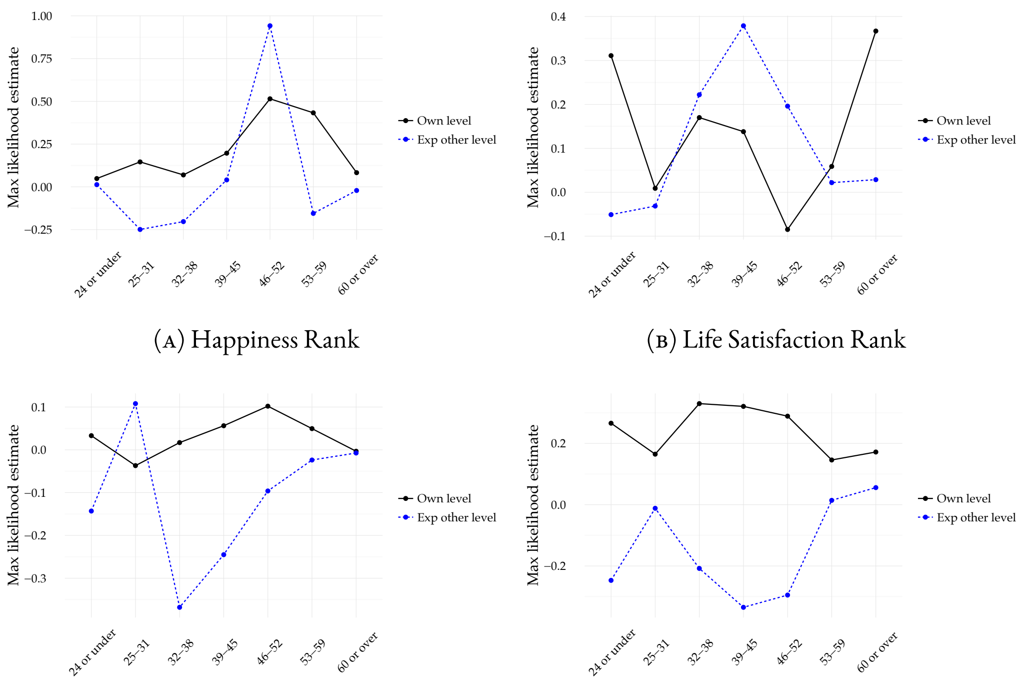
TABLE 17: Ordered logit models for well-being rank, with linear specification estimated for each age band.

Age Band	n		Happiness Rank	Life Sat Rank	Worthwhile Rank	Anxiety Rank
≤ 24	40	$\hat{\beta}_1$	0.0487	0.311*	0.0331	0.266**
		$\hat{\beta}_3$	0.0131	-0.0509	-0.143	-0.247
25 - 31	40	$\hat{\beta}_1$	0.146	0.00895	-0.0369	0.165
		$\hat{\beta}_3$	-0.249	-0.0314	0.108	-0.0117
32 - 38	40	$\hat{\beta}_1$	0.0699	0.170*	0.0169	0.330***
		$\hat{\beta}_3$	-0.203	0.222	-0.368**	-0.208
39 - 45	40	$\hat{\beta}_1$	0.197	0.138	0.0563	0.321**
		$\hat{\beta}_3$	0.0408	0.379	-0.245	-0.335
46 - 52	41	$\hat{\beta}_1$	0.515**	-0.0847	0.102	0.289**
		$\hat{\beta}_3$	0.942**	0.196	-0.0961	-0.295*
53 - 59	40	$\hat{\beta}_1$	0.433	0.0588	0.0496	0.146
		$\hat{\beta}_3$	-0.155	0.0219	-0.0238	0.0142
≥ 60	40	$\hat{\beta}_1$	0.0831	0.367***	-0.00327	0.172
		$\hat{\beta}_3$	-0.0210	0.0287	-0.00742	0.0558

Standard errors and constants omitted for brevity. *** p<0.01, ** p<0.05, * p<0.1

are large. For example, the reduction in odds of a high happiness ranking caused by a point increase in Neuroticism is almost three times the magnitude of the increase in odds due to a point increase in happiness level. Surprisingly, however, we do not observe the converse for Extraversion. The most noteworthy personality trait influence, aside from Neuroticism, is that of Agreeableness. More Agreeable people place a higher priority on worthwhileness. It is not immediately clear why this would be the case.

Healthier people also place a higher priority on worthwhileness, relative to the other aspects. This is somewhat intuitive. Worthwhileness of life is a long



(A) Happiness Rank

(B) Life Satisfaction Rank

(C) Worthwhileness Rank

(D) Anxiety Rank

FIGURE 13: Plot of ordered logit estimates for β_1 (solid black lines) and β_3 (dotted blue lines) by age band, using data from Table 17.

term evaluation of life quality and value. One would expect that this is only considered of primary importance for those individuals who are free of more primitive concerns like insufficient income or ill health. This is corroborated by the finding that those with a higher level of income also attach a higher priority to worthwhileness, though the magnitude of this relationship is much weaker than that of health level.

Males, those with children, and individuals with a higher level of education place higher priority on happiness and life satisfaction, and lower priority on worthwhileness and anxiety. The preference given to happiness and life satisfaction by males is consistent with evidence from survey 1 (see Table 13). Those who are married or living with a partner place a higher priority on happiness, relative to the other aspects. The unemployed place higher priority on happiness and anxiety. This may be due to the fact that these are shorter term hedonic considerations. As argued for health level and income, longer term evaluations are likely to be considered less important than an immediate improvement in situation for someone who is unemployed.

Perhaps the most surprising results are for smoking frequency. Recall that individuals who smoke more frequently are found to be more impulsive (Reimers et al., 2009), in that they prefer smaller-sooner payoffs over larger-later ones. Those who smoke more frequently place a higher priority on happiness and worthwhileness than they do on life satisfaction and anxiety. The result for happiness is again intuitive - those who are more impatient are likely to place a higher value on hedonism. However, it is not clear why smokers give a higher priority to worthwhileness, and a significantly lower priority to anxiety.[11]

2.5 DISCUSSION

The non-linear life cycle pattern for well-being priorities obtained from survey 1 does not match the flat profile found in survey 2 when fitting a basic polynomial in age to the data. Yet, there is some indication from the additional results in survey 2 that there may be some general themes underlying the prioritisation process.

Survey 1 shows that middle-aged individuals give a higher weight to happiness and anxiety, and lower weights to life satisfaction and worthwhileness. Despite not finding direct confirmation for these patterns in survey 2, Figure 12 shows that a similar (though not statistically significant) grouping of aspects emerges after separating for gender. The age-band separated models in Table 17 show evidence of a greater correspondence between happiness and anxiety levels, and their respective ranking around middle age.

The reason why this grouping may be occurring was alluded to in Section 2.4.3. The 4 aspects of well-being can be grouped by those that measure Affective Well-being (AWB), and those that measure Cognitive Well-being (CWB). The former is related to shorter term mood; the latter to a more holistic evaluation of life (Luhmann et al., 2012). The distinction between these forms of well-being is

11 Whilst a direct measure of discounting was also collected, the data were noisy, with two respondents exhibiting multiple switching points in the task. Inclusion of this variable did not add any meaningful information, and so these additional regressions are omitted from the chapter. They are available upon request.

Table 18: Ordered logit regressions for the determinants of well-being rank, inclusive of full set of control variables.

	Happy Rank	Satis Rank	Worth Rank	Anx Rank
Age	-0.026	0.0219	0.00698	-0.0238
Age^2	0.000161	-0.000226	-2.62E-05	0.00035
Happiness Level	0.117	-	-	-
Happiness Other	-0.0621	-	-	-
Life Satisfaction Level	-	0.182***	-	-
Life Satisfaction Other	-	0.109	-	-
Worthwhileness Level	-	-	-0.0928	-
Worthwhileness Other	-	-	-0.0951	-
Anxiety Level	-	-	-	0.181***
Anxiety Other	-	-	-	-0.101
Agreeableness	-0.207	-0.0319	0.187*	-0.0638
Conscientiousness	-0.149	-0.132	0.0105	0.133
Extraversion	-0.0301	0.052	-0.0255	0.0614
Neuroticism	-0.348**	-0.0675	-0.0764	0.199*
Openness	-0.0503	-0.135	0.0612	0.108
Married/Cohabiting?	0.126	-0.0721	-0.0473	-0.0756
Has children?	0.0917	0.225	-0.0446	-0.385
Employed?	-0.295	0.169	0.209	-0.0276
Male	0.334	0.323	-0.0517	-0.598**
Health Level	-0.11	-0.245	0.462***	-0.214
Income Band	-0.0415	-0.0461	0.0498	0.018
Freq of Smoking	0.197**	-0.0749	0.131*	-0.170**
Education Level	0.0559	0.0351	-0.122	-0.0757
Constant 1	-2.856	-4.203**	-1.217	-1.452
Constant 2	-0.995	-2.441	0.893	0.938
Constant 3	0.596	-0.674	3.173*	2.103
Observations	281	281	281	281

Standard errors omitted for brevity. *** p<0.01, ** p<0.05, * p<0.1

important, as the determinants of each form differ. The key difference between the two is the role that hedonic adaptation (Brickman and Campbell, 1971; Diener, Lucas, and Scollon, 2006) plays in AWB. It is widely accepted in psychology that most positive or negative shocks to happiness are transitory.[12] However, when it comes to CWB measures such as life satisfaction, factors such as income and significant life events can have a permanent impact, despite having a lower short-term variance (Luhmann et al., 2012).

Anxiety and happiness can both be thought of as affective states, and hence forms of AWB. These are feelings and emotions that take place in the short-run, i.e. in response to a particular stimulus or situation. In economic parlance, one might call these 'flow' measures of well-being. In contrast, life satisfaction and worthwhileness of life are wider in their scope of consideration. They require the respondent to take into account their entire life history (or at least a significant portion of it). We may therefore consider them to be 'stock' measures of well-being. It should be noted, however, that happiness cannot be considered exactly equivalent to flow utility (Kimball and R. Willis, 2006). Instead, Kimball and R. Willis (2006) split affect into *baseline mood* (i.e. long-run happiness), and *elation* (i.e. short-run happiness).

Analysis of the APS well-being data, along with the data from survey 2, confirms this grouping. Table 19 shows the results of a factor analysis on the levels of the four well-being aspects, using the principal factors method with an oblique promax rotation of power four.[13] A rule of thumb states that loadings above 0.32 are significant at the 1% level for sample sizes above 300 (Yong and Pearce, 2013). For the very large APS sample, this threshold is likely to be lower.

The factor loadings for both sets of data support an underlying relationship between life satisfaction and worthwhileness of life, and between happiness and anxiety. Happiness has a significant loading on both factors, which suggests that the measure of happiness in generalised studies captures a more holistic assessment of well-being than short term transitory changes alone. This supports the claim of Kimball and R. Willis (2006).

12 Though Easterlin (2005) explains that the data rule out a 'setpoint' of happiness, in the sense that there does not appear to be complete adaptation.

13 An oblique rotation was used as opposed to an orthogonal rotation, since we would not expect cognitive and affective forms of well-being to be completely independent from each other. Applying a varimax rotation does not yield qualitatively different results.

Figure 14 plots loadings over the life cycle when we perform factor analysis separately on each age band. There are more age bands in the APS data, due to a higher upper age limit. We see that the loadings on each factor are relatively stable over the life cycle for each aspect of well-being for the APS data.

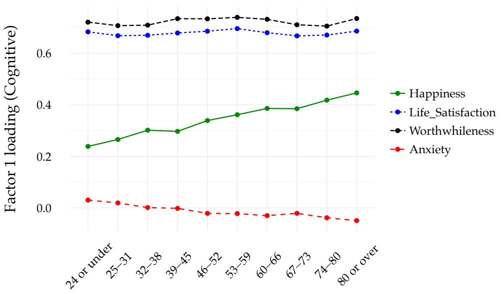
There are two notable exceptions to this. First, the loading of happiness on the factor representing CWB is increasing over the life cycle in the APS data. Whilst the loadings from survey 2 exhibit more noise, we can see some indication of the same trend. If we relate this to the framework proposed by Kimball and R. Willis (2006), then the baseline mood component of happiness appears to be dominating the elation component as one ages. Despite not finding evidence of prioritisation differences across the life cycle, if happiness is perceived as being closer to life satisfaction for those that are older, then there may be an implicit increasing preference for CWB over the life cycle.

According to Socioemotional Selectivity Theory (Carstensen, 2006), the young pursue goals that optimise the future. Close proximity to the end of life for older individuals leads them to pursue shorter term goals (Löckenhoff, 2011). This means that even though their discount factors may be lower (Green et al., 1996), the young appear to have longer time horizons than the middle-aged. In light of this, one might expect an increasing preference for AWB policies with age. However, it is also the case that older adults tend to focus further into the past and less into the future than younger adults (Löckenhoff and Rutt, 2015). It is possible that this backward looking evaluation of life may be contributing to an increasing emphasis on CWB with age.

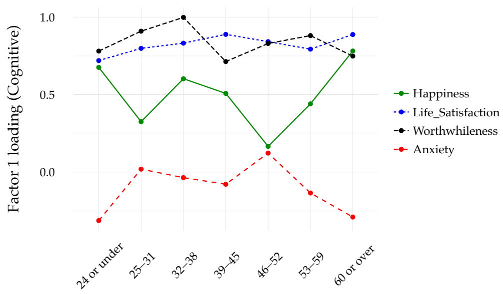
Second, the loadings of happiness and anxiety on the factor representing AWB switch polarity for those aged 74 and above (the survey 2 loadings in Figure 14 (D) are extremely noisy, though one might argue a similar pattern may be present). In other words, happiness begins to correspond to negative affect, and anxiety begins to correspond to positive affect. This is somewhat puzzling. It is unclear whether this can be related to research on older individuals, or whether it is merely an anomaly in the data. The latter seems unlikely, given that the total number of observations from individuals 74 and above in the APS data is 19,308.

TABLE 19: Factor analysis of well-being levels from APS 2013-14 (n=165,122) and survey 2 (n=281), showing the rotated loadings on cognitive and affective well-being.

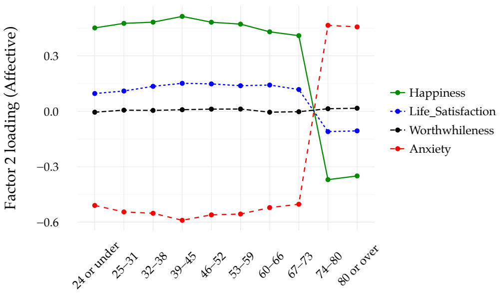
	APS 2013-14		Survey 2	
	Factor 1 "Cognitive"	Factor 2 "Affective"	Factor 1 "Cognitive"	Factor 2 "Affective"
Happiness yesterday	0.3460	0.4550	0.4181	0.4833
Life satisfaction	0.6763	0.1379	0.8048	0.1243
Worthwhileness of life	0.7236	0.0042	0.8218	0.0082
Anxiety yesterday	-0.0156	-0.5365	-0.0346	-0.5277



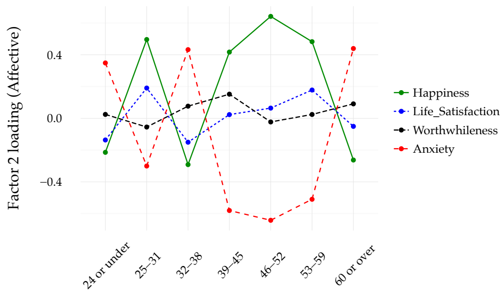
(A) Factor 1 ("Cognitive") - APS data



(B) Factor 1 ("Cognitive") - Survey 2 data



(C) Factor 2 ("Affective") - APS data



(D) Factor 2 ("Affective") - Survey 2 data

FIGURE 14: Graphs showing factor loadings for each well-being aspect over the life cycle. Data for (A) and (C) is from APS 2013-14. The minimum number of observations for an age band in the APS data was 7,638 for those over 80. Data for (B) and (D) is from survey 2.

2.6 SUMMARY AND CONCLUSION

In summary, this study finds the following. First, there is no consistent evidence that supports changing well-being priorities over the life cycle (based on cross-sectional data containing a mixture of birth cohorts). Despite this, older individuals may implicitly be exhibiting a preference for cognitive well-being over affective well-being, due to an increasing factor loading of happiness on cognitive well-being with age.

Second, when an age-stratified sample is used, the mean rank ordering of well-being aspects (from highest to lowest rank) is: life satisfaction, worthwhileness of life, anxiety, and happiness yesterday. The ranking is stable across age groups. On average, individuals overestimate the rank they believe others will give to happiness, and underestimate the rank they believe others will give to worthwhileness and anxiety.

Third, an individual with a higher level of happiness, life satisfaction, or anxiety, is more likely to give a higher priority ranking to that respective aspect. This relationship is strongest during middle age for happiness and anxiety. Beliefs about others' levels of well-being generally have less of an impact on prioritisation than own levels of well-being. The fact that life cycle levels of well-being show a mid-life dip, but priorities do not, suggests that levels are being moderated by other factors that also determine prioritisation.

Fourth, individuals with the following characteristics show a clear prioritisation preference for one aspect over the other three: more Agreeable people (worthwhileness); more Neurotic people (anxiety); those married or cohabiting (happiness); healthier people (worthwhileness). In addition, more frequent smokers (a proxy for impatience) prefer happiness and worthwhileness to life satisfaction and anxiety.

The literature on well-being prioritisation is still in its infancy, but its overall goal is of prime importance: to inform optimal resource allocation when seeking to improve society. As the focus of the developed world shifts from increasing raw incomes to improving the general well-being of its inhabitants, this line of enquiry promises to become increasingly pertinent for public policy. It is important that we understand which aspects people value, how they value them, and why they form these valuations. The findings from this study contribute to this understanding by providing a first attempt to identify determinants of prioritisation over the life cycle.

Part II

NEW BEHAVIOURAL IDEAS FOR ECONOMIC SETTINGS

3

DO PEOPLE ADJUST FOR EXTREME REVIEW SCORE BIAS?

An important implication of the internet on modern economic life is the increasing reliance on online reviews to inform consumption decisions. Yet, extremely positive or negative reviews may be subject to a large degree of bias, as well as conflicts of interest. I introduce a model that proposes individuals weight extreme review scores to adjust for this potential bias. A randomised experiment on 501 individuals finds insufficient evidence that extreme review scores are being weighted when evaluating the quality of a good. Hence, individuals are susceptible to being influenced by deliberately falsified extreme reviews, which is likely to reduce consumer surplus. I also find that personality traits have no significant moderating effect on product quality evaluation.

3.1 INTRODUCTION

The importance of customer reviews for products and services has increased, due to the increasing proportion of transactions we make online. In the U.S., after adjustment for seasonal variation, the proportion of total retail sales accounted for by e-commerce has more than doubled, to 7.5%, in less than a decade (Figure 15).[1]

Word-of-mouth recommendations have existed since the dawn of communication. In modern times, officially accredited ratings have been in use for some time (e.g. in the financial industry). Independent product reviews and consumer advice have long been dispensed through the media. However, the internet has resulted in the standardisation (and abundance) of word-of-mouth. Online purchasing has placed an emphasis on peer reviews from consumers themselves. Detailed ratings and reviews by other customer can be found next to the vast major-

1 These data can be found at census.gov/retail/mrts/www/data/pdf/ec_current.pdf

ity of goods available for purchase. As a result, in a market environment where physical access to a product is not always possible, a purchasing decision is likely to be highly influenced by reviews (e.g. Luca, 2011).

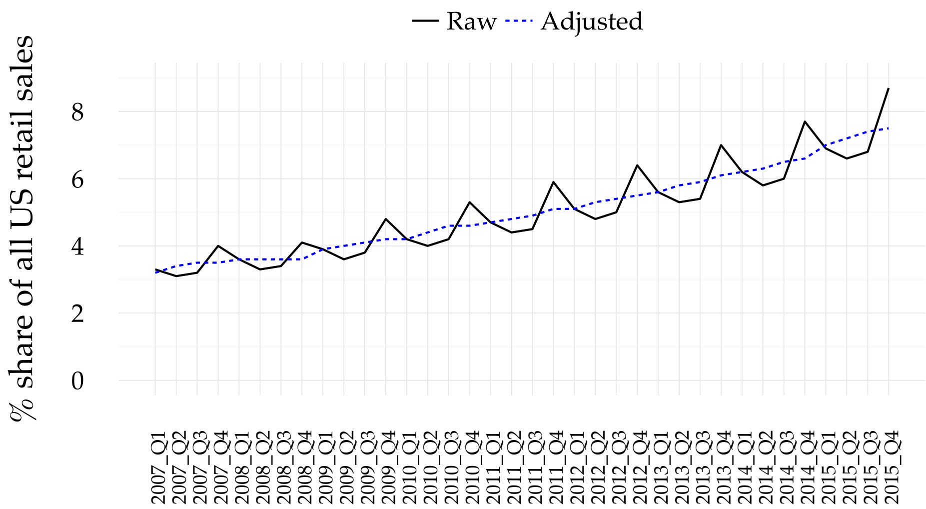


FIGURE 15: E-commerce sales in the U.S. have steadily increased as a proportion of total retail sales. Source: U.S. Census, via Department of Commerce. The dotted 'adjusted' trend removes seasonal fluctuations, likely due to increased consumer activity during the holiday period.

The reliance on review scores means that the set of strategies for sellers is now vastly different than it was prior to internet shopping. In particular, some firms have taken to writing anonymous glowing reviews of their own products, and even buying the services of 'fake' reviewers.[2] Though less publicised, it has also been shown that falsified negative reviews are left for competing goods (Mayzlin, Dover, and Chevalier, 2014). This has led to legal action being taken by online marketplace websites that wish to keep reviews on their platform free of bias.[3] In developed countries, such as the UK and Australia, the legal system allows for action to be taken if deliberate attempts to mislead are detected (Hunt, 2015). Whilst automated methods of detection have been proposed (e.g. Lim et al.,

2 One firm, Taser, has even defended its practice of allowing staff to write reviews without disclosing affiliation (arstechnica.com/the-multiverse/2015/12/bad-reviews-for-taser-documentary-on-amazon-itunes-seem-to-come-from-taser-employees/)

3 For example, the online retailer Amazon sued over 1000 professional 'fake reviewers' in 2015 (nbcnews.com/tech/internet/amazon-files-suit-against-1-000-people-fake-reviews-n447101)

2010; Mukherjee, B. Liu, and Glance, 2012) as a result of sentiment analysis and opinion mining in computer science (see B. Liu and L. Zhang, 2012, for a review), it remains extremely difficult to identify and remove reviews posted by every individual that has a conflict of interest.

Therefore, consumers must take the potential for bias into account when valuing a good based on review scores. However, researchers do not yet know how consumers use review scores to value a good in terms of its quality, nor in monetary terms. Hence, this chapter seeks to determine whether individuals implicitly place a weight on extreme review scores, in order to adjust for this potential bias.

I focus on extreme scores (those at the top and bottom end of the scoring range), since these are the most likely to be falsified, given the aim is to alter the perception of a good in the most dramatic way possible. I also investigate whether individual differences (in the form of personality traits) mediate behaviour in this setting. It is possible that more sophisticated forms of falsification may involve leaving a number of non-extreme reviews. Given the general J-shaped distribution of review scores in practice (discussed later), however, this is unlikely to have much of an impact for the majority of goods with a reasonable number of reviews.

In order to test this hypothesis, I first develop a weighted-mean model that applies weights to review scores at the top and bottom ends of the scoring range. I then perform a simple randomised experiment, using real goods taken from the websites Amazon.co.uk and TripAdvisor.co.uk, in order to test whether individuals exhibit this weighting pattern in practice.

At a high level, the experiment involves treatment conditions where the review scores for various goods are manipulated. Reviews at the extremes of the range were either partially or entirely removed from the review score distribution shown to individuals (in a manner similar to the judging process used in figure skating competitions). If individuals are applying a weight to extreme review scores in the way specified by the model, the quality of a good should be perceived (in most cases) as being higher in one of the treatments than in the control condition.

The data obtained from the experiment shows that the weighted-mean model is better than the mean review score, and a model based on range-frequency theory, at predicting the effects of review score manipulation on perceived good quality. However, the overall predictive power of this model is still relatively poor. This suggests that consumers are not fully adjusting for extreme review bias. They may therefore be vulnerable to being manipulated into the purchase of potentially low quality products, thus harming welfare. This is especially likely to be the case for individuals low in the personality factor Agreeableness, or high in the personality factor Neuroticism.

3.1.1 *Literature*

The economic potential of computers to make product evaluations cheap and ubiquitous was recognised by Avery, Resnick, and Zeckhauser (1999). Reviews can be thought of as public goods, as they are non-rivalrous and are (usually) non-excludable. Since it is costly to purchase a good and evaluate it early rather than waiting for more reviews, this generates an opportunity cost. Therefore, the market will not produce reviews efficiently by itself. Avery, Resnick, and Zeckhauser (1999) propose a pricing mechanism to resolve this inefficiency. However, this requires a benevolent broker and two of three possible conditions to be satisfied. In practice, online product reviews are not centrally organised in this way. Hence, we are likely to be in a state of the world where the number of individuals reviewing, and the information disseminated, is sub-optimal. Another strand of theoretical literature focuses on developing optimal mechanisms that exploit cascades and herding to maximise welfare by withholding a subset of the available information from certain individuals (Kremer, Mansour, and Perry, 2014).

Hu, Pavlou, and Jennifer Zhang (2006) show that the only way review scores signal true quality is if all consumers leave a review score, or that those consumers that do review are equally likely to 'moan' about a bad product as they are to 'brag' about a good one. However, there is endogeneity in terms of which individuals choose to leave reviews, leading to under-reporting (Koh, Hu, and Clemons, 2010). Hence, it is likely that neither of the requisite conditions for an unbiased signal will be satisfied. This is corroborated by the fact that experimental review score distributions are unimodal (Hu, Jie Zhang, and Pavlou, 2009).

A more applied analysis of product reviews has predominantly been confined to the business and management literatures. In one notable exception, Mayzlin, Dover, and Chevalier (2014) infer, using a difference-in-differences approach, that more fake positive and negative reviews can be found for hotels on TripAdvisor than for the same hotels on Expedia. This is due to the fact that Expedia requires reviewers to have booked into the hotel through their website, whereas TripAdvisor does not.

The type of good under consideration has been found to be important in determining the usefulness of extreme reviews. Mudambi and Schuff (2010) find that extreme review scores for experience goods (goods that must be experienced before they can be reasonably valued, e.g. music) are rated as being less helpful than moderate reviews by customers on Amazon.com. However, the opposite effect was found for books. Mudambi and Schuff (2010) suggest that the difference may be explained by the prior attitude of the consumer towards the product. This implies that personality traits and other individual characteristics might account for some of the variation in review score perception.

In this chapter, I do not consider information about written reviews or reviewer reputation. However, these factors may be important in determining the value of a particular review in any individual's belief updating process (e.g. Hu, L. Liu, and J. J. Zhang, 2008). Mudambi and Schuff (2010) find that longer reviews for search goods (goods that can be compared easily using objective attributes e.g. cameras) are considered more helpful. This effect is smaller for experience goods, potentially due to the stronger misalignment between text reviews and review scores (Mudambi, Schuff, and Z. Zhang, 2014). Other work has focused on disclosure of reviewer identity being positively related to the perceived helpfulness of a review, and also quantity of sales (Forman, Ghose, and Wiesenfeld, 2008). Hu, L. Liu, and J. J. Zhang (2008) show that the impact that a product review has on sales diminishes the longer the product has been on the market.

Literature in computer science has focused on developing algorithms that attempt to elicit the 'useful' information component from review score data. This problem has proved challenging to solve, since the precise proportion of fake reviews is unknown.[4] Using a behavioural program designed to identify and remove spam reviewers makes a larger impact on overall review scores than merely discarding reviews flagged as unhelpful (Lim et al., 2010). However, it is not clear how this affects a consumer's valuation of a good.

The distribution of review scores is likely to play a factor in how a product is evaluated. In their classic economic theory paper, Rothschild and Stiglitz (1970) explain that if two random variables have the same mean, the one with the larger variance may be preferred by some risk averse individuals. Therefore, it is not clear whether a larger spread of review scores is more or less helpful for a consumer's valuation of a product.

Park and Sabourian (2011) show that in financial markets, herding should only take place in theory if private information follows a U-shaped distribution (i.e. one that emphasises extreme outcomes are more likely). Contrarianism (i.e. behaving in a manner that is against the crowd) occurs only when private information follows a hump-shaped distribution.

These results suggest that people may be prone to following the crowd in reviewing products when their prior signal is that a product can either be 'good' or 'bad'. Hu, Pavlou, and Jennifer Zhang (2006) find that review score distributions for around half of the products on Amazon.com are bimodal. Furthermore, extremely high ratings are more common than extremely low ratings, creating a J-shaped distribution (Hu, Jie Zhang, and Pavlou, 2009). As the number of reviews increases, there will not necessarily be convergence towards a true score for the product. Both of these findings point towards U-shaped signals facilitating herding behaviour towards the extremes.

The aforementioned literature does not attempt to elicit perceptions of the value or quality of a good, based on review scores. However, the psychology literature has produced substantial work that has built upon Range Frequency Theory (Parducci, 1965). This posits that people take information about the rank position of a good within a distribution, and the range of the distribution, in order to form a valuation for that good. The valuation is determined by taking a linear combination of the range and rank effects.

4 Hu, Bose, et al. (2012) estimate that 10.3% of products on Amazon.com are subject to manipulated reviews.

Subsequent work has confirmed predictions of the theory hold in experimental data. Parducci (1968) found that the average moral judgement on an act of 'bad behaviour' is harsher when other acts in the set of scenarios are milder. In other words, when a set of scenarios has a positively skewed distribution (i.e. a lower frequency of extreme scenarios), then a particular scenario which is near the upper end of that distribution will be given a higher (harsher) judgement valuation than when that same scenario is part of a negatively skewed set.

Parducci (1968) also found that when subjects were given a sequence of money payoffs from distributions with the same expected value, they were more satisfied when the values were drawn from a negatively skewed distribution (i.e. when the mean is to the right of centre) than a positively skewed one. Range frequency effects have been shown to hold in various contexts, such as in the perception of drink sweetness (Riskey, Parducci, and Beauchamp, 1979).[5]

Range Frequency Theory suggests that the skewness of the review score distribution is likely to influence an individual's valuation for that good. A key difference between the present study and the experiments on range frequency is that the distributions of review scores are explicitly observable to consumers. One would therefore expect an even stronger effect of distribution on valuation.

The remainder of the chapter is organised as follows. Section 3.2 introduces the weighted-mean model, and describes a model based upon Range Frequency Theory. Section 3.3 outlines the experimental design, and provides predictions for the experiment using the models described in Section 3.2. Section 3.4 presents and discusses the results of the experiment. Finally, Section 3.5 concludes.

3.2 MODELS

The central question of interest in this study is: given that there is some degree of prior public knowledge that online reviews for goods may be fake or biased, are consumers taking this information into account in their evaluation of a product's quality? There are many possible approaches that one might take in order answer this question. In this chapter, I consider one particular approach – namely that consumers adjust extremely high and extremely low review scores

5 See Tripp and Brown (2016) for a summary of the findings in this area.

by applying a weight to them. This weight binds more strongly when the total number of extreme reviews is low.

One might question why it is assumed that reviews at the extreme ends of the scale are those most likely to be fake or biased. The logic behind this assumption is as follows. Upon seeing a review score distribution and a mean review score, perhaps the simplest and most naive evaluation for the quality of that good would be the mean score. One might think of this as a 'level-0' approach in the context of level-k reasoning (Stahl and Wilson, 1994), where no adjustment is made. If an individual or firm wanted to positively (negatively) influence perception of a good's quality, then their best response to a level-0 consumer would be to give the highest (lowest) possible rating to the good. A more sophisticated consumer, knowing this best response, would therefore apply a weight in the first instance to review scores at the highest and lowest ends of the review scale. Of course, it is possible that levels of sophistication go even further than this. However in this study, I aim to test whether or not consumers are applying this first-order 'level-1' response.

An implicit assumption that arises from the former discussion is that individuals will apply the same weighting approach to any review score distribution they see. Given that most online product review distributions follow the same 'J-shape' (see the discussion in Section 3.1.1), it is plausible that individuals learn to form one strategy over time, which would then be applied upon exposure to more unusual (e.g. 'U-shaped') review score distributions.[6]

Suppose that a product can be given an integer review score r, where $r \in \{1, 2, 3, ..., R\}$.[7] The mean review score for a product is given by:

$$\mu = \frac{1}{N} \sum_{r=1}^{R} r n_r \tag{16}$$

where n_r represents the number of reviews with score r, and $N = \sum_{r=1}^{R} n_r$ (i.e. the total number of reviews for that product).

6 The case when different models are used depending on the shape of the review score distribution is left for future research.
7 Ideally, R is odd, so that there is a clear middle score.

I propose a weighted version of the mean score from which to obtain a quality evaluation for a good:

$$\mu_w = \frac{1}{N}\sum_{r=1}^{R} w(.)rn_r \tag{17}$$

There are two processes that generate the weighting function $w(.)$. First, because I assume 'extreme reviews' (i.e. reviews with scores of either 1 or R) are most likely to be biased or fabricated, reviews with those scores are weighted. In particular, reviews with a score of R will be given a negative weight, whilst reviews with a score of 1 will be given a positive weight (with a weight larger than 1). This is because if an extremely high review was not genuine (and one was aware of this), one would expect a consumer to *reduce* their valuation of the good, relative to the value suggested by the raw mean review score. In contrast, if an extremely low review was not genuine, then a consumer would *increase* their valuation of the good, relative to the value suggested by the raw mean review score.

Second, this weighting should only bind for sufficiently small numbers of reviews with a particular score. As n_r increases, r is increasingly likely to be a true reflection of the quality of a good.[8] Based on these two processes, we can explicitly define the weighting function as follows:

$$w(r,k,\alpha,\beta) = 1 + g(r)f(n_r,\alpha,\beta,T) \tag{18}$$

The function $g(r)$ represents the first process. The simplest way to model this is to use a piecewise function, where scores of 1 are positively weighted, scores of R are negatively weighted, and all other scores receive a weighting of 1 (i.e. they are unweighted). I take the positive weight for $r = 1$ to be $\frac{R+1}{2}$, which is the midpoint of the review scale. This essentially has the effect of 'cancelling out' reviews with score 1, by pushing them towards a neutral, middle value. I take the negative weight for $r = R$ to be -1. This is the simplest integer weight

8 An astute consumer may realise that if the frequency of r scores is sufficiently high, that it could also be the result of herding. In this case we could add the possibility of the review score weightings being 'reactivated' after a certain frequency is reached. I leave this as a future extension.

that reduces the overall weighted mean in the presence of suspicious reviews at the uppermost end of the review scale. The functional form for $g(.)$ is therefore:

$$g(r) = \begin{cases} \frac{R+1}{2} & \text{if } r = 1 \\ -1 & \text{if } r = R \\ 1 & \text{otherwise} \end{cases} \tag{19}$$

The function $f(n_r, \alpha, \beta, T)$ represents the second process. This generates what may be thought of as an 'unreliability' score, bounded by 0 and 1. For a high enough number of reviews n_r, $f(.) = 0$ so that $w(.) = 1$. I use T to denote the threshold, such that for $n_r > T$, $f(.) = 0$.[9] On the other hand, as $n_r \to 0$, we have that $f(.) \to 1$. In other words, the fewer the number of reviews at a given score, the less reliable that score is deemed to be as an estimate of the true quality of the good. The beta density function allows us to model this, whilst providing the flexibility to easily change the shape of $f(.)$ by changing the parameters α and β. The general form of $f(.)$ can be derived as:

$$f(n_r, \alpha, \beta, T) = c\left(\frac{n_r}{T}\right)^{\alpha-1}\left(1 - \frac{n_r}{T}\right)^{\beta-1} \tag{20}$$

where c is a normalising constant, given by:

$$c = \frac{1}{\left(\frac{n_r^*}{T}\right)^{\alpha-1}\left(1 - \frac{n_r^*}{T}\right)^{\beta-1}} \tag{21}$$

and n_r^* is the maximiser of the beta density function:

$$n_r^* = \frac{T}{\frac{\beta-1}{\alpha-1} + 1} \tag{22}$$

In order to compare valuations generated by the weighted-mean model with a metric more sophisticated than the mean, I use a model based on the Range-

9 Note that if n_r is large enough for all r, we have $\mu = \mu_w$.

Frequency model (RF). Riskey, Parducci, and Beauchamp (1979) describe the category rating of a stimulus as being of the general form:

$$RF(w) = wR_e + (1-w)F_e \tag{23}$$

where R_e represents the range effect (the position of a draw in the range of the distribution), and F_e represents the frequency effect (the percentile rank of the draw).[10] The relative weighting placed on each of these effects is $w \in [0,1]$.

There are two issues with using this model directly. First, the model evaluates a draw (stimulus) from a distribution. Whilst consumers viewing a review score distribution do not draw a stimulus from that distribution per se, it seems reasonable to take the mean rating as being the stimulus, since this is the most accessible information signalled to them. Second, the model assumes individuals receive stimuli from the domain of the distribution, which is generally discrete. However, using the mean score as the stimulus will mean that the stimulus lies within the continuous interval $[1, R]$. I make the discrete distribution continuous by assuming a straight line connects n_r for each r (in a construct that resembles an upper envelope). See Figure 16 for illustration.

Hence, in the context of the current study, the range effect is given by:

$$R_e = \frac{\mu - 1}{R - 1} \tag{24}$$

and the frequency effect is given by:

$$F_e = \frac{n_{\lfloor \mu \rfloor} + h}{N} \tag{25}$$

where h is a measure of the change in the density of the distribution, given a change in the domain from $\lfloor \mu \rfloor$ to μ:

$$h = (\mu - \lfloor \mu \rfloor)(n_{\lceil \mu \rceil} - n_{\lfloor \mu \rfloor}) \tag{26}$$

An assumption I make in applying both models is that individuals all behave according to one model, with one set of parameters. It is a simplifying assump-

10 I have used different notation to Riskey, Parducci, and Beauchamp (1979), so as to avoid confusion with parameters in the weighted-mean model.

tion, in order to assess the relative performance of the models as general descriptors of behaviour. However, it is of course quite likely that parameters vary between individuals. It is also possible that there are different 'types' of individuals, that behave according to different models. Whilst the latter point is partially addressed later in the chapter by testing hypotheses concerning two personality factors, the broad issue of heterogeneity is left as an extension.

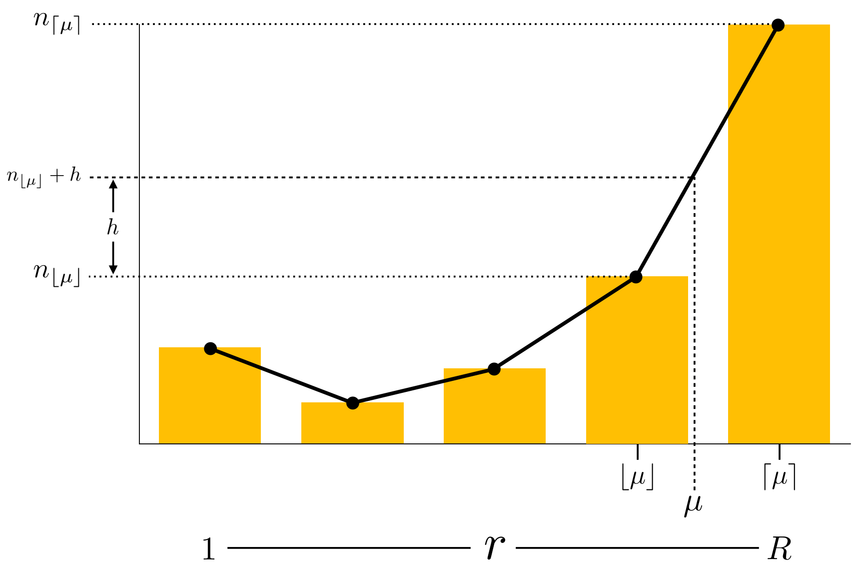


FIGURE 16: An illustration of the parameters used to calculate the range (R_e) and frequency (F_e) effects, given a review score distribution.

3.3 DESIGN AND PREDICTIONS

I have argued that, due to the presence of fake reviews and the increased likelihood that an extreme review score is subject to large amounts of bias, individuals may be implicitly applying weights to reviews with these scores. Therefore, the central hypothesis to be tested is the following:

Hypothesis 3.1 *Individuals implicitly overweight 1-star reviews, and negatively weight 5-star reviews.*

Two sub-hypotheses can be drawn from the central hypothesis above, since it may be the case that individuals are only applying weights to one of the two extremes of the review score range:

Hypothesis 3.2 *Individuals implicitly overweight 1-star reviews (but do not weight 5-star reviews).*

Hypothesis 3.3 *Individuals implicitly apply a negative weight to 5-star reviews (but do not weight 1-star reviews).*

If the weighted-mean model outperforms predictions generated by the raw mean and range-frequency approaches, this would support the central hypothesis. Therefore, the experimental design is based on a need to differentiate the predictions generated by each model.

3.3.1 *Experimental Design*

Review score distributions found online are, more often than not, negatively skewed. If valuations are judged by individuals as being dependent on the mean review score, then a mean-preserving removal of extreme scores to make the distribution less negatively skewed should have no effect on subjective evaluations for that good. However, if individuals are suspicious of the abundance of high review scores because they suspect them as false or biased, they may actually value the good more highly in the *manipulated* distribution.

The design of the experiment is based upon these manipulations of review score distributions for different types of good. Individuals are asked to rate a good in terms of its quality. I also ask for their maximum Willingness to Pay (WTP) for each good (though these data are likely to be noisier, given that they will be influenced by income and personal preferences). To maintain external validity, real goods (and their review scores) are obtained from the websites Amazon.co.uk and TripAdvisor.co.uk.[11] Each product or service on these websites can be reviewed on a 1-5 scale by registered users (i.e. $R = 5$). Since anyone with an account can review any product, review falsification and bias is possible.

11 Amazon is the largest online retailer in the UK. TripAdvisor is the largest travel-oriented review website.

10 goods were selected: 5 search goods, and 5 experience goods. For each type, goods were chosen to cover the following criteria: highly rated with high N; poorly rated with high N; highly rated with low N; poorly rated with low N; and middle rated with a roughly even split of bottom and top reviews (i.e. a U-shaped score distribution). Highly (poorly) rated in this context refers to mean scores that are above (below) the mid-point of the review scale.

For each good, a brief description of the good is shown, along with one or two images. Branding is stripped from the images in order to minimise the effects of prior good knowledge (or preferences towards certain brands) on valuation. Along with each good, the review score distribution is shown, along with mean review score. The mean review score is the only summary statistic that is explicitly available to participants.[12] This information is presented in a format which is in keeping with the style of the original website. See Figure 17 for an example. Participants report the quality of each good on a 0-100 point scale. Their WTP can be any non-negative dollar amount.

12 Amazon shows the mean numerically, and also visually. TripAdvisor only shows the mean visually. I preserve this difference in the experiment to maximise external validity.

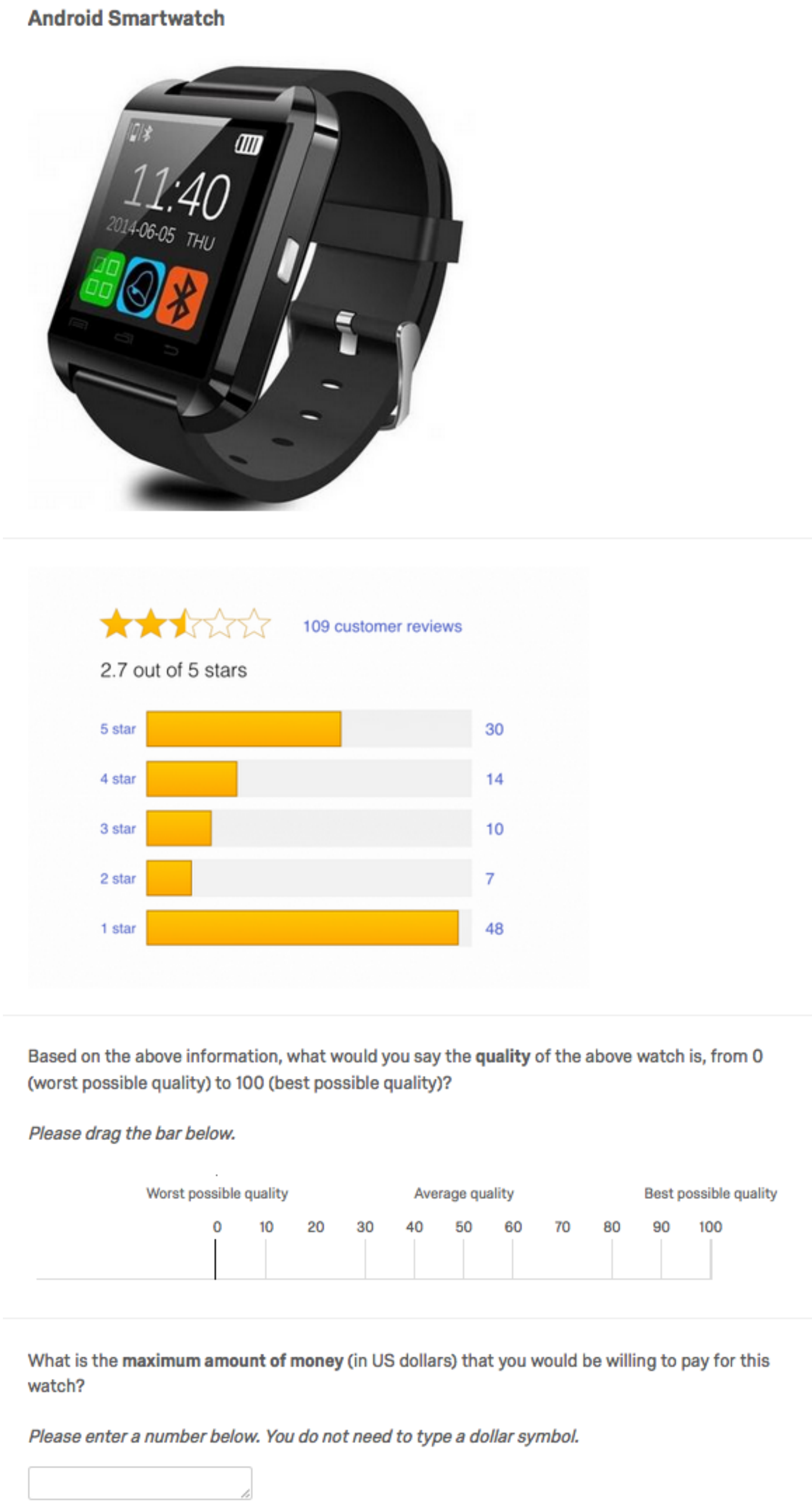


FIGURE 17: Example of the information and questions shown to a participant for a good. Shown is the original review condition for good 1.

An individual will see one of three possible conditions for each good. The control condition (o) shows all the information as described above, with the original review score distribution. Two manipulations are applied to review scores in order to generate two different treatment conditions. The first treatment, referred to as *mean-preserving* (m), removes *some* 1-star and 5-star scores from the distribution, whilst keeping the mean identical. Where more than one mean preserving option is available, the most aggressive transformation is used. If individuals are only looking at the mean score to base quality judgements upon, there should be no difference between reported quality in this treatment and the control group. However, if they weight extreme scores, then the removal of these scores would result in a more positive perception of quality.

The second treatment, referred to as *extreme* (e), removes *all* 1-star and 5-star reviews.[13] This necessarily changes the mean. The direction of movement depends on the shape of the distribution of review scores. For example, goods with a negatively skewed review score distribution will have a lower mean in the extreme treatment, relative to the control. If individuals only consider the mean, then they will have a lower valuation. However, if they weight extreme reviews according to the weighted-mean model, then the removal of these scores should increase their valuation. A summary of the goods used, and the review distributions for each treatment can be found in Table 20.

Each individual is shown all 10 goods, one by one, in a randomised order. For each good, a subject is randomly shown either the control review scores, the mean-preserving treated review scores, or the extreme treated review scores. Therefore, every subject sees a variety of treatments across goods, but only one treatment per good. Approximately a third of subjects sees one condition for each good. We can therefore compare the mean quality and willingness to pay for each good individually, between the three conditions.

The reason this design is preferred over the more simple assignment of one individual per condition for all goods, is to minimise changes of behaviour arising in the extreme condition. If individuals were to see 10 consecutive goods with no 1-star or 5-star reviews, they may have changed their usual strategy to evaluate

13 The removal of lowest and highest scores is similar to the process used by judging in figure skating competitions, used to avoid bias caused by disproportionately extreme opinions.

Table 20: Summary of the 10 goods used in the experiment, with original and treated review score distributions.

				Number of reviews for each star rating (1*, 2*, 3*, 4*, 5*)		
Good	Description	Type of good	Original mean score	Original distribution	Mean-preserving condition	Extreme condition
1	Smartwatch	Search	2.73	(48, 7, 10, 14, 30)	(31, 7, 10, 14, 17)	(0, 7, 10, 14, 0)
2	Smartphone	Search	4.36	(23, 32, 38, 94, 370)	(5, 32, 38, 94, 276)	(0, 32, 38, 94, 0)
3	Headphones	Search	2.50	(9, 6, 2, 2, 5)	(4, 6, 2, 2, 2)	(0, 6, 2, 2, 0)
4	LCD TV	Search	4.50	(2, 2, 3, 6, 39)	(1, 2, 3, 6, 32)	(0, 2, 3, 6, 0)
5	Laptop	Search	3.16	(86, 23, 29, 35, 102)	(17, 23, 29, 35, 21)	(0, 23, 29, 35, 0)
6	3* Hotel	Experience	2.19	(485, 184, 248, 129, 64)	(346, 184, 248, 129, 5)	(0, 184, 248, 129, 0)
7	4* Hotel	Experience	4.06	(18, 30, 59, 230, 206)	(5, 30, 59, 230, 164)	(0, 30, 59, 230, 0)
8	Programming Book	Experience	2.50	(14, 2, 4, 5, 5)	(9, 2, 4, 5, 2)	(0, 2, 4, 5, 0)
9	Parenting Book	Experience	4.33	(3, 3, 3, 5, 34)	(1, 3, 3, 5, 24)	(0, 3, 3, 5, 0)
10	Restaurant	Experience	2.93	(48, 24, 21, 23, 43)	(18, 24, 21, 23, 15)	(0, 24, 21, 23, 0)

Notes: The review score distributions are given in the format (a, b, c, d, e), where a represents the number of 1* reviews, b represents the number of 2* reviews, and so on. The mean-preserving treatment has reduced numbers of 1* and 5* reviews, whilst keeping the mean identical to the original distribution. The extreme treatment removes all 1* and 5* reviews from the original distribution.

products. This would have created an additional confounding factor, reducing the validity and comparability of the results.

In order to assess whether individual differences in personality have mediating effects on good valuation, I ask participants to complete the 20 item mini-IPIP personality inventory (Goldberg et al., 2006), based upon the Big Five factors. This measure is short enough to prevent survey fatigue, whilst still providing a high degree of reliability. The order of personality items was randomised for each subject.

3.3.2 *Predictions*

In order to generate predictions from the models discussed in Section 3.2, we need to determine plausible parameter values. With each model, and for each good, we can predict a rank ordering of the three experimental conditions, based upon how highly the good will be valued in each condition. This can then be compared to the experimental data. In particular, by assessing the accuracy of the weighted-mean model predictions, relative to the valuation implied by the mean review score, we can determine whether the main hypothesis holds.

To generate valuations from the weighted-mean model, I let $\alpha = 1$ and $\beta = 4$. The α and β parameters are chosen so that the the shape of function $f(.)$ is decreasing in the number of reviews with a given score, n_r.

Four different variants of the weighted-mean model are computed. The first (referred to as 'μ_w') takes the threshold $T = N$. This means that the weighting will *always* be different from 1 for extreme reviews. The weighting will be stronger when n_r represents a larger proportion of N.

The second (referred to as 'μ_w threshold') utilises a fixed threshold value for T, depending on the type and source of the good. In a dataset by Julian McAuley (see McAuley, Pandey, and Leskovec, 2015; McAuley, Targett, et al., 2015), the 95th percentile of total number of reviews for 7,824,482 electronics products on Amazon, from May 1996 to July 2014, is 55. For books, the 95th percentile of 22,507,155 goods is 32. Hence, I use $T = 55$ for the first five goods in the experiment, and $T = 32$ for goods 8 and 9. These represent sensible upper thresholds. For TripAdvisor, based on a sample of 1850 hotels, the mean number of reviews for each hotel is 58.86 (Wang, Lu, and Zhai, 2010). There is no information provided from which to infer percentiles. However, if we assume a similar distribution for number of reviews per product on Amazon and per hotel on TripAdvisor, then the 95th percentile would be 3.34 times the mean. Based on this approximation, I use a threshold of $T = 197$ for the three goods from TripAdvisor.

The third (referred to as 'μ_w 1-star') takes $T = N$, as in the first variant. It truncates the shape of $g(.)$, so that the only 1-star reviews are weighted (positively). The weight applied to $r = 5$ is fixed at 1 (i.e. 5-star reviews remain unweighted).

The fourth (referred to as 'μ_w 5-star') is the converse of the previous case. Only 5-star reviews are weighted (negatively). The weight applied to $r = 1$ is fixed at 1 (i.e. 1-star reviews remain unweighted).

Finally, I compute valuations using the range-frequency based model, as discussed in Section 3.2. The weighting parameter w has been found to be close to 0.5 in experimental data (Parducci and Wedell, 1986). Therefore, I take $w = 0.5$, placing an equal weight on the range and the frequency effects.[14]

Scores from each model are computed for each good and condition. An ordinal ranking of valuations by treatment is formed for each good. Model scores are rounded to 1 decimal place. If scores for a good are the same for two treatments at this level of precision, they are taken to be valued equally.

A summary of the ordinal rank predictions for each treatment, using each model, is given in Table 21. With few exceptions, the weighted-mean model gives higher value to a good which has had its scores treated according to the extreme condition. This is no surprise for poorly rated goods, since the treatment will increase the mean review score. However, a highly rated good which is valued most highly in the extreme condition would suggest that individuals are implicitly correcting for extreme review bias. It is these highly rated goods, with negatively skewed review score distributions, that reflect the majority of goods found on online review websites. Hence, these goods (2, 4, 7, and 9) are particularly important in testing the main hypothesis.

14 Parducci and Wedell (1986) find that w can be greater when end points are fixed, and there is more limited scope for spacing. However, without any reliable justification, a value of 0.5 seems more sensible as a prior assumption.

Computed valuations from the four variants of the weighted-mean model are plotted in Figure 18, alongside the raw means. The range-frequency model is excluded, since it produces scores on a [0,1] interval, which would be too subtle to see in comparison. Each symbol in Figure 18 represents a different model. The absolute weighted-mean values are not important, but the relative ordering of valuations between treatments is. This allows us to see how different the orderings generated by the weighted-mean models are to those implied by the raw mean review score.

For example, let us look at panel A of Figure 18. The solid black dots plot the raw mean scores for good 1 (a smartwatch) in the original case, in the mean-preserving treatment case, and in the extreme treatment case. The hollow blue circles plot the weighted mean generated by the weighted-mean model in each of the three conditions. The red triangles plot the weighted mean generated by the weighted-mean model with a fixed threshold, and so on. We can see that the raw mean suggests we should value the smartwatch equally in the original and mean-preserving conditions. However, all but one of the weighted-mean models predict that we would value the smartwatch more in the mean-preserving condition. Of these, the μ_w and μ_w 5-star models value the smartwatch highest in the extreme condition. Table 21 represents this prediction for these two models (in the top row) with the abbreviated notation e>m>o.

TABLE 21: Predictions for the ordering of perceived quality over treatments from each model.

	Ordering of conditions based on predicted values (highest to lowest)					
Good	Mean score	μ_w	μ_w threshold	μ_w 1-star	μ_w 5-star	RF
1	e>o=m	e>m>o	e>o=m	m>o>e	e>m>o	e>o=m
2	o=m>e	o>m>e	o=m>e	m>o>e	o=m>e	o=m>e
3	e>o=m	e>m>o	e>m>o	o=m>e	e>m>o	e>o=m
4	o=m>e	o=m>e	e>o>m	o=m>e	o=m>e	o=m>e
5	o=m>e	e>m>o	e>o=m	m>o>e	e>m>o	m=e>o
6	e>o=m	e>m>o	e>m>o	e>o>m	e>m>o	e>m>o
7	o=m>e	o>m>e	o=m>e	o>m>e	o>m>e	o=m=e
8	e>o=m	e>m>o	e>m>o	e>o=m	e>m>o	e>m>o
9	o=m>e	o=m>e	e>o=m	m>o>e	o=m>e	o=m>e
10	e>o=m	e>m>o	e>m>o	o=m>e	e>m>o	e>o=m

Notes: o = original condition, m = mean-preserving treatment, e = extreme treatment, RF = range-frequency model.
The predictions are shown with the most highly valued treatment on the left. For example, "e>o=m" means the good will be valued most highly when an individual is shown the extreme treatment, followed by both of the other treatments, which are valued equally.

The goods evaluation task requires individuals to react to information in front of them, as well as think about what each review score actually represents (and potentially, how it was generated). Therefore, individual differences in personality may influence valuations. Gill and Prowse (2014) find that more Agreeable and less Neurotic individuals earn more in a p-beauty contest game, since they choose numbers closer to equilibrium and operate on a higher level within the context of a level-k learning model. Their results are consistent with research by Nettle and Liddle (2008) and DeYoung et al. (2010) that suggests higher Agreeableness is linked with a having a better 'theory of mind', which allows these individuals to perform better in situations where they have to predict and interpret the actions of others.

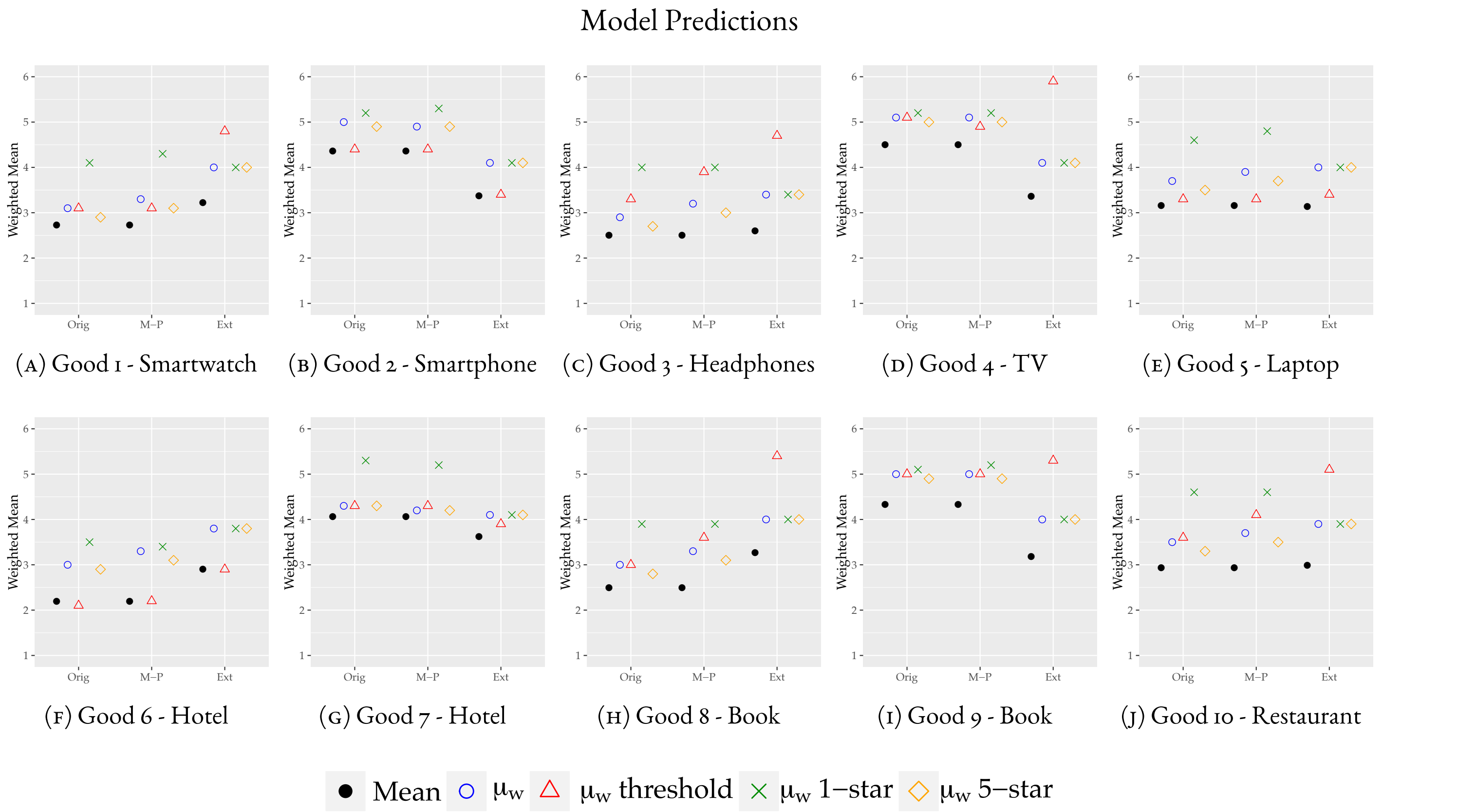


FIGURE 18: Predicted ordering for the valuation of goods in each treatment, according to the raw mean, and the four variants of the weighted-mean model. See text for a detailed description.

Hypothesis 3.4 *Individuals who are more Agreeable will evaluate a good based on a more complex metric than the mean review score (i.e. either the weighted-mean or range-frequency models).*

Because they perceive the possibility of bias, one might expect that individuals with a better theory of mind would be more likely to apply weights to a review score.

More Neurotic individuals respond more strongly to negative affect (McCrae and John, 1992). Gill and Prowse (2014) find that higher Agreeableness and lower Neuroticism load onto one common factor in their strategic setting. In this context, it is not clear whether those lower in Neuroticism would behave similarly to those higher in Agreeableness.

However, more Neurotic individuals would likely pay more attention to negative signals than less Neurotic individuals. Higher Neuroticism has been linked with greater activation of the right insula in the brain (Paulus et al., 2003). This region provides emotional responses that signal aversion, in order to minimise harm. Further studies have confirmed links between higher Neuroticism and greater aversion to risk and loss (e.g. Bibby and Ferguson, 2011). Based on this research, we might expect that highly Neurotic individuals negatively weight 5-star reviews, but do not adjust for deliberately biased 1-star reviews.

Hypothesis 3.5 *Individuals who are highly Neurotic will value goods in closer correspondence to the predictions of 'μ_w 5-star' than those who are low in Neuroticism.*

3.4 DATA AND RESULTS

The experiment was conducted online, using participants from MTurk. 501 observations were collected specifically for this experiment. For each good, treatment conditions were randomised, so that approximately the same number of individuals saw each condition for each good.[15] Mean time for overall completion was 12.76 minutes. Participants were paid $1.50 for successful completion of the experiment. No task dependent incentive was given. The mean age of

15 They were not exactly equal due to the procedure used.

participants was 37.7, and 51.7% were male. 69.5% had at least an undergraduate degree. Almost all participants were U.S. nationals.

3.4.1 *Results for Quality Evaluation*

Table 22 shows the observed ordering of treatments for each good. The ordering of experimental conditions for a particular good is based on the mean reported quality for that good. In each of the subsequent columns, a 'Yes' is shown whenever the actual ordering in column 2 matches the corresponding model prediction from Table 21. These model predictions (apart from those for the RF model) can be seen visually in Figure 18. In the analysis to follow, each of the predicted orderings implied by the weighted-mean model can be contrasted with the actual observed ordering by comparing each graph from Figure 18 with the corresponding graph in Figure 19.

We see that no model prediction matches the actual ordering for every good. The weighted-mean model with fixed threshold (μ_w threshold) predicts correctly most often. It is more successful than both the raw mean (3rd column), and the range-frequency approach (rightmost column), since the goods predicted correctly by μ_w threshold form the union of the goods predicted correctly by the mean and RF. Three of the four goods that μ_w threshold correctly predicts the ordering for are experience goods, suggesting that more weighting may be occurring for these goods than search goods.

However, even the μ_w threshold model is only correct for 4 of 10 goods (5th column in Table 22), so none of the models tested are excellent predictors of behaviour. Furthermore, only one of these 4 correct predictions (good 7) corresponds to a negatively skewed review distribution. Since most product reviews are distributed in this fashion, the weighted-mean model is, overall, not very successful in predicting behaviour.

Despite this, the mean quality for the mean-preserving treatment is significantly different to the mean quality in the control condition (i.e. where the mean quality for the mean-preserving condition lies outside the 95% confidence interval of the original condition) for 7 of 10 goods.[16] Therefore, individuals are

16 This can be seen in the 2nd column of Table 22, for all goods where the 'o' and 'm' conditions are not denoted as being equal to each other.

not basing their evaluation purely on the raw mean review score. In addition, 3 out of 5 experience goods have higher quality ratings for the mean-preserving treatment when compared to the control, though this is only for the goods with a poor overall review score and the good with a U-shaped review distribution.

The μ_w 1-star and μ_w 5-star versions were also unsuccessful as predictors of quality evaluation. The μ_w 1-star model's predicted ordering did not match the actual ordering for any of the goods. The μ_w 5-star model's predicted ordering only matched the actual ordering for two goods. Hence, two things appear to be true. First, if individuals are taking extreme review bias into account, they are doing so at both ends of the review score scale. Second, individuals seem to be more likely to (negatively) weight 5-star reviews than (positively) weight 1-star reviews. This may reflect the J-shaped review score distribution observed in practice for most goods online. The fact that many more 5-star reviews are placed than 1-star means that individuals are more likely to have devised an implicit mechanism to deal with this abundance.

TABLE 22: Summary of results for mean quality, and accuracy of model predictions. The predicted order for each model can be found in Table 21.

	Result	Did model prediction match result?					
Good	Actual order (quality)	Mean	μ_w	μ_w threshold	μ_w 1-star	μ_w 5-star	RF
1	e>o=m	Yes	No	Yes	No	No	Yes
2	o>m>e	No	Yes	No	No	No	No
3	o=m=e	No	No	No	No	No	No
4	o>m>e	No	No	No	No	No	No
5	m>o=e	No	No	No	No	No	No
6	e>m>o	No	Yes	Yes	No	Yes	Yes
7	o=m>e	Yes	No	Yes	No	No	No
8	e>m>o	No	Yes	Yes	No	Yes	Yes
9	o>m>e	No	No	No	No	No	No
10	m=e>o	No	No	No	No	No	No

Key:
= - the means of two treatments are not significantly different at the 5% level.
o - original review scores; m - mean preserving treatment; e - extreme treatment, RF = range-frequency model.
The results are shown with the highest valued treatment on the left. For example, "e>o=m" means the good had significantly higher mean quality in the extreme treatment than the other two treatments (which are not valued significantly differently).

Mean quality for each good and treatment, along with 95% confidence intervals, is plotted in Figure 19. If we compare these plots to the model predictions in Figure 18, the weighted-mean models appear to do worse for goods which have mixed reviews, and for goods that have a low overall number of reviews (goods 3-5 and 8-10). No model does well at predicting the results for the two goods with mixed reviews (5 and 10), though if we compare panels E and J in Figure 18 with the corresponding panels in Figure 19, μ_w 1-star appears qualitatively closest.

Although 7 of 10 goods have statistically significant differences in quality for the control and mean-preserving conditions, the relative difference in quality rating in these cases is small. This is unsurprising, given that for many goods, mean-preserving transformations did not substantially alter the overall shape of the review distribution. In general, there is little difference in quality ordering for search goods and experience goods with similar original review score distributions. This is in contrast to the finding by Mudambi and Schuff (2010) that reviews with moderate scores are rated by consumers as being more helpful in the case of experience goods. A more detailed discussion of the results for specific goods follows.

The largest difference between the mean-preserving and original treatment can be seen for goods 5 (the laptop) and 10 (the restaurant). These goods both had a U-shaped review score distribution originally. In the mean-preserving treatment, reduction of 1-star and 5-star reviews turned these distributions into hump shapes. The fact that quality was perceived as being higher when review scores were distributed with a hump shape appears to support the case for weightings on extreme scores. However, the fact that the extreme treatment does not share the same preference suggests that the mean review score is still important.[17]

Goods 1, 2, 6, and 7 were selected to represent products with low and high mean review scores, given a relatively high N. For goods 1 and 6, which were rated poorly, reported quality was significantly greater in the extreme treatment (panels A and F in Figure 19). This is as we would expect, given that the mean score is also increased as a result of this treatment. However, this was not the case for goods 2 and 7, which were highly rated. The extreme treatment in these cases was ranked significantly lower than the original (panels B and G in Figure

17 The extreme treatment was preferred to the original for good 10 (where the mean score was increased), but was not preferred for good 5 (where the mean score was decreased).

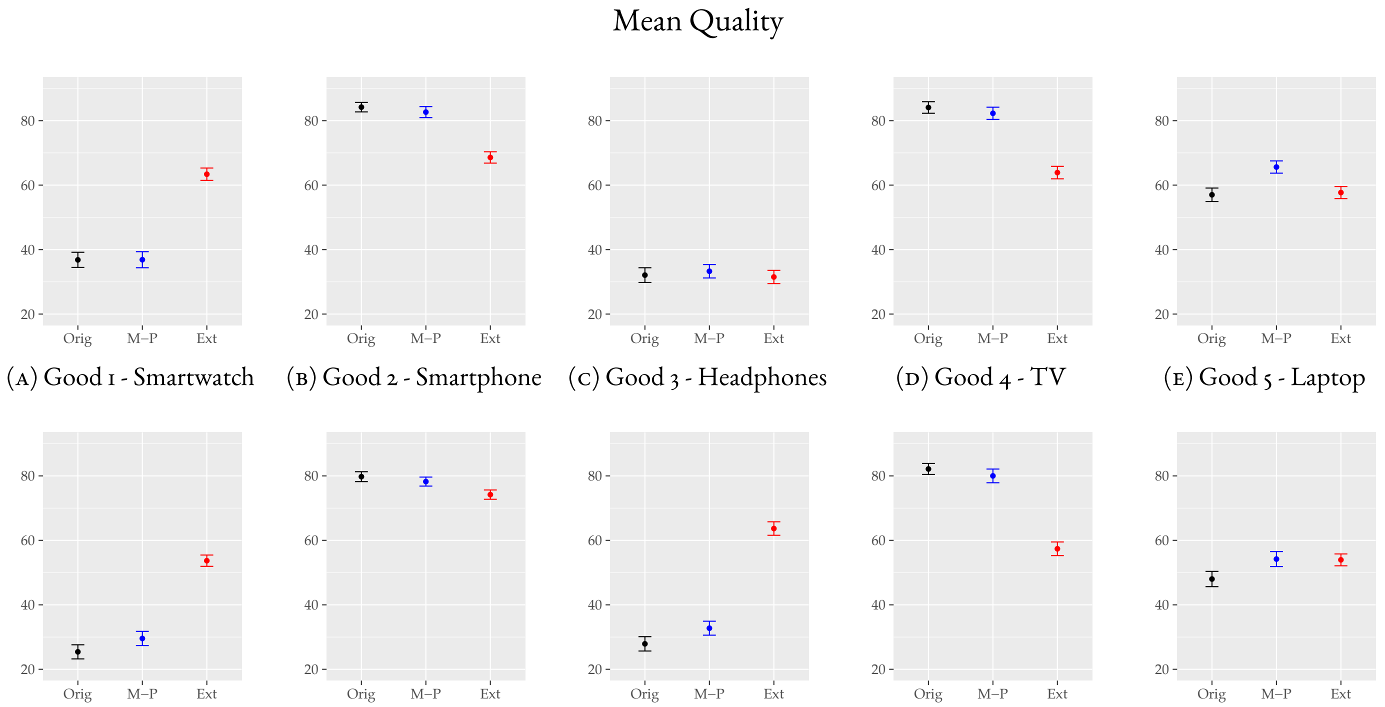
19), likely due to the fact that the extreme transformation resulted in a reduction of the mean score. For these goods, there is minimal reason to believe that individuals are rating search goods and experience goods differently.

For goods 2 and 7, the mean-preserving treatment was ranked lower than the original (though only significantly for good 2). For good 6, the mean-preserving treatment was ranked significantly higher than the original (column 2 in Table 22). These results imply that, when mean scores are equal, individuals seem to be focusing on the number of 5-star reviews, possibly in relation to the number of 1-star reviews. Contrary to the weighted-mean models, this result suggests that individuals may be placing *positive* emphasis (i.e. applying a weight greater than 1) on 5-star review scores, relative to moderate scores.

Goods 3, 4, 8, and 9 were selected to represent products with low and high mean review scores, given a low N. Apart from good 3, we see that the overall number of reviews does not have an effect on the quality ordering of goods between treatments. The highly rated goods with low N (4 and 9) have the same ordering as the corresponding highly rated goods with high N (2 and 7): the good in the control condition is rated as being of higher quality than in the mean-preserving treatment, and the good is valued lowest in the extreme treatment (panels D and I in Figure 19). Similarly, the poorly rated, low N good 8 shares the same ordering as the poorly rated, high N good 6.

The only anomaly among these goods is good 3 (headphones), where there was no significant difference between quality ratings in any of the three conditions (panel C in Figure 19). Though the overall number of reviews for the original product was low (24), this was also the case for the corresponding experience good, good 8 (book), which originally had 30 reviews (panel H in Figure 19). One possible reason is that the number of reviews with the middle three star ratings (i.e. 2, 3, or 4 stars) for good 3 was positively skewed, whereas it was negatively skewed for good 8. Given that the range frequency model pays considerable attention to the skewness of a distribution, it is surprising that it did not predict the ordering for good 3 accurately, given that it correctly predicted the ordering for good 8 (see the final column of Table 22).

In sum, there is some (but limited) support for Hypothesis 3.1, particularly for experience goods. The weighted-mean model with a fixed frequency threshold outperforms the predictions generated by the raw mean score and range-



(A) Good 1 - Smartwatch
(B) Good 2 - Smartphone
(C) Good 3 - Headphones
(D) Good 4 - TV
(E) Good 5 - Laptop
(F) Good 6 - Hotel
(G) Good 7 - Hotel
(H) Good 8 - Book
(I) Good 9 - Book
(J) Good 10 - Restaurant

FIGURE 19: Mean reported quality for each good and treatment (original, mean-preserving, extreme), with 95% confidence intervals. The 5 goods on the top row are search goods, the 5 on the bottom row are experience goods.

frequency approach. There is no support for Hypotheses 3.2 and 3.3. In other words, if individuals are implicitly weighting extreme reviews, overall they are doing so for both extremely positive and extremely negative ones.

3.4.2 *Results for Personality*

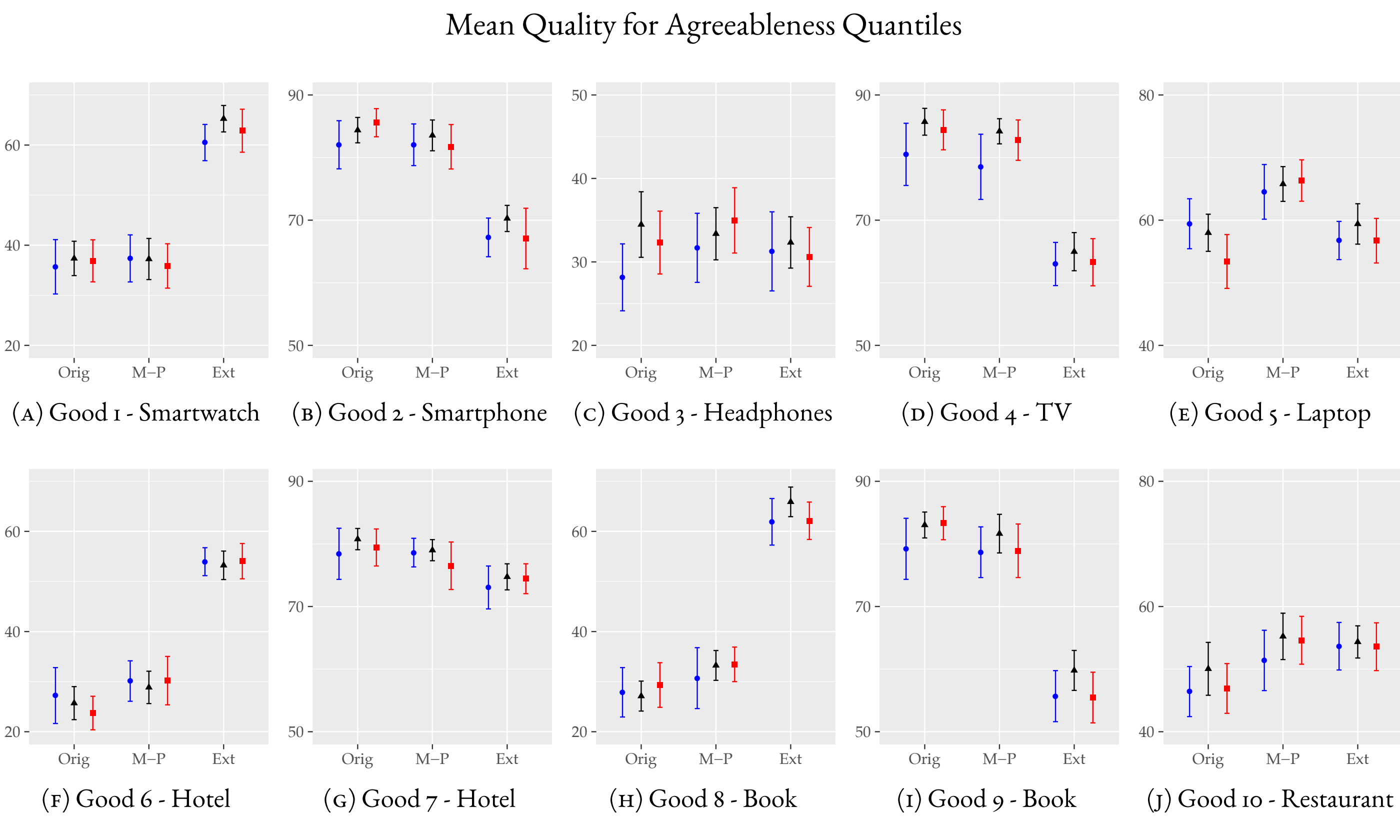
We have two hypotheses to test for personality effects in this context. Hypothesis 3.4 states that more Agreeable individuals should follow the predictions of either the weighted-mean or range-frequency models more closely than the prediction of the raw mean score. Hypothesis 3.5 states that more Neurotic individuals will specifically follow the predictions of μ_w 5-star more closely than the raw mean score.

In order to test these hypotheses, I split the sample into quantiles by personality trait. Individuals are classified into low, medium, and high levels of a factor in the following way. Those with a trait score in the bottom quartile of the sample (25th percentile or lower) are classified as having a low score. Those with a trait score in the second or third quartiles (between the 75th and 25th percentiles) are classified as having a medium score. Those with a trait score in the top quartile (75th percentile or higher) are classified as having a high score.

Figure 20 plots the mean reported quality for individuals with low, medium, and high levels of Agreeableness. At first glance, there do no appear to be large differences in quality evaluation for the majority of goods, given different levels of Agreeableness.

By listing the observed treatment ordering for those with high and low levels of Agreeableness, we can test Hypothesis 3.4 more explicitly. Table 23 lists the actual ordering for quality evaluation, separating for high and low Agreeableness. Neither the best performing weighted-mean model (μ_w threshold), nor the RF model predicts ordering for highly Agreeable individuals well. The RF model is successful more often than the raw mean or weighted-mean, but this still only amounts to prediction success in 4 of the 10 goods.

However, even though these more sophisticated models are not good predictors, it is still true that individuals low in Agreeableness are much closer to the predictions given by the raw mean than individuals high in Agreeableness. This suggests that there may be a theory of mind effect present: individuals high in



(A) Good 1 - Smartwatch (B) Good 2 - Smartphone (C) Good 3 - Headphones (D) Good 4 - TV (E) Good 5 - Laptop

(F) Good 6 - Hotel (G) Good 7 - Hotel (H) Good 8 - Book (I) Good 9 - Book (J) Good 10 - Restaurant

FIGURE 20: Mean reported quality for each good and treatment (original, mean-preserving, extreme), separated by Agreeableness. Blue circles are quality evaluations for individuals in the bottom quartile for Agreeableness, black triangles are for individuals in the middle two quartiles, red squares are for individuals in the top quartile.

Agreeableness appear to be doing something more sophisticated than looking at mean score alone. It so happens that neither of the models tested in this study does a good job in capturing their actual process.

There is evidence that for search goods with a low number of overall reviews (goods 3 and 4), those who are low in Agreeableness perceive the good to be of lower quality in the original and mean-preserving conditions than those with medium or high Agreeableness. This appears to be consistent with Evans and Revelle (2008), who find that Agreeableness leads to trust in situations where there is greater uncertainty. However, this relationship does not hold for low frequency experience goods.

Table 23: Highly Agreeable individuals' quality evaluation is not well captured by the weighted-mean model, or the range-frequency model.

	Results		Did model prediction match result?					
	Actual order (quality)		Raw mean		μ_w threshold		RF	
Good	Low A	High A	Low A	High A	Low A	High A	Low A	High A
1	e>o=m	e>o=m	Yes	Yes	Yes	Yes	Yes	Yes
2	o=m>e	o>m>e	Yes	No	Yes	No	Yes	No
3	o=m=e	o=m=e	No	No	No	No	No	No
4	o=m>e	o=m>e	Yes	Yes	No	No	Yes	Yes
5	m>o=e	m>o=e	No	No	No	No	No	No
6	e>o=m	e>m>o	Yes	No	No	Yes	No	Yes
7	o=m>e	o=m=e	Yes	No	Yes	No	No	Yes
8	e>o=m	e>o=m	Yes	Yes	No	No	No	No
9	o=m>e	o>m>e	Yes	No	No	No	Yes	No
10	m=e>o	m=e>o	No	No	No	No	No	No

Key:
= - the means of two treatments are not significantly different at the 5% level.
o - original review scores; m - mean preserving treatment; e - extreme treatment.
The results are shown with the highest valued treatment on the left. For example, "e>o=m" means the good had significantly higher mean quality in the extreme treatment than the other two treatments (which are not valued significantly differently).

Figure 21 plots mean quality graphs for individuals with low, medium, and high Neuroticism. As with the graphs plotted for Agreeableness, there are few

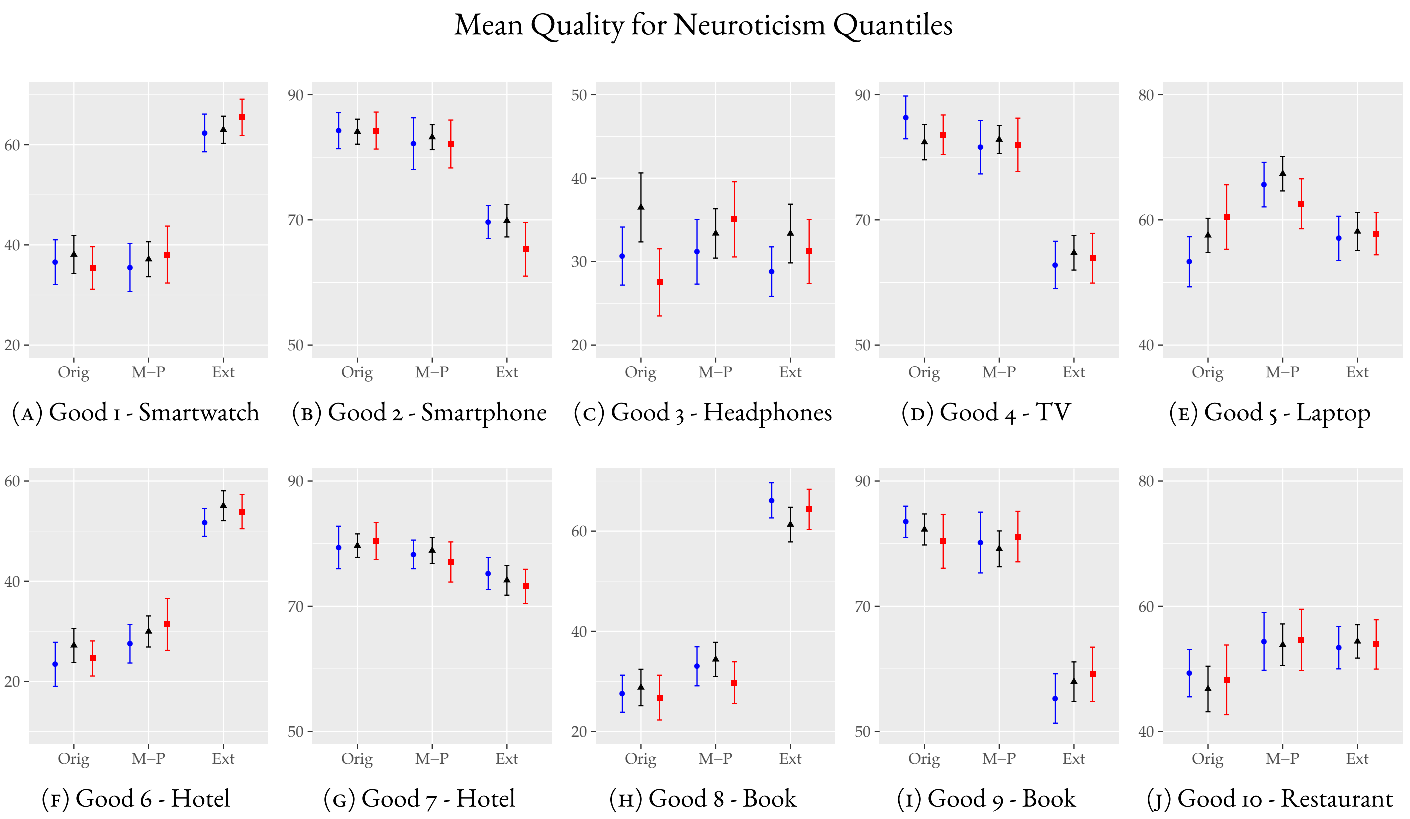


FIGURE 21: Mean reported quality for each good and treatment (original, mean-preserving, extreme), separated by Neuroticism. Blue circles are quality evaluations for individuals in the bottom quartile for Neuroticism, black triangles are for individuals in the middle two quartiles, red squares are for individuals in the top quartile.

goods that exhibit large differences in quality evaluation between different levels of Neuroticism.

For good 3 (headphones), individuals with a low level of Neuroticism show no significant difference in quality ratings for each of the three conditions. Those with a medium level of Neuroticism show a preference for the good under the original review distribution condition over the mean-preserving (significant at the 5% level) and the extreme (not significant at the 5% level) treatments. Those with a high level of Neuroticism, on the other hand, show a significant preference for the mean-preserving treatment over the others. In particular, those with high Neuroticism also exhibit a reduced preference for the good in the original distribution.

For good 5 (laptop), those with relatively low or medium Neuroticism value the good most highly in the mean-preserving treatment. Those high in Neuroticism do not show a significant difference between the original and mean-preserving conditions.

I summarise the differences in ordering between individuals with low and high levels of Neuroticism in Table 24. As discussed previously, the differences in quality evaluation for goods 3 and 5 are the most prominent. Nevertheless, 7 of the 10 goods exhibit different orderings at the 5% level when comparing individuals with low Neuroticism to those with high Neuroticism (all but goods 1, 2, and 10).

Hypothesis 3.5 states that those with high Neuroticism will correspond more closely to the predictions of the μ_w 5-star model than those with low Neuroticism. Table 24 confirms this. First, we see that the μ_w 5-star model predicts quality ordering correctly for only 2 of 10 goods when looking at individuals with low Neuroticism; whilst it predicts ordering correctly for 5 of 10 goods when looking at individuals with high Neuroticism. Whilst a 50% success rate is still no better than a coin flip on average, it is still clear that the model performs better on highly Neurotic individuals.

Second, for highly Neurotic individuals, the μ_w 5-star model predicts ordering correctly more often than the μ_w threshold model (which is the best performing model overall). This provides support to the argument that individuals high in Neuroticism are more likely not to adjust for deliberately biased 1-star reviews.

Table 24: Highly Neurotic individuals are better captured by the μ_w 5-star model than those low in Neuroticism.

	Results		Did model prediction match result?			
	Actual order (quality)		μ_w 5-star		μ_w threshold	
Good	Low N	High N	Low N	High N	Low N	High N
1	e>o=m	e>o=m	No	No	Yes	Yes
2	o=m>e	o=m>e	Yes	Yes	Yes	Yes
3	o=m=e	m>o=e	No	No	No	No
4	o>m>e	o=m>e	No	Yes	No	No
5	m>o=e	o=m=e	No	No	No	No
6	e>o=m	e>m>o	No	Yes	No	Yes
7	o=m>e	o>m>e	No	Yes	Yes	No
8	e>m>o	e>o=m	Yes	No	Yes	No
9	o>m>e	o=m>e	No	Yes	No	No
10	m=e>o	m=e>o	No	No	No	No

Key:
= - the means of two treatments are not significantly different at the 5% level.
o - original review scores; m - mean preserving treatment; e - extreme treatment.
The results are shown with the highest valued treatment on the left. For example, "e>o=m" means the good had significantly higher mean quality in the extreme treatment than the other two treatments (which are not valued significantly differently).

In sum, differences in mean quality across treatments for different levels of Agreeableness and Neuroticism are neither large nor consistent enough to conclude that these personality factors have a significant influence on the evaluation of a good. However, there are differences in ordering between individuals with low and high levels of a trait that provide some support for Hypotheses 3.4 and 3.5.

Whilst it appears that highly Agreeable people are using a more complex metric than the mean review score, none of the models tested in this study are able to adequately explain their method. Highly Neurotic individuals appear more susceptible to falsified 1-star reviews, since they do not appear to weight these in the same way as they weight scores for 5-star reviews.

It is apparent from the graphs in Figures 20 and 21 that there is a larger level of disparity in quality evaluation for different personality levels when the overall number of reviews is lower, or when goods have a U-shaped review score distribution with a mean score close to the midpoint of the scale. Furthermore, any differences appear to be greater for search goods than experience goods.

3.4.3 *Results for Willingness to Pay*

As one might expect, the data on WTP for each good was considerably noisy, relative to the data on quality. At the end of the rating task, individuals were provided with a free text entry box in which to explain the reasoning behind their decisions.[18] From some of the comments, it is apparent that individuals (at least, in part) chose their WTP based upon their preferences towards a particular good, wealth constraints, and other inferences about the good which were not directly based upon the information presented in front of them. This was not the case for quality, where individuals refer more to the review data provided. For example, one respondent explains:

> "I don't wear watches so, I didn't want to spend too much on watches. I'm not fond of touch screen smart phones either, so even though the reviews might be quite favorable, I know I might be offering way less than what it's worth. Same thing with the Windows laptop, 14 inch is my minimum requirement, so I might've offered way less than it's worth. I roughly estimated 4 stars equals about 80 percent, although i might be off here and there on certain goods (and I take into consideration if there were some really poor rating too)."

However, Figure A1 shows that there is still a general correspondence to the ordering and patterns observed in the quality data. Therefore, despite the lack of reliability, the conclusions that can be drawn from quality and WTP are qualitatively similar. For WTP, the range-frequency model predicts ordering correctly for 4 of the 10 goods, which is the same as the standard μ_w model. Nevertheless,

18 Completion of this was optional, though everyone in the sample provided a response.

the raw mean review score is the best overall predictor of rank ordering (see Table A1). This supports the case against Hypothesis 3.1: individuals do not appear to be weighting extreme reviews.

3.5 CONCLUSION

Due to the increase in e-commerce, online reviews have become increasingly important to the valuation of goods. This has opened up the possibility of exploitation through the generation of fake reviews. This study has hypothesised that to compensate for this possibility, consumers may be applying weights to reviews which are scored at the extreme ends of a scale.

Overall, I find that quality evaluations for goods correspond more closely to the weighted-mean model developed in this chapter than the raw mean review score, or a model based on range-frequency theory. Nevertheless, the predictive power of the weighted-mean model devised in this chapter is still relatively poor. Treatment ordering based on quality is predicted correctly (at best) for only 4 of 10 goods, and treatment ordering based on WTP was also predicted correctly (at best) for only 4 of 10 goods. If all individuals were consistently adjusting for extreme reviews by overweighting 1-star reviews and negatively weighting 5-star reviews, one would have expected predicted orderings to be correct for most goods. Hence, individuals appear not to be fully adjusting review scores to take into account the possibility of deliberately biased reviews.

In addition to this, the study finds that personality differences have some influence on how individuals evaluate the quality of goods. Personality differences have more influence on the evaluation of good quality when a product has fewer overall reviews, or when a product has mixed review scores.

Whilst individuals low in Agreeableness appear to evaluate quality predominantly using the mean review score, those high in Agreeableness do not. This supports research on highly Agreeable individuals having better theory of mind, since they appear to be doing something more sophisticated than merely looking at the mean score. However, none of the models tested in this study can explain the evaluation behaviour of highly Agreeable individuals satisfactorily.

Highly Neurotic individuals are better predicted by an asymmetric version of the model that only applies a weight to 5-star reviews. This suggests that those high in Neuroticism are more susceptible to reduce their valuation of a good following exposure to reviews at the bottom end of the scoring range. Therefore, they may be more prone to exploitation by firms who seek to harm the reputation of competing products by leaving false bad reviews.

Personality differences, however, are only one aspect of individual heterogeneity. Since the results are based on averages across individuals, it is also possible that some individuals are weighting review scores, but that others are not. Alternatively, individuals may be weighting review scores, but with differing parameter values. Accounting for explicit heterogeneity in the model is left as an extension for future work.

There is some evidence (as discussed in Section 3.3, in the comparison between quality evaluation in the mean-preserving condition and the control condition) that suggests individuals may be placing more emphasis on extreme reviews than middling reviews in their evaluation of a good's quality. One possible interpretation of these results is that consumers see a product as being either 'good' or 'bad'. This would imply that middling review scores have relatively little value in signalling product quality.

Due to the potential utility loss from overvaluing a low quality good, or undervaluing a high quality good, it seems unlikely to be rational to form incorrect valuations. Therefore, if it is true that individuals are aware of potential review bias but are not weighting extreme reviews, then it may point to the presence of a cognitive bias.

Another possibility, whilst less likely, is that some individuals are simply unaware of the possibility of falsified reviews. This may be true for individuals with low Agreeableness, for whom the mean review score is a good predictor of quality evaluation. It may be possible to distinguish between these two explanations by performing an additional experiment in which another treatment primes 50% of the sample with information and news articles about fake reviews. If individuals in this primed group were to evaluate goods in a way that better matched the predictions of the weighted-mean model, this would provide support for the idea that awareness of the possibility of deliberately biased reviews is lacking.

In both of these cases, consumer surplus is likely to be reduced. The explicit measurement of this welfare loss would depend on the cost attached to goods that consumers bought because of unfairly inflated reviews, but then returned or were unhappy with, combined with the cost of goods forgone that consumers never bought due to unfairly deflated reviews. This would be difficult to calculate, given that most of these figures are unobservable. It is left as an open avenue for future research.

4

BEHAVIOURAL FOUNDATIONS OF INDUSTRIAL COMPOSITION: AN EXPLORATORY ANALYSIS

Traditionally, the industrial sectors in which a country specialises have been thought of as being determined largely by resource endowments. However, characteristics of workers within an economy are likely to influence its composition. Using data from the UK and Germany, I show that Big Five personality factors are predictive of future industry change - by as much as 16 percentage points for Agreeableness. Differences in traits may help to explain differences in industrial composition between countries, but further work needs to be done to verify this. This work highlights the relevance of personality data to the analysis of traditional economic issues.

4.1 INTRODUCTION

The industrial output of a country is largely determined by its endowment of resources, and its relation to global demand. However, is it also possible that the individual characteristics of workers can help to determine industry composition? I hypothesise that personality trait differences in labour across countries may be an important, yet previously unconsidered, component of industrial differences. Though there is a well-established literature on comparative advantage (e.g. Dornbusch, Fischer, and Samuelson, 1977; Leamer, 1995; Ricardo, 1821; Roy, 1951), and a burgeoning one on the economics of personality (e.g. Almlund et al., 2011; Borghans et al., 2008; Boyce and Wood, 2011; Boyce, Wood, and Brown, 2010; Nyhus and Pons, 2005, 2012; Uysal and Pohlmeier, 2011), this appears to be the first attempt to combine the two ideas.

The aim of this chapter is to provide an initial exploration of the relationship between personality and industrial composition. The results suggest that worker personality is related to industrial composition through schooling choice.

There is also an early indication that differences in worker personality, and differences in industrial composition between countries may be related.

The Big Five is the most widely studied measure of personality.[1] Nevertheless, relatively little work has analysed cross-country or cross-cultural Big Five data.[2] Costa, Terracciano, and McCrae (2001) look at cross-country data to show gender differences in personality are more pronounced in the Western world. McCrae and Terracciano (2005) compare the Big Five across countries and find internal validity across all countries tested, apart from Botswana. Big Five items from the NEO-PI-R measure (Costa and McCrae, 1992) identify the same traits across countries; even where the language and culture differs from the US. The Big Five, therefore, appear to measure something inherently human, as opposed to merely an artefact of Western culture.

Terracciano et al. (2005) find that national character stereotypes (such as the Germans being highly Conscientious and Canadians being highly Agreeable) do not represent the individual personality traits of the people in those countries. Instead, they appear to be separate social constructs that represent a culture but that do not appear to determine the type of people within it. Na et al. (2010) draw a similar conclusion. This is an important result, because it lends strength to the argument that personality traits are robust to one's environment. In determining an empirical strategy, this suggests that the industrial culture of a country is not likely to impact the personalities of the workers within it.

Comparative advantage and specialisation has been explained most prominently by Heckscher and Ohlin (see Leamer, 1995) on a cross-country level, and by Roy (1951) on an individual level. The Heckscher-Ohlin approach operates under the premise that whilst factors of production are relatively immobile between countries, goods can be freely traded.[3] Therefore, differences in production advantage between countries would lead them to produce more of the good that they were most efficient at producing (by allocating factors accordingly). Trading these goods leads to greater combined output than a single country trying to produce all goods themselves. Markets reach equilibrium through the

1 A more thorough review of Big Five research in economics can be found in Chapter 1.

2 This is, at least in part, explained by the limited availability of personality data from representative surveys in general.

3 Many would challenge this notion today.

relative prices of final goods translating into wage and rent changes for labour and capital respectively.

The Heckscher-Ohlin approach is more macro-oriented, and does not address the important role of individual characteristics. The original theoretical ideas on self-selection in occupational choice by Roy (1951) have been developed by others to allow for more rigour and clarity in hypothesis testing (notably Almlund et al., 2011; Borjas, 1987; Heckman and Sedlacek, 1985; R. J. Willis and Rosen, 1979). In particular, Heckman and Sedlacek (1985) expand the model to include utility (among other things) rather than computing choices based on pure income maximisation. This family of models takes skills and ability into account, which in combination with wage, determine the choice threshold for an individual. However, I propose that an individual will gain utility directly from a good personality match.

Personality has already been shown to have links with occupational choice. Woods and Hampson (2010) find that Openness and Conscientiousness in childhood (as well as gender) are the strongest personality-based predictors of adult occupational choice. However, since the cohort were born in the 1950s and 1960s, Big Five structures had to be derived from the data they obtained. Hence these results may be subject to measurement error. To the best of my knowledge, no work has been done to extend these findings in order to discover the role of personality in shaping overall industrial composition, and in influencing cross-country compositional differences.

The rest of this chapter is organised in the following way. Section 4.2 briefly outlines theoretical ideas and uses these to serve as the basis for hypothesis formation. Section 4.3 tests these hypotheses using empirical evidence from the UK and Germany. Section 4.4 provides a discussion on robustness of results. Section 4.5 concludes. Additional tables can be found in the Appendix.

4.2 HYPOTHESES

Classic economic literature, following Roy (1951), has explained occupational choice at an individual level using (expected) income. Output prices and wages adjust to meet demand and supply requirements. As a result, the problem reduces to one of income maximisation. However, income maximisation is not the only motive for occupational choice. Many individuals have an intrinsic inclination or desire above and beyond their wages to do their job.[4]

Figure 22 proposes an outline of the high-level channels that determine career choice. I do not claim to exhaust all of the intricate interdependencies here. Values and individual preferences are shaped by many external and internal factors, not all of which can be expressed in this diagram. I do claim, however, that in broad terms, an individual's career choice is influenced directly by *income*, *personality* and *values*.

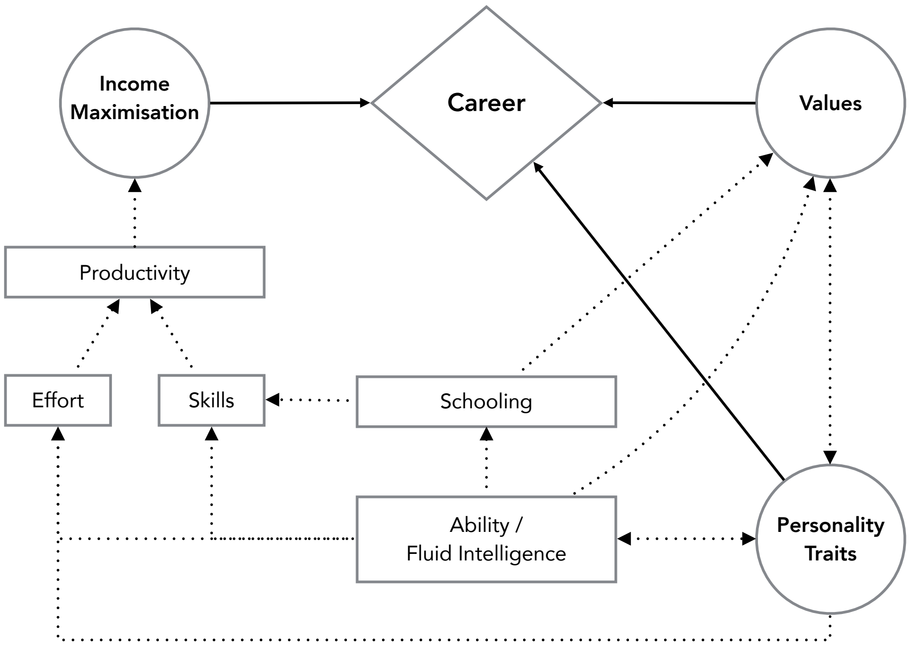


FIGURE 22: A diagram showing proposed high-level determinants of an individual's career choice.

I conjecture that indirect links and other external influences eventually affect career choice through one of these three channels. Once values have been

4 Empirical evidence of this phenomenon can be found in previous research on job satisfaction, e.g. A. E. Clark (1996). Morgan, Dill, and Kalleberg (2013) show the significance of intrinsic motivation in explaining job satisfaction among frontline healthcare workers, such as nurses. Job attrition, however, is still best explained by extrinsic rewards (or lack thereof).

formed, they have a direct influence on career selection: "I don't want to work for a company that sells tobacco products because I don't approve of their contribution to society". Personality traits also have a similarly direct influence. For example, "I want to work in sales because I know I'm quite chatty, I love interacting with people, and like the fast pace and high pressure," might be a conclusion drawn by a highly Extraverted person. Hence, I form a first hypothesis about personality traits and job selection:

Hypothesis 4.1 *Big Five personality traits are related to the occupation chosen by an individual.*

This hypothesis is not new, and the literature (e.g. Woods and Hampson, 2010) does provide some support. However, we should confirm this for the remainder of the argument to be consistent.

Given that each industry has a distinctly different profile of individual occupations contained within it, career choice at the individual level should shape the industrial composition of a country. Hence, I form a second hypothesis:

Hypothesis 4.2 *Big Five personality traits have a significant impact on the industry an individual works in, and therefore on the industrial composition of a country.*

In selecting an individual job, it may be that the relative importance of personality and values are low if one is simply working to survive. This can prove problematic because not everyone will be in their optimal career at any given time. However, at least for the developed world, an assumption I make is that, over time, people will gravitate closer to their 'ideal' occupational area. Empirically, this may suggest a need to look at workers above some age threshold in order to observe a tangible relationship between an individual's personality and an industry. It also implies that we can test this idea by determining whether personality can predict an individual worker's future movement across industries.

Finally, I consider the issue of industry differences across countries. If personality traits affect the industrial composition of an economy, then it follows that countries will end up with different industrial compositions given a different starting distribution of personalities. Whilst there are other factors influencing a country's specialisation, such as natural resource levels and climate, we would expect personality traits to be independent of other country fixed effects.[5] Hence, the final hypothesis is:

Hypothesis 4.3 *Differences between personality traits across countries are associated with differences in their industrial compositions.*

4.3 DATA AND EVIDENCE

This section represents a first, exploratory attempt at testing these admittedly ambitious hypotheses. I use a combination of representative survey data from the UK and Germany. At the time of data collection, these countries were both part of the European Union. They are similar enough in their development level that we are unlikely to have to worry about large structural differences when explaining the results.

For the UK, I use the 2005 wave of the British Household Panel Survey (BHPS), from University of Essex Institute for Social and Economic Research (2010). For Germany, I use both the 2005 and 2009 waves of the SOEP (Wagner, Frick, and Schupp, 2007). These waves are selected as they contain a shorter version of the Neuroticism-Extraversion-Openness Five-Factor-Inventory (NEO-FFI), originally by Costa and McCrae (1992). The short form used in both surveys was developed by Benet-Martínez and John (1998).

The NEO-FFI is the most widely used and highly regarded measure of the Big Five personality factors. The reduced form represents a compromise between accurate measurement of Big Five traits and the desire for a shorter list of items to reduce survey fatigue. Each individual rates the strength of 15 statements on a 7-point scale (the statements are listed in Table A2). Three statements correspond to each factor. The statements use simplified vocabulary, relative to the NEO-FFI, in order to assist with comprehension after translation to multiple languages. I compute the mean of these three scores to obtain one score per trait for each individual.

5 These factors can be subsumed into a country dummy variable.

After refining the dataset to capture only the working population, the number of individual observations for SOEP 2005 was 1,556.[6] This is a slightly smaller sample size than would be desirable. Hence, since the industry breakdown for SOEP 2005 and 2009 is not vastly different, I pool data from the two waves together. Doing this results in a total of 12,637 observations for Germany. Refining the 2005 BHPS results in 7,017 observations for the UK.

4.3.1 *The Big Five and Occupational Choice*

First, I test Hypothesis 4.1 to see whether Big Five factors influence career choice at the occupational level. For the UK, 317 different occupations are represented in the data. For Germany, there are 295. Of these, I eliminate all that have less than 20 occurrences in order for probit estimation to be effective.[7] This leaves 83 occupations for the UK and 143 for Germany.

In addition to standard demographic control variables, I include a series of variables that capture at least part of the *values* channel on job choice. Heckman, Stixrud, and Urzua (2006) explain that personality traits have an influence on schooling choice. I echo this sentiment in Figure 22. This being the case, including schooling choices as a regressor would likely absorb some of the variation due to personality. Therefore, I do not include education variables in the regressions.

6 I remove those who are unemployed, and also those for which there is missing data on personality, educational status, or industry of employment.

7 Failing to do this means that probit regressions for occupations where a low number of workers are reported in the data will have regressors perfectly predicting choice in a spurious manner.

Table 25: The importance of Big Five personality factors in occupational choice regressions: UK and Germany

	Percentage of occupations with 95% significance in:						
	Any B5 factor	Placebo trials	A	C	E	N	O
UK, no values	56.6	22.3	12.0	12.0	10.8	9.6	32.5
UK, no values, age ≥ 30	50.6	-	8.4	10.8	7.2	9.6	31.3
UK, with values	47.0	-	9.6	9.6	7.2	8.4	16.9
Germany, no values	63.6	22.3	11.9	25.9	16.8	14.0	40.6
Germany, no values, age ≥ 30	62.9	-	10.5	23.8	15.4	14.7	36.4
Germany, with values	50.3	-	7.0	19.6	15.4	5.6	21.0

For example, in the right part of the table, the top-right number of 32.5 means 'in the UK, when omitting value variables from the specification, Openness is significantly related to choice for 32.5% of occupations'. In the left part, the number 56.6 means personality is significantly related to choice for 56.6% of occupations, compared with only 22.3% using randomly generated data.

Table 25 shows a summary of results from probit regressions performed for each occupation.[8] Age, age squared, gender, and marital status were included as independent variables alongside Big Five personality factors in all regressions. I perform each set of these regressions twice - with and without controls for values. Values are captured by variables such as closeness to political parties, and membership of a particular club or society.

In the regressions that did not include value variables, over half of the occupations in both the UK and Germany had at least one significant Big Five factor at the 95% level. Whilst this is not proof of a causal relationship, it does indicate that personality traits appear to be an important determinant of career choice. Results are particularly strong for Germany. This is potentially due to sample size - the minimum number of observations for a German probit (omitting value variables) was 6,104, as opposed to 3,462 for the UK. I repeat the 'no values' regressions for each country for workers that are at least 30 years of age, since personality is more stable after young adulthood (Lucas and Donnellan, 2011). The relationships between Big Five and choice in this case are only marginally weaker than when all workers are included. This adds some robustness to the findings.

8 Detailed estimation results for each occupation are available upon request.

Openness is the Big Five factor most closely associated with occupational choice. This is consistent with personality theory, since Openness is related to intellectualism, and therefore underpins schooling choices. For all regressions, gender is still one of the strongest predictors of occupational choice. Woods and Hampson (2010) reach a similar conclusion, but with longitudinal analysis rather than the cross sectional analysis presented here. They do not measure actual occupation choice, but occupational environment choice, using RIASEC classifications (Holland, 1997).

Due to the multiple comparisons problem, it is not clear whether the significance percentages in the first column of Table 25 are occurring by chance alone. The multiple comparisons problem arises when testing multiple hypotheses for a successful outcome. The probability of success can be high by pure chance if the number of hypotheses being tested simultaneously is large (see Abdi, 2007). In our case, our hypothesis is that personality is related to occupational choice, but we are simultaneously testing this hypothesis 83 times for the UK and 143 times for Germany. In order to assess the validity of our result, we need to be able to show success is more frequent than we would observe due to random noise alone.

Therefore, I ran a 100 trial placebo simulation of the simple regression specifications for each country. From a population of 2 million randomly generated observations, I draw a sample of 20,000 for each trial and run the required number of probit regressions.[9] The average proportion of probits that had at least one significant personality factor at the 95% level over the 100 trials was 22.3% for both the UK and German simulations (Table 25). As the simulations were performed only on the model with the fewest variables, this figure is an upper bound. Therefore, quantity of data at the occupational level notwithstanding, we can reject the null hypothesis that personality has no influence on occupational choice. This result is consistent with general findings from previous literature (e.g. Heckman, Stixrud, and Urzua, 2006; Woods and Hampson, 2010).

9 Specifically, in place of dummy and categorical variables, I create a random variable with equal probability on each of the binary values or categories. For age, I draw random observations from a truncated normal distribution with mean, standard deviation, and range defined by my sample data. For personality traits, I also draw from a normal distribution, truncated between 1 and 7, with a mean of 4 and variance 1. 'Jobs' are assigned uniformly from a pool of 200.

The evidence suggests that there is likely to be a different weighting between the personality, values (such as political preference), and income motivators for choosing each occupation. There also appears to be some overlap between the Big Five and values, since predictive power of Big Five factors falls, in general, when introducing value regressors. The exact nature of this relationship is not clear and is beyond the scope of the present work.

An interesting finding is that those occupations where the Big Five seems to have the highest predictive power have a component of specialist skill involved. This supports the idea that the personality motive is much more likely to play a significant role in occupational choice where the cost of entry into that occupation (psychological as well as material) is non-trivial. An intuitive argument to explain occupations where value variables and Big Five variables lacked in predictive power is that there are some occupational categories that are not end career goals. They may be representing an interim occupation (a 'stop gap'), where the income motive dominates other channels shown in Figure 22.

4.3.2 *Personality and Industrial Composition*

4.3.2.1 *Predicting Industrial Composition*

In order to test Hypothesis 4.2, I use predictions from probit regressions run for the reduced set of 83 occupations for the UK and 143 for Germany. In order to assess the relative predictive power of the Big Five personality factors on industrial composition, I use predicted probabilities from three probit specifications. First, I include only the Big Five factors as independent variables. Second, I include only a gender dummy, age and age-squared. This regression allows us to compare the predictive power of personality to the classical exogenous demographic variables used in empirical analysis of individual differences. Third, I include only two education dummies - whether the individual has a degree, and whether they have a vocational qualification.

The method for obtaining the predictions was as follows. For each of the three probit specifications, predicted probabilities were obtained, holding independent variables at their mean levels for each occupation. The mean of these predictions was taken over the individuals from each occupation j to obtain

$\hat{p}_{sj}$ for specification $s \in \{Big\ Five, Demographic, Education\}$. As the predicted probabilities were very close to 0.5, I subtracted this common component from each $\hat{p}_{sj}$, in order to highlight the differences between the predictions. Each occupation can potentially map onto multiple industries. Therefore, weights were inferred from the sample. Occupation j is mapped onto industry $i \in \{1, 2, ..., 10\}$ with weight w_{ij}, hence:

$$\sum_{i=1}^{10} w_{ij} = 1 \tag{27}$$

The 10 industry categories were obtained from the German 2011 census.[10]

The overall prediction for each industry, given specification s, is:

$$\hat{P}_{si} = \sum_{j=1}^{J} w_{ij}[\hat{p}_{sj} - 0.5] \tag{28}$$

where $J = 83$ for the UK and 143 for Germany. $\hat{P}_{si}$ has no interpretation by itself. Instead, it is converted to a percentage in order to allow for a valid comparison between predictions and true sample data:

$$\hat{P}^{\%}_{si} = \frac{\hat{P}_{si}}{\sum_{i=1}^{10} \hat{P}_{si}} \tag{29}$$

Table 26 gives a breakdown of the raw industrial compositions for the UK and Germany from their 2011 censuses, and compares these to $\hat{P}^{\%}_{si}$ from the three probit specifications, as described above. Whilst the bulk of the prediction in absolute terms comes from the relative weights w_{ij}, one can see that there is variation between the three specifications (albeit small). This allows us to compare them in order to assess the relative strength of personality traits in determining industrial composition.

Figure 23 plots the predictions and census data from Table 26 in graphs for the UK and Germany respectively. If we view each industry in isolation, we see

10 The UK 2011 census has a finer classification, but it is relatively simple to subsume these into the same 10 industry categories used in the German census.

TABLE 26: Personality predictions of industry composition outperform predictions based upon demographics or education for some industries.

		UK				Germany			
	Industry	Census	Big Five	Age, Sex	Education	Census	Big Five	Age, Sex	Education
1	Agriculture, Forestry, Fishing	0.86	1.33	1.26*	1.35	1.69	1.31	1.28	1.44*
2	Mining, Manufacturing	9.11	11.02*	11.71	11.10	19.08	20.53	20.90	20.30*
3	Utilities, Water, Sewerage	1.28	1.06*	1.03	1.04	1.35	1.73	1.71*	1.72
4	Construction	7.70	6.58*	9.88	6.21	5.63	5.56*	6.80	5.72
5	Wholesale, Retail, Motoring, Accommodation, Food	21.51	19.11	18.19	21.99*	17.04	15.44	15.98*	15.88
6	Transportation, Storage, Information & Comms	8.93	7.51	7.32	7.91*	8.29	6.50	6.19	6.54*
7	Financial & Insurance	4.32	5.69	5.05*	5.54	3.15	4.57	4.40	4.33*
8	Real Estate, Professional, Scientific, Admin & Support	12.91	8.55*	8.39	7.92	12.78	8.53*	8.22	8.43
9	Public Admin, Defence	6.00	8.04	7.30*	7.33	7.45	7.62	7.41*	7.54
10	Education, Health, Arts & Recreation, Other Services	27.39	31.10	29.86	29.60*	23.53	28.22	27.12*	28.09

** indicates the closest prediction for a given industry (viewed independently). Numbers are percentage shares for each industry. Census data is from 2011 for both countries.*

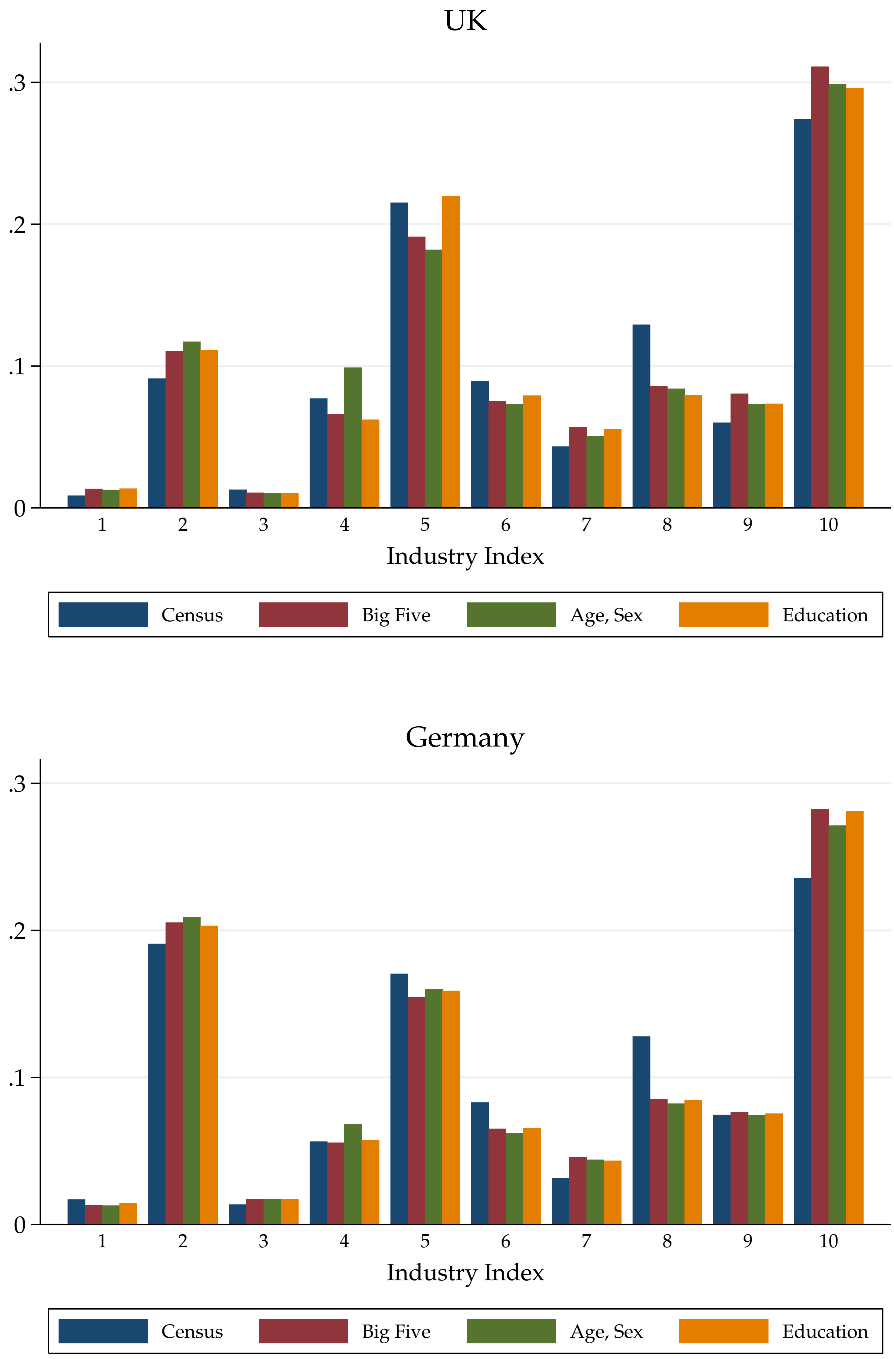


FIGURE 23: Predicted industry proportions from three specifications compared with actual 2011 Census data.

that the Big Five predictions are closest to more individual industries than the other two specifications for the UK. The Big Five predictions are outperformed by both the demographic and education specifications in Germany. However, from these results, it appears that certain industries are more closely associated with personality differences. In particular, we see that for both countries, Construction, and the combined sector containing Real Estate, Professional, Technical, Scientific and Admin & Support services, have relative sizes best predicted by the Big Five.

In order to obtain a more sophisticated comparison between the predictions, we need to compare them more holistically. I compare distributional differences between each prediction specification and actual census data using the ES test (Epps and Singleton, 1986). This is similar to, but generally more powerful than, a two-sample Kolmogorov-Smirnov test (Goerg and Kaiser, 2009). The results are presented in Table 27.[11] It is clear that the null hypothesis of no difference between actual and predicted distributions was not rejected for any of the three specifications. However, examination of the magnitudes of the ES test statistics reveals the relative performance of each specification. The education specification provides by far the best distributional prediction for the UK, whilst the demographic specification appears to be the best for Germany. In particular, the education specification outperforms the Big Five specification in both countries.

TABLE 27: Personality does not predict the overall distribution of industries better than education.

	UK		Germany	
Specification	ES test statistic	p-value	ES test statistic	p-value
Big Five	1.080	0.89737	0.920	0.92164
Age, Sex	1.289	0.86316	0.811	0.93703
Education	0.353	0.98616	0.867	0.92918

Results of test comparing the difference between prediction distributions and the true census distribution of industries.

11 ES statistics were evaluated at default parameter values ($t_1 = 0.4, t_2 = 0.8$), and a small sample correction factor of 0.6014 was applied. See Goerg and Kaiser (2009) for further details.

4.3.2.2 *The Education Channel*

Referring back to Figure 22, it may be the case that the channel linking personality with ability and schooling has a much stronger role to play in determining industrial composition than any direct influence personality may have. If true, this may ease the mind of labour economists who have spent their careers studying the effects of schooling on occupational choice. However, it would also serve to highlight the important role that personality plays as a more primitive determining factor of observed individual choice.

In their meta-analysis of the Big Five and academic performance, Poropat (2009) finds that Conscientiousness, Openness, and Agreeableness (in that order) have the strongest correlations with Grade Point Average (GPA) scores. All three of these factors have positive effects on GPA, and Conscientiousness has an association similar in magnitude to that of intelligence. This association appears to become stronger for later levels of education, whereas the effects of Agreeableness and Openness weaken. Conscientiousness, and to a degree Openness, therefore are likely to be driving choices pertaining to higher education.

To test this, I first run OLS regressions to measure the relationship of Big Five factors with respect to whether an individual has a degree, or a vocational qualification.[12] Since heteroskedasticity is always present when estimating a linear probability model, robust standard errors are used (Angrist and Pischke, 2008). The results are shown in specifications (1) to (4) of Table 29. Conscientiousness is highly significant both on whether an individual had a degree and whether they had a vocational qualification, for both the UK and Germany. Other Big Five factors were also highly significant, particularly Openness, which we would expect from the previous research on academic performance.

Parameter estimates are unlikely to be reliable for simple OLS, since ability/intellect is unobserved, but is certainly correlated with schooling choice, and very likely to be correlated with personality factors (especially Openness). Previous research has shied away from using an Instrumental Variables (IV) approach, primarily because reliable instruments for personality are difficult to find. Bowles, Gintis, and Osborne (2001) explain that one can use childhood personality, or

12 In the UK, having a degree simply means an undergraduate or postgraduate university degree. In Germany, the distinction is more nuanced. A degree includes one obtained from a 'fachhochschule' (an applied sciences institution); any other university degree; or a doctorate.

personality prior to labour market entry as an instrument for current personality. However, since childhood personality is unstable, this may prove to be a weak instrument. On a more practical note, save for ambitious long-term panels, data on pre-labour market personality is often non-existent, as is the case with the data used in the present study. The alternative is to construct instrumental variables from other observable data, but the best way to do this is not clear.

I attempt to utilise the variables related to individual values used for some of the occupational probit regressions. These predominantly capture political attitudes, preferences related to leisure activities, and membership of social groups (such as volunteering organisations or religious groups). A first stage regression using the education dummies as dependent variables identified those value variables that had no significant relationship with schooling choice (i.e. those that satisfy the exclusion restriction). These were then used as instruments for the Big Five variables. IV estimates are shown in specifications (5) to (8) of Table 29.

In general, the instruments used were quite weak. The correlations between instruments and the Big Five were stronger for Germany than for the UK. The best correlation between a personality factor and an IV for the UK was 0.1212, whereas it was -0.3039 for Germany. This is likely to explain the weaker significance of Big Five coefficients in the IV regressions for the UK. However, it does appear that if the transmission mechanism maps personality onto schooling choice, then onto occupation choice and industry, it may be more likely to work through an academic educational channel rather than a vocational one.

The final interesting observation from Table 29 concerns the direction of the personality coefficients. Looking at specifications (5) and (7), with the degree dummy as the dependent variable, we see that the coefficient estimates for Conscientiousness, Extraversion, and Neuroticism are negative (although the latter is non-significant for Germany). In the case of Conscientiousness, this is quite surprising, since previous research has emphasised the positive relationship of Conscientiousness with academic performance. This may be due to the mediating influence of age or gender, but research has shown that these tends to lower the magnitude correlations, and not flip their direction completely (Poropat, 2009).

A more plausible explanation could be that *academic performance* does not proxy well for *academic choice*. In other words, more Conscientious students who are already in a particular degree program may do better, but would be less likely to have elected to be in that program to begin with. This is supported by the positive (although non-significant) coefficients for these three personality factors in the vocational regressions (6) and (8). The implication is that more Conscientious, more Extraverted, and more Neurotic people may find it more beneficial for their underlying personality to opt for a more vocational career. This is not entirely unintuitive. More Conscientious people may tend to like to see some tangible outputs, which they might not find in theoretical study. More Extraverted people may be better suited to industry, and less suited to the self-study that academic learning implies. More Neurotic people may find academic study and assessment to be too stressful. Though an in depth discussion here would take us too far from the aim of this chapter, this is an interesting finding that warrants further research.

TABLE 28: Mean probit prediction errors from schooling regressions

	Mean prediction error, using:	
	UK Data	German Data
Full specification		
UK Degree	0.308	0.188
UK Vocational	0.471	0.441
Ger Degree	0.229	0.054
Ger Vocational	0.438	0.153
Big Five only		
UK Degree	0.312	0.200
UK Vocational	0.488	0.457
Ger Degree	0.227	0.057
Ger Vocational	0.441	0.154

Notes: Bold indicates out-of-country prediction.
All errors are significantly below 0.5, $p<0.001$.

We can obtain an idea of how universal our relationships between personality and schooling are by using German schooling regressions to predict the probability of an individual in the UK data to have a particular qualification, and vice versa. I run probit versions of regression specifications (1) - (4) from Table 29,

Table 29: The relationship between education level and Big Five personality factors.

	OLS				IV			
	UK		Germany		UK		Germany	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Has degree?	Has voc qual?	Has degree?	Has voc qual?	Has degree?	Has voc qual?	Has degree?	Has voc qual?
Age	0.0210***	0.0525***	-0.0117***	0.00678***	0.0306	0.0489**	-0.00916***	0.0038
Age^2	-0.000272***	-0.000607***	0.000106***	-6.13e-05***	-0.000403*	-0.000500***	8.06e-05***	-3.14e-05
Male?	-0.0455***	0.00624	-0.00184	0.0115**	-0.797**	0.345	0.00517	0.897
Agreeableness	-0.0143***	0.00587	0.00201	0.00487*	-0.107	0.0707	0.0730***	1.075
Conscientiousness	-0.0340***	0.0259***	-0.0119***	0.0145***	-0.600*	0.0916	-0.0928***	0.559
Extraversion	-0.0297***	0.00259	0.000923	0.00312	-0.365*	0.202	-0.0205**	1.078
Neuroticism	0.00629	-0.000399	-0.00105	-0.00430*	-0.681*	0.349	-0.00416	0.386
Openness	0.0823***	-0.00227	0.00810***	0.00617***	0.591**	-0.161	0.0430***	-0.0927
Constant	-0.145**	-0.796***	0.343***	-0.227***	5.343*	-3.173	0.341**	-15.71
Observations	7017	7017	12637	12637	3252	3252	10537	10537
R^2	0.058	0.04	0.067	0.014	-	-	-	-

Robust standard errors used, but omitted for brevity. *** p<0.01, ** p<0.05, * p<0.1

and calculate predicted probabilities. I also run probit versions of (1) - (4) but using only the Big Five variables as regressors.

The mean differences between predictions and actual educational dummies are shown in Table 28. Bold numbers in the table indicate where an out-of-country prediction has been made i.e. from a model inferred using the other country's data. A simple t-test rejects the null hypothesis that the mean error is equal to 0.5 for all prediction errors in the table, with $p < 0.001$ in all cases. This means that our out-of-country predictions of educational qualification status are better than random. One can see that this is particularly true when predicting degree status, since the prediction error is close to 0.2 when predicting German data from a UK regression or UK data from a German regression. In other words, this is closer to 0 (perfect prediction) than it is to 0.5 (completely random prediction). This represents greater predictive power than the vocational qualification regressions provide, as we would expect from the results in Table 29. Furthermore, we see that there is very little difference between the full specification, and the specification that only takes personality into account. This suggests that demographic variables have less power in predicting education choice, relative to the Big Five factors.[13]

Therefore, the relationship between Big Five factors and higher education choice appears to be relatively stable across similar countries. The implication is that the influence of personality on schooling choice is likely to have some form of general validity, and that this relationship is not restricted to a single country.

4.3.2.3 *Industry Switching*

We can test whether there is a predictive link between personality traits and industrial distribution within a country by looking at whether personality has an impact on the likelihood to switch to a different industry.[14] If personality does determine industry choice and all individuals are rational and choose optimally, then everyone should be in the correct industry already. There would be no switching, and we would observe equilibrium. However, in labour market re-

13 This is confirmed by comparing probits only including demographic variables with probits only including Big Five variables.

14 Causality is more difficult to establish, due largely to the lack of availability of large scale panel data on personality over time.

ality, it is likely that individuals have imperfect information about the true type of an industry, and will switch in future if they are mismatched today.

The following specification is used for estimation:

$$\begin{aligned} SWITCHED_w = \; & \beta X_w + \gamma BIGFIVE_w + \delta VALUES_w \\ & + \theta INCOME_w + \epsilon_w \end{aligned} \tag{30}$$

Regressors are all taken from 2005 data. The dependent variable is a dummy with a value of 1 if worker w has a different industry category in a given year, compared with 2005. X includes demographic factors and a constant.

In Table 30, I estimate the above regression for the German SOEP, with all years of the panel post-2005. Neuroticism is predictive of industry switching 3-4 years into the future. Openness has a significant influence on switching beyond 5 years. However, sample sizes for the 2010 and 2012 regressions are very low. The most consistent predictor of industry change is gross wage. These results suggest that although there does not appear to be any systematic pattern, personality does appear to play a role in determining which industry an individual ends up in.

To explore this relationship in more depth, I focus on data from 2009, which is the year with the largest pool of observations. One might expect that younger individuals with less information about occupations and industries are guided more by their personality traits than individuals later in their life cycle. At the same time, younger individuals are also more likely to be motivated by higher incomes since they have had less time to accumulate wealth. Therefore, I split the data into two groups by age. The mean age of the full German sample from both years is 42 (to the nearest year). Table 31 shows results from OLS regressions for individuals less than the mean age, and those greater than or equal to the mean.[15] I also estimated each specification using a probit model to ensure consistency. Estimates were virtually identical between the two estimation methods. Therefore, only OLS estimates are reported.

15 I omit the degree dummy variable from these regressions. Repeating the regressions with the dummy included did not change the results.

TABLE 30: Predictive power of the Big Five on German industry switching.

	Dependent variables: did individual change industry in year below, relative to 2005?						
	2006	2007	2008	2009	2010	2011	2012
Age	0.0177	0.0108	-0.0140	-0.0162	-0.0606	-0.00534	0.0344
Age^2	-7.84e-05	-6.41e-05	0.000151	0.000240	0.000990	4.90e-05	-0.000617
Male	0.129	0.0508	0.0515	0.0542	0.151	0.0504	0.0773
Been married?	0.00454	-0.0575	0.126	-0.0166	-0.114	-0.0478	-0.0989
Agreeableness	0.000702	0.0381*	0.0506	0.0232	0.0198	0.0102	-0.120
Conscientiousness	0.0424	-0.0172	0.0650	0.00239	0.00854	0.00245	0.153*
Extraversion	-0.0399	0.0124	0.0331	-0.0128	-0.00738	-0.0228	-0.0743
Neuroticism	0.0254	0.0189	0.0846**	0.0303*	0.0354	0.0185	0.0216
Openness	-0.00702	-0.0160	-0.0101	0.0209	-0.00276	0.0423**	0.167**
Satisfaction with income	0.00777	-0.0133	-0.0208	-0.0261***	-0.0276	-0.00362	-0.0228
Gross wage in 2004-05	-6.11e-05*	-4.73e-05***	-8.19e-05*	-4.22e-05**	-0.000100**	-6.65e-05***	-8.87e-05
Political attitude	0.0161	0.00698	0.0216	0.00115	-0.0320	0.00782	0.0481
Freq of sport activity (-ve)	0.0630**	-0.000217	0.0153	-0.00843	-0.0551	-0.0215	-0.0982*
Freq of artistic activity (-ve)	-0.0435	-0.0185	-0.0151	-0.00810	0.0270	0.0293	0.116
Optimism towards future (-ve)	-0.0237	-0.00873	-0.0390	-0.0328	-0.127*	-0.0235	0.0635
Constant	-0.334	0.0157	-0.226	0.595*	1.804*	0.369	-0.809
Observations	163	568	148	622	88	470	68
R^2	0.112	0.042	0.110	0.058	0.190	0.077	0.371

Robust standard errors are omitted for brevity. *** p<0.01, ** p<0.05, * p<0.1

TABLE 31: The predictive power of the Big Five on German industry switching, by age.

	Dep variable: Changed industry between 2005 and 2009?					
	Age $<$ 42	Age $\geq$ 42	Age $<$ 42	Age $\geq$ 42	Age $<$ 42	Age $\geq$ 42
Age	0.0320	0.0706	0.0271	0.0506	0.0328	0.0227
Age^2	-0.000538	-0.000536	-0.000437	-0.000336	-0.000550	-9.24e-05
Male	0.0652	-0.0167	0.0730	-0.0407	0.0772*	-0.0325
Been married?	-0.00734	0.0795	-0.0181	0.0501	-0.0160	0.0590
Agreeableness	0.0652***	-0.0811*	0.0554**	-0.0854*	0.0530**	-0.0965*
Conscientiousness	0.00529	-0.0123	0.0110	-0.0130	0.00230	-0.00101
Extraversion	-7.87e-05	-0.0520	-0.00597	-0.0696*	-0.00759	-0.0789*
Neuroticism	0.0242	0.00472	0.0353*	0.000898	0.0450**	0.00366
Openness	0.0242	0.0384	0.0133	0.0737*	0.0130	0.0767
Satisfaction with income	-0.0308***	0.00108	-0.0359***	0.00446	-0.0377***	0.00174
Gross wage in 2004-05	-5.48e-05***	-3.79e-05	-6.40e-05***	-3.61e-05	-4.88e-05**	-4.84e-05
Political attitude	-	-	-0.00155	0.0235	-0.00276	0.0184
Freq of sport activity (-ve)	-	-	-0.00351	-0.0299	-0.00159	-0.0324
Freq of artistic activity (-ve)	-	-	-0.0194	0.0356	-0.0184	0.0423
Optimism towards future (-ve)	-	-	-0.0705**	0.0738	-0.0525	0.0646
Current health (-ve)	-	-	-	-	0.000146	-0.0272
Satisfaction with health	-	-	-	-	0.0199	-0.00553
Life satisfaction	-	-	-	-	0.00642	-0.00135
Constant	-0.474	-1.248	-0.110	-1.073	-0.389	-0.101
Observations	474	167	461	161	471	149
R^2	0.077	0.075	0.093	0.095	0.090	0.092

Robust standard errors are omitted for brevity. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

First, one notices that the relationship between switching and income is strongly significant only for those younger than the mean. The higher the income and the more satisfied one is with their income, the less likely one is to switch industries. This is consistent with the wealth accumulation argument made earlier.

Second, for this lower age group, individuals with higher Agreeableness are more likely to switch industries. A one point higher Agreeableness score (for an individual under 42) increases the likelihood of being in a different industry in 4 years by 5-7%, depending on the specification used. Theoretically, it is not immediately obvious why this is the case. One possible explanation is that Agreeable people are more mobile, due to their more accommodating nature. Neuroticism has weak positive significance on the probability of switching in the final two specifications for the younger age group. This is theoretically easier to justify. Since Neuroticism is related to the degree of responsiveness to negative affect, an industry mismatch is likely to have more of a detrimental effect on utility for individuals that are more Neurotic.

Results for those aged 42 and over are weaker. Income and optimism appear to have little predictive power. Personality traits do seem to have effects of some significance, although the relationship is not strong. If we look at the fourth regression in Table 31, we see that the predictiveness of Agreeableness is reversed, relative to the younger age group. This is more intuitive. Less Agreeable people are less likely to have a need to please others, and so we might expect them to be more ready to leave an industry which they perceive as not suiting their best interests. However, since the sample size is quite low for this age group, we cannot rule out the fact that this could have been a spurious result.

In the last pair of regressions in Table 31, I add variables corresponding to health and life satisfaction to determine whether these are the missing determinants of switching for the older age group. None of the coefficients for these variables were significant. However, for the younger age group, including health and life satisfaction variables increases the predictive power of Neuroticism. A worker with one extra unit of Neuroticism has a 4.5% higher probability of switching industries in future. From the same regression we also see that a unit increase in Agreeableness for a worker means that they would be 5.3% more likely to switch industry in future.

A similar analysis was performed for the UK. Table 32 shows OLS linear probability regressions for three separate years after the personality data was obtained. As was the case with Germany, the strongest predictive factors on the probability of switching industries in future are due to labour income levels and satisfaction with this income. However, personality also has strong predictive power.

Conscientiousness is a strong negative predictor of switching in all three future years. A one point increase in one's Conscientiousness reduces the probability of switching industry by just over 1%. This effect becomes less significant (though the point estimates are similar) when value-related variables are added. After controlling for factors such as political preference and group membership, Neuroticism and Openness, in particular, positively predict industry switching in the following year. Neuroticism is also predictive in following years, though to a lesser extent

From these regressions, it appears that values and preferences have more predictive power than personality. In particular, political preference, trade union membership, desire for one's own business and sports club membership significantly predict industry switching. However, as with the argument put forward related to schooling, from a theoretical basis, it is quite likely that personality has a causal effect on the formation of some of these values.

Performing simple pairwise correlations between the Big Five and these value variables finds significant relationships at the 1% level. The strongest of these is a positive correlation between the desire to own one's own business and Openness. Therefore, although the predictive power of personality is not consistently high, we may not be capturing all of its impact due to input into preferences.

Finally, the UK regressions were separated for age. Results of the regressions performed without value variables are given in Table 33.[16] Income and values are the most significant predictors of industry switching. Personality is strongly predictive when values are omitted. The strength of these relationships, however, are smaller than those in Germany. With the age separated regressions, it is Conscientiousness that appears to have the strongest relationship with future industry switching in the majority of cases, though Agreeableness also appears strongly for the older age group. This is slightly different to the results observed for Germany, where Agreeableness dominates in general. Therefore, it is not clear that the same personality traits have the same effects on industry movement in different countries.

16 Remaining regression results are excluded due to brevity and a lack of additional insight. The findings are qualitatively similar to the results in Table 32.

Table 32: Predictive power of the Big Five on UK industry switching, using OLS.

	Dependent variables: did individual change industry in year below, relative to 2005?					
	2006	2007	2008	2006	2007	2008
Age	-0.0126***	-0.0202***	-0.0212***	-0.00634	-0.0165***	-0.0206***
Age^2	0.000134***	0.000215***	0.000214***	5.89e-05	0.000175**	0.000213***
Male	0.0156	0.0110	0.00590	0.0109	0.00642	0.00545
Been married?	-0.00189	0.00950	0.00975	0.000284	0.00633	0.0247
Agreeableness	0.0128**	0.0122*	0.00892	0.00218	-0.00145	0.000162
Conscientiousness	-0.0119**	-0.0142**	-0.0152**	-0.00855	-0.0119	-0.0148*
Extraversion	0.00363	0.00583	0.00697	0.00869	0.0127*	0.0127*
Neuroticism	0.00578	0.00387	0.00365	0.0117**	0.0105*	0.00729
Openness	0.00380	0.00203	0.00332	0.0148**	0.0103	0.00763
Wage in 2004-05	-4.75e-07**	-8.51e-07**	-9.93e-07**	-1.03e-06***	-1.40e-06***	-1.42e-06***
Satisfaction with pay	-0.0140***	-0.0195***	-0.0172***	-0.0131***	-0.0171***	-0.0159***
Closest to:						
- Tory party	-	-	-	0.0898***	0.0676***	0.0626***
- Labour party	-	-	-	0.0465***	0.0411**	0.0299
- Lib Dem party	-	-	-	0.0624***	0.0496**	0.0448
Belong to a social class?	-	-	-	-0.0142	0.00262	0.000607
Member of:						
- Trade union	-	-	-	-0.0346**	-0.0454***	-0.0564***
- Environmental group	-	-	-	0.0501	0.0485	0.119**
- Parents association	-	-	-	0.00714	-0.0223	0.00431
- Tenants group	-	-	-	0.00425	0.00127	0.0575
- Religious group	-	-	-	-0.0150	-0.0166	-0.0146
- Voluntary service grp	-	-	-	-0.0108	0.0391	0.0319
- Sports club	-	-	-	-0.00904	-0.0340**	-0.0322*
- Women's group	-	-	-	0.0444	0.00926	-0.0912
Would like own business	-	-	-	0.0120	0.0664***	0.0675**
Constant	0.433***	0.688***	0.766***	0.226*	0.548***	0.690***
Observations	4,994	4,921	4,783	2,778	2,723	2,654
R^2	0.015	0.025	0.028	0.028	0.040	0.043

Robust standard errors are omitted for brevity. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE 33: Predictive power of the Big Five on UK industry switching, by age.

	Switched 2006?		Switched 2007?		Switched 2008?	
	Age < 40	Age ≥ 40	Age < 40	Age ≥ 40	Age < 40	Age ≥ 40
Age	-0.0467***	-0.0131	-0.0487***	-0.0271	-0.0585***	-0.0200
Age^2	0.000729***	0.000138	0.000690***	0.000272	0.000830***	0.000185
Male	0.00671	0.0234*	-0.0133	0.0360**	-0.00479	0.0193
Been married?	-0.00227	0.0192	0.0155	0.0169	0.0219	0.00639
Agreeableness	0.00492	0.0213***	0.00424	0.0201**	0.00333	0.0142
Conscientiousness	-0.0132	-0.0126*	-0.0184*	-0.0119	-0.0181*	-0.0130
Extraversion	0.00354	0.00325	0.00186	0.00881	0.00578	0.00743
Neuroticism	0.00517	0.00606	-0.00158	0.00853	0.00297	0.00422
Openness	0.00483	0.00338	0.00258	0.00270	-0.000397	0.00820
Wage in 2004-05	-5.08e-07*	-3.58e-07	-8.28e-07*	-9.01e-07**	-9.27e-07*	-1.09e-06***
Satisfaction with pay	-0.0144**	-0.0148***	-0.0173***	-0.0221***	-0.0146**	-0.0197***
Constant	0.955***	0.385	1.192***	0.792	1.357***	0.725
Observations	2,446	2,548	2,429	2,492	2,375	2,408
R^2	0.016	0.010	0.027	0.016	0.030	0.013

Robust standard errors are omitted for brevity. *** p<0.01, ** p<0.05, * p<0.1

4.3.3 *Cross-Country Big Five Differences and Differences in Industrial Composition*

4.3.3.1 *Cross-country Big Five comparisons*

I first conduct a simple comparison of the means of the Big Five traits for both the full sample and the minimum age sample using t-tests. Schmitt et al. (2007) explain that cultural biases that affect responses are difficult to control for and rule out completely. The primary form of this is known as acquiescence bias - a predisposition to agree with statements. However, they reference a number of studies that support the fact that a comparison of means across cultures is valid. Since the personality measure used in the BHPS and the SOEP is identical, save for language, we can compare mean trait scores between the UK and Germany directly.

Table 34 shows that mean Big Five scores are significantly different between the two countries for all five factors ($p < 0.01$ in all cases). I also performed the Kolmogorov-Smirnov and Epps-Singleton test of distributional difference. Both tests show a highly significant difference in personality trait distributions between countries for all Big Five factors.[17]

In particular, mean Conscientiousness is very different between the UK and Germany (as confirmed by the large t-statistic). This can be seen visually in Figure 24. To the best of my knowledge, this is the first report of such a large difference in Conscientiousness between these countries. The difference is also large for Extraversion, which is less obvious from visual inspection alone. We can see from the graphs that Conscientiousness seems to have a larger rightward (negative) skewness in the German sample, relative to the UK. Explicitly, the skewness of Conscientiousness is -0.74 for the Germany, but -0.36 for the UK.

This finding is in contrast to previous research, which finds *similar* trait means for historically and geographically close countries. Schmitt et al. (2007) use a slightly more comprehensive 44 item personality inventory, but with smaller sample sizes.[18] They find Conscientiousness scores of 46.52 for Germany and 46.89 for the UK (where mean = 50, standard deviation = 10). This is opposed

17 The details are omitted, as they are not important for the present analysis.
18 In particular, 483 from the UK and 790 from Germany. This was a convenience sample of college students and those from the wider community.

TABLE 34: There is a significant difference in the means of personality traits between the UK and Germany.

	Agreeableness	Conscientiousness	Extraversion	Neuroticism	Openness
Full Sample					
UK mean	5.428	5.417	4.578	3.622	4.592
UK s.d.	(0.947)	(0.964)	(1.119)	(1.236)	(1.095)
German mean	5.312	5.883	4.868	3.729	4.492
German s.d.	(0.972)	(0.892)	(1.150)	(1.203)	(1.180)
t-statistic	-8.145***	33.340***	17.220***	5.883***	-5.994***
Age $\geq$ *30*					
UK mean	5.437	5.482	4.500	3.602	4.551
UK s.d	(0.945)	(0.965)	(1.130)	(1.238)	(1.103)
German mean	5.309	5.954	4.829	3.731	4.481
German s.d	(0.976)	(0.861)	(1.146)	(1.202)	(1.189)
t-statistic	-7.962***	30.258***	17.223***	6.297***	-3.665***

t-tests are performed under the assumption of different variances for the UK and Germany.

*** p<0.01, ** p<0.05, * p<0.1

to my finding that Germans have significantly higher mean Conscientiousness. Furthermore, they find the UK has slightly *higher* acquiescence bias. This suggests that respondents from the UK would be expected to answer more strongly in agreement with positively worded Conscientiousness items, therefore theoretically biasing the mean UK Conscientiousness score *upwards*.[19] If we were to adjust for this bias here, it would suggest an even greater difference between the UK and Germany.

19 There were two positively worded items and one negatively worded item for Conscientiousness in the survey.

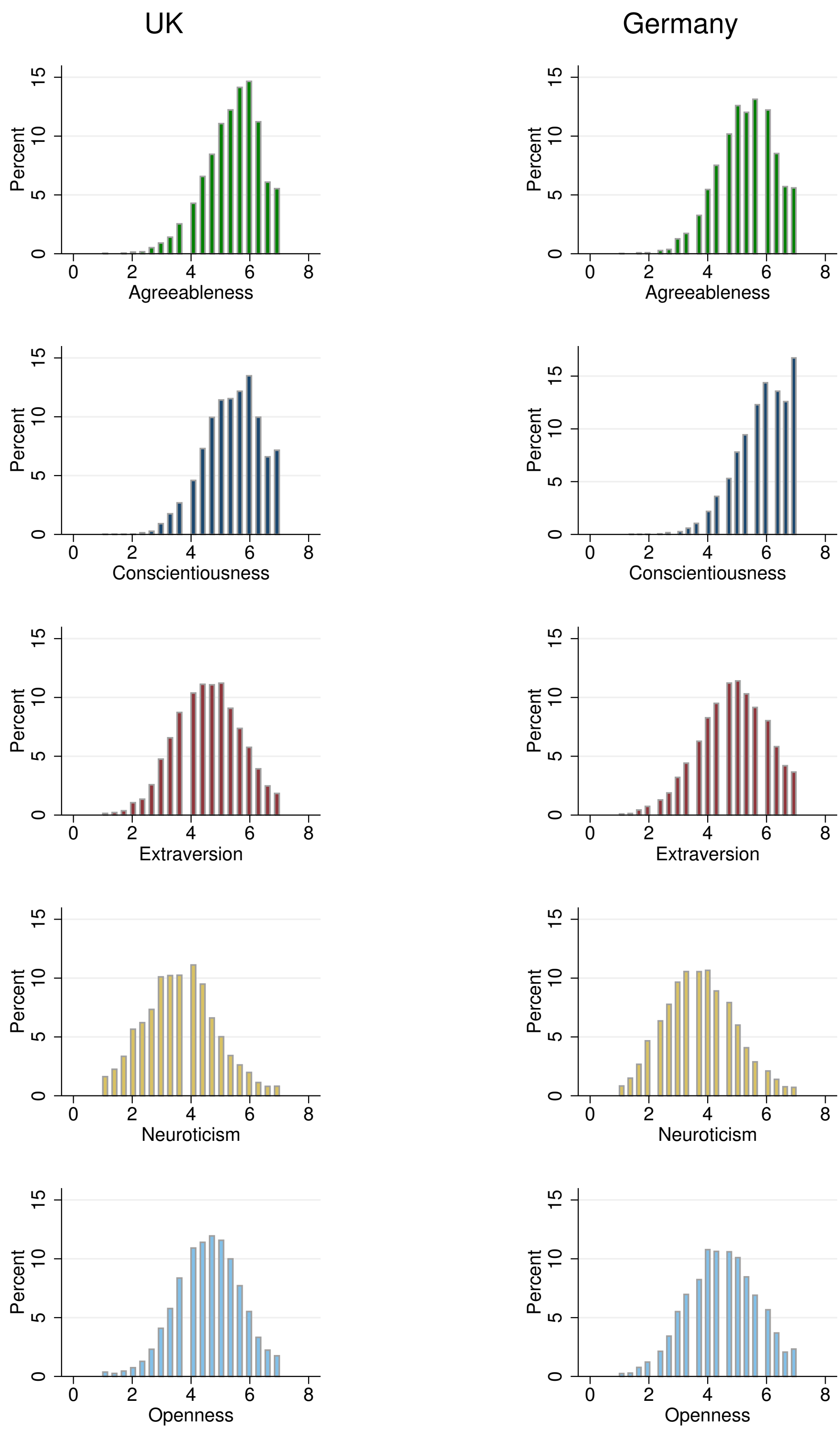


FIGURE 24: A comparison of the different distributions of Big 5 factors in the UK and German workforces. UK data is from BHPS 2005; n=7,017. German data is from SOEP 2005 & 2009 combined; n=12,637.

One key difference from existing research is that the sample in the present study is considerably larger, representative, and concerns only those in the labour force. However, even setting aside data concerns, previous research would still predict similar means due to the relative historical and geographic similarities between the UK and Germany. Therefore, the finding of significant mean trait score differences for all Big Five factors may be considered as surprising. Since Conscientiousness has strong predictive implications for a number of economic variables (such as unemployment probability and job satisfaction), large cross-country differences may be of particular interest to economists.

4.3.3.2 *Paired Regression*

We can now test the final hypothesis of this chapter - whether personality trait differences predict differences in industrial composition between the UK and Germany (Hypothesis 4.3). In order to test this, we need to obtain trait differences at the individual level. This requires the pairing of observations from the two datasets.

There are two main ways to achieve this. The first is through propensity score matching (see Rosenbaum and Rubin, 1983). This involves fitting a binary choice model with a dummy differentiating between two groups as the dependent variable (in our case, this would be a country dummy). The regressors are variables that should be taken into account to indicate matching proximity. Predicted probabilities from this model are calculated for each observation - these are the propensity scores. Finally, observations are paired using an appropriate algorithm to compare propensity scores. The second is by exact matching. This directly matches observations based on specified observed covariates, but is more computationally expensive as a result. As a result of this I focus on the propensity score method.[20]

A logit model is estimated that has a country dummy as the dependent variable, and independent variables that reflect the individual characteristics we want to take into account for optimal pairing. While it is tempting to include every conceivable covariate as a regressor in calculating the propensity score, this is not advisable as it can reduce precision of estimates (Caliendo and Kopeinig, 2008).

20 I discuss a variant of exact matching in Section 4.4.

Therefore, I include age, gender, whether an individual has been married, and education dummies. To match the resulting propensity scores, I use the one-to-one nearest neighbour algorithm, as implemented by Leuven and Sianesi (2014). Calipers of 1 and 0 were used in order to provide added precision, but there was no difference in the matches from reducing the caliper value.[21] This suggests that the algorithm provided relatively good matches.

Absolute differences between the personality traits were calculated and used as independent variables. In addition, all five factors were taken as a single vector for each individual. The Euclidian distance between the two personality vectors in a pairing was calculated, and this was used as an alternative independent variable.[22] Table 35 shows results of linear probability regressions that estimate whether the fact that two individuals in a pair work in a different industry is related to differences in personality traits.

Overall model fit, in general, was quite poor for all specifications estimated, suggesting (as we would expect) that personality is not the only determinant of industry differences. When estimating both the vector difference and component difference specifications on the full paired sample, coefficients were not significantly different from zero. However, as we have seen from previous analysis, personality influences are stronger when separating young from old. The regressions were repeated for those below 40 and those greater than or equal to 40 years of age.

First, we notice that the vector personality difference is significantly associated with greater industry difference for both age groups. This suggests Big Five personality difference *does* have some influence on industry differences across countries. A one point increase in the difference between the personality vectors of a British and a German individual corresponds to a 1% higher likelihood of the individuals working in different industries. To put this into perspective, if two individuals differed exactly by one point on all five personality dimensions, they would be approximately 2.24% more likely to be in different industries than if their personalities were identical. If we compare someone with an extreme personality (i.e. with all trait scores either 1 or 7) to someone with a cen-

21 A caliper defines a maximum distance between propensity scores, beyond which a match is not made.

22 Regressions using relative differences and quadratic terms were also performed - the fit of these models was similar to the absolute difference models.

TABLE 35: Relationship between Big Five differences and industry differences, using propensity score matched data from UK and Germany.

	Dep variable: Are paired individuals in different industries?					
	Full Sample		Age < 40		Age $\geq$ 40	
Abs diff in A	-0.00264	-	0.00883	-	-0.0124	-
	(0.00562)		(0.00727)		(0.00790)	
Abs diff in C	-0.00586	-	0.00390	-	0.0125	-
	(0.00568)		(0.00741)		(0.00784)	
Abs diff in E	-0.00304	-	0.00644	-	0.00879	-
	(0.00486)		(0.00641)		(0.00662)	
Abs diff in N	0.00135	-	0.000137	-	0.0125**	-
	(0.00444)		(0.00595)		(0.00621)	
Abs diff in O	0.000358	-	0.00888	-	0.00376	-
	(0.00469)		(0.00643)		(0.00654)	
Euclidian dist between B5 vectors	-	-0.00269	-	0.0105**	-	0.0111**
		(0.00401)		(0.00536)		(0.00557)
Constant	0.829***	0.827***	0.806***	0.806***	0.782***	0.779***
	(0.0131)	(0.0138)	(0.0179)	(0.0188)	(0.0186)	(0.0200)
Observations	7,017	7,017	3,475	3,475	3,542	3,542
R^2	0.000	0.000	0.001	0.001	0.003	0.001

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

tred personality (all trait scores at 4), they would be approximately 6.7% more likely to be in different industries.

We would expect that older individuals, having found a better industry match as a result of industry movement when young, would show stronger personality difference influences on industry difference, relative to the young. This is consistent with what we see in Table 35. In particular, one-point differences in Conscientiousness and Neuroticism both account for 1.25% higher likelihoods of being in different industries for those at least 40 years of age. Whilst the Conscientiousness coefficient is not significant, we would have expected it to be among the largest due to the large differences that were found in this factor between the UK and Germany. It is also consistent with the fact that Conscientiousness plays a strong role in determining education choice. For the same reason, we might also have expected a stronger relationship for Openness differences. However, this was not found in the data.

The strong influence of Neuroticism, relative to other factors, is more surprising. However, it is consistent with intuition. Since Neurotic people are more sensitive to negative affect, it is likely that individuals with high values of this trait would avoid highly stressful work environments. With the white-collar nature of the majority of jobs in modern society, Neuroticism differences are likely to play a more crucial role than they may have in a subsistence economy, for example. Absolute difference in Agreeableness acts in the opposite direction, which is contrary to what we would expect. However, the coefficient is not significantly different from zero, and so the point estimate may be incorrect. Overall, it appears that differences in personality between two countries are related to differences in industrial composition, albeit weakly.

4.4 ROBUSTNESS AND LIMITATIONS

The majority of studies on personality focus on correlations and some predictive effects. Conti and Heckman (2014) warn that many studies claiming causality of personality have been premature in doing so. They explain that causality is difficult to establish and often neglected due to the combination of a lack of theory regarding individual choices, and an unclear treatment of endogeneity issues. Whilst the aim of this chapter is to provide an initial exploration of a new relationship, I address two relevant forms of endogeneity for a subset of my results.

First, the personality item responses may be subject to measurement error. This is a common concern with all such subjective or self-reported behavioural response, where the true values are unobserved. Heineck and Anger (2010) calculate Cronbach's α to measure the internal reliability of the Big Five items in the 2005 SOEP wave. They find relatively low internal consistency scores for the items in the survey. I calculate Cronbach's α scores for my data and obtain similar results. For the pooled 2005 and 2009 SOEP data, I obtain scores of 0.50, 0.58, 0.69, 0.62, and 0.59 for Agreeableness, Conscientiousness, Extraversion, Neuroticism and Openness respectively. For the 2005 SOEP alone in my sample, I obtain corresponding scores of 0.49, 0.61, 0.63, 0.69, and 0.61. For the 2005

UK BHPS data, I obtain corresponding scores of 0.56, 0.53, 0.60, 0.69, and 0.66. One rule of thumb for psychometric tests is that good consistency is indicated by α scores of at least 0.7. This is what is usually found in the personality literature (Heineck and Anger, 2010). However, as explained at the beginning of Section 4.3, some consistency is inevitably sacrificed for the convenience of having a shorter list of items.

Heineck and Anger (2010) use α scores as reliability scores for each of the Big Five variables to perform Errors-in-Variables (EIV) regressions, which corrects for the measurement error bias that one would observe in OLS estimates. We can use EIV as an alternative to the linear probability OLS models that were estimated in order to test the robustness of the results.

First, I perform EIV versions of models estimated in Table 31. The results are shown in Table A3. If we compare the age-separated EIV regressions with those from Table 31, the story is very different. Whilst we found significant personality coefficients before, the EIV dramatically increases the magnitudes of these coefficients. Recall that in Table 31, we find that a one-point increase in Agreeableness for those below the age of 42 increases the probability of switching by 5.3%. According to the EIV regression, this coefficient has now been magnified to a 15.6% increase. This suggests that OLS estimates are likely to be biased towards zero as a result of measurement error. These magnitude increases are dramatic in the final EIV regression for the older age group. The R^2 for this regression is unusually high at 0.336, as many of the value variables are now also significant. The result that Agreeableness, Extraversion and Openness have effects on the probability of changing industry near or in excess of 50% warrants serious attention. However, the sample size is low, and so we should be cautious of this result as it is possible that we have an unrepresentative subsample.

In Table A4, I estimate EIV coefficients for the UK BHPS OLS regressions from Table 32. As with the SOEP EIV regressions, the absolute value of coefficient estimates tends to be larger in the BHPS EIV regressions, compared to OLS. In the regressions that omit value-related variables, the marginal (negative) impact of Conscientiousness on industry switching has increased from approximately 1.8% to over 5% in absolute terms. However, in the regressions that include values, some of the Big 5 EIV estimates are less significant than they were using OLS. This could be due to the fact that EIV does not allow for robust standard errors.

Standard errors are larger with the EIV estimates than OLS, and heteroskedasticity may be present. In sum, the EIV results suggest that measurement error is likely to be biasing coefficients towards zero.

The second endogeneity concern is whether reverse causality between personality and industry exists, affecting parameter estimates. Theoretically, in the present analysis, it appears difficult for industry to influence personality. This is because the relationships tested here are either cross-sectional, or where personality measures precede industry choice. Borghans et al. (2008) suggest reverse causality is possible even when the outcome variable is measured after personality. The reasons cited for this are related to anticipation or expectation of future outcomes causing changes in current personality. In the case of industry choice, this explanation seems unlikely. One would expect to change industry in future *because* of personality issues in the current period (as well as value/income issues). Married with the findings from Terracciano et al. (2005) and Na et al. (2010), this suggests reverse causality should not be an issue in our analysis of industry switching. Controlling for age and age-squared in every regression mitigates problems that could arise due to personality traits evolving over the life-cycle. In an ideal scenario, we would want strong instrumental variables to mitigate endogeneity problems. Whilst I have attempted to use value variables as instruments for personality, they make for relatively weak instruments. As discussed in Section 4.3, it is difficult to find suitable instruments for personality.

A final note on endogeneity concerns omitted variable bias. Following the mechanism described in Figure 22, one variable that we have not accounted for is ability. Whilst this is generally assumed to be unobservable, measures such as IQ tests and the Raven Progressive Matrix test have been used to shed some light on innate fluid intelligence (see Almlund et al., 2011). These measures are unfortunately not available in the data and therefore cannot be included in the regressions. Although intelligence is a separate concept to personality, the most obvious link between the two would be through Openness. One would expect intelligence to be positively correlated with Openness. However, since the precise interactions of all other variables with intelligence are unknown, it is difficult to make a confident assessment of the size or direction of the bias.

Finally, I look at an alternative to propensity score matching to test Hypothesis 4.3. The approach used is a variant of exact matching, due to Blackwell et al. (2009), known as Coarsened Exact Matching (CEM). CEM takes the same covariates that are used to generate the propensity score, but assigns observation into strata based on exact matching. Continuous variables, or variables with a large number of distinct values, are matched 'coarsely' based on a specified number of intervals. In our case, only age needs to be coarsened, as the remaining covariates are simply indicator variables. The disadvantage of this approach is that observations which cannot be matched exactly are lost. Additionally, each stratum does not necessarily contain equal numbers of British and German individuals. Therefore, I wrote a script in R (R Development Core Team, 2016) to randomly match individuals within each stratum and perform robust linear probability regressions.[23] This process was repeated 500 times, and means of regression coefficients and robust t-statistics were recorded. Results are shown in Table A6.

Due to the randomness of the matching process within a stratum and a reduced number of observations, results using this method were far weaker than when matching using the propensity score. The strongest personality difference influence from this method is in Openness, rather than Conscientiousness. Directions of estimates for these two factors are consistent between methods, but not for the remaining three. Determining which method best measures the impact of differences in personality across countries on outcomes is left for future work.

4.5 SUMMARY AND CONCLUSION

In this chapter, I offer a new perspective on the determinants of comparative advantage. Unlike long periods of history where production was primarily driven by natural resource availability, modern industry is largely service-oriented. Hence, personal characteristics are increasingly relevant.

The main findings are as follows. First, Big Five personality traits help to explain the probability of an individual working in a given occupation. Openness

23 The code is available upon request.

plays a particularly important role due to its influence on schooling choice and connection with ability.

Second, Big Five personality influences industry choice through educational choice, and predicts the likelihood of future career switching between industries. This suggests that personality is likely to be a determinant of an economy's industrial composition (though certainly not the strongest one). A one point increase in Agreeableness or Neuroticism increases the probability of switching industry in four years time by approximately 5 percentage points in Germany, for those below the mean age of 42. These influences are stronger when adjusted for measurement error. For example, the predictive power of a one point increase in Agreeableness on industry switching increases from approximately 5 percentage points to 15.6 percentage points.

Third, a one point increase in the personality difference between matched individuals from the UK and Germany leads to an increase in probability of working in different industries by 0.01. Hence, differences in personality are correlated with differences in industry composition. When separating for individual traits, the strongest personality influences come from differences in Conscientiousness and Neuroticism for those at least 40 years of age (who are more likely than younger individuals to be in equilibrium regarding their career choice).

Whilst traditional market forces and structural differences may be more successful in predicting relative industry differences between countries, this chapter finds that Big Five personality traits have a role to play alongside them. These findings represent a first step towards understanding the role of personality traits in explaining the differences in industrial activity across economies. Although more work needs to be done in this area, this chapter highlights the potential importance of personality for macroeconomic problems, and not just economic issues at the individual level.

A

APPENDICES

A.1 APPENDIX FOR CHAPTER 2

Script for survey 1

WHERE SHOULD SOCIETY FOCUS ITS EFFORTS TO IMPROVE WELLBEING?

We are interested in people's opinions on the quality of a society.

The UK government is collecting information on the four well-being questions on the following page. These measure happiness, satisfaction with life, how worthwhile life is, and people's anxiety. We would like to know your view on the relative importance of these for assessing how well a society is doing.

We would like you to imagine that you have 100 points to allocate as an indication of the importance of measures of well-being. How would you personally allocate the 100 points across the four measures below? [for example, if you believe all four are equally important, you would allocate 25% to each of the four measures]:

- Happiness – "Overall, how happy did you feel yesterday?": Personally I would allocate ...... % of my efforts to improving this.
- Satisfaction – "Overall, how satisfied are you with your life nowadays?": Personally I would allocate ...... % of my efforts to improving this.
- Worthwhile – "Overall, to what extent do you feel that your life is worthwhile?": Personally I would allocate ...... % of my efforts to improving this.
- Anxiety – "On a scale where nought is "Not at all anxious" and ten is "Completely anxious". Personally I would allocate ...... % of my efforts to improving this.

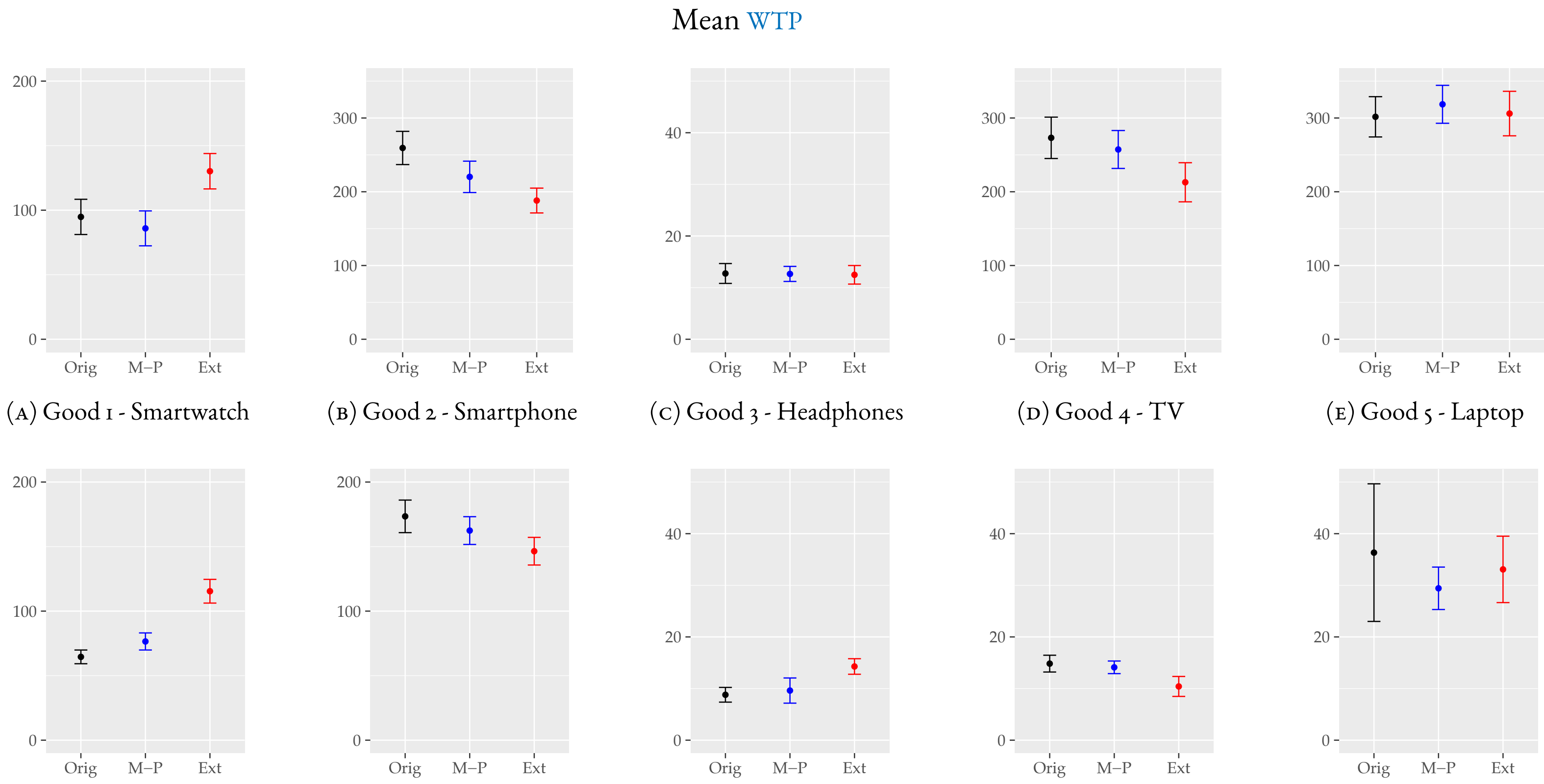
PLEASE REMEMBER THAT YOUR FOUR CHOSEN NUMBERS SHOULD ADD UP TO 100%. THANK YOU FOR YOUR VIEWS

A.2 APPENDIX FOR CHAPTER 3

TABLE A1: Summary of results for WTP, and accuracy of model predictions. Predicted orderings for each model can be found in Table 21.

	Result	Did model prediction match result?					
Good	Actual order (WTP)	Mean	μ_w	μ_w threshold	μ_w 1-star	μ_w 5-star	RF
1	e>o=m	Yes	No	Yes	No	No	Yes
2	o>m>e	No	Yes	No	No	No	No
3	o=m=e	No	No	No	No	No	No
4	o=m>e	Yes	Yes	No	Yes	Yes	Yes
5	o=m=e	No	No	No	No	No	No
6	e>m>o	No	Yes	Yes	No	Yes	Yes
7	o=m>e	Yes	No	Yes	No	No	No
8	e>o=m	Yes	No	No	Yes	No	No
9	o=m>e	Yes	Yes	No	No	Yes	Yes
10	o=m=e	No	No	No	No	No	No

Key:
= - the means of two treatments are not significantly different at the 5% level.
o - original review scores; m - mean preserving treatment; e - extreme treatment; WTP - (maximum) willingness to pay.
For example, "e>o=m" means the good has significantly higher mean WTP in the extreme treatment than the other two treatments (which are not significantly different from each other at the 5% level).



(a) Good 1 - Smartwatch (b) Good 2 - Smartphone (c) Good 3 - Headphones (d) Good 4 - TV (e) Good 5 - Laptop

(f) Good 6 - Hotel (g) Good 7 - Hotel (h) Good 8 - Book (i) Good 9 - Book (j) Good 10 - Restaurant

Figure A1: Mean of maximum willingness to pay for each good and treatment (original, mean-preserving, extreme), with 95% confidence intervals.

A.3 APPENDIX FOR CHAPTER 4

Table A2: Big Five items used in the BHPS and SOEP surveys.

I see myself as someone who:	Factor	Polarity
is sometimes rude to others	A	-
does a thorough job	C	+
is talkative	E	+
worries a lot	N	+
is original, comes up with ideas	O	+
has a forgiving nature	A	+
tends to be lazy	C	-
is outgoing, sociable	E	+
gets nervous easily	N	+
values artistic, aesthetic experiences	O	+
is considerate and kind	A	+
does things efficiently	C	+
is reserved	E	-
is relaxed, handles stress well	N	-
has an active imagination	O	+

TABLE A3: EIV regressions to test robustness of SOEP industry switching results from Table 31.

	Changed industry between 2005 and 2009?			
	Age < 42	Age ≥ 42	Age < 42	Age ≥ 42
Age	0.0424	0.142	0.0357	0.127
Age^2	-0.000717	-0.00123	-0.000581	-0.00103
Male	0.105**	-0.0432	0.123*	-0.122
Been married?	-0.00247	0.133	-0.0213	0.215
Agreeableness	0.175**	-0.328**	0.156**	-0.551***
Conscientiousness	-0.0297	0.0233	-0.00742	-0.121
Extraversion	0.00181	-0.203*	0.00145	-0.499***
Neuroticism	0.0557	-0.0653	0.0798*	-0.227*
Openness	0.0317	0.180	0.00517	0.636**
Satisfaction with income	-0.0362***	0.0144	-0.0393***	0.0487**
Gross wage in previous year	-4.41e-05**	-8.10e-05**	-5.37e-05**	-0.000146***
Political attitude	-	-	0.000846	0.0774**
Freq of sport activity (-ve)	-	-	0.000693	-0.0556
Freq of artistic activity (-ve)	-	-	-0.0291	0.235**
Optimism towards future (-ve)	-	-	-0.0820**	0.295***
Constant	-1.186*	-1.542	-0.835	-0.903
Observations	474	167	461	161
R^2	0.104	0.159	0.116	0.336

Standard errors are omitted for brevity. *** p<0.01, ** p<0.05, * p<0.1

Table A4: EIV regressions to test robustness of BHPS industry switching results from Tables 32 and 33.

	Changed industry between 2005 and 2008?			
	Age < 40	Age ≥ 40	Age < 40	Age ≥ 40
Age	-0.0557***	-0.0203	-0.113***	-0.0149
Age^2	0.000797***	0.000192	0.00177***	0.000132
Male	-0.00382	0.0282	0.00403	0.0188
Been married?	0.023	0.0111	0.0514	0.0364
Agreeableness	0.0224	0.0850*	0.0229	0.00642
Conscientiousness	-0.0521	-0.0843*	-0.044	-0.0364
Extraversion	0.0152	0.0156	0.0500*	0.0118
Neuroticism	0.00257	0.00446	0.0356*	9.56e-06
Openness	-0.000819	0.00933	0.00549	0.014
Labour income in prev yr	-9.27e-07**	-9.78e-07**	-2.42e-06*	-1.14e-06**
Satisfaction with pay	-0.0133**	-0.0205***	-0.00772	-0.0194***
Closest to Tory party?	-	-	0.0757**	0.0652**
Closest to Labour party?	-	-	0.0101	0.0286
Closest to Lib Dem party?	-	-	0.0648	0.0212
Belong to a social class?	-	-	0.000859	-0.00527
Member of trade union	-	-	-0.0772***	-0.0410*
Member of environmental group	-	-	0.113	0.131**
Member of parents association	-	-	-0.0285	0.00389
Member of tenants group	-	-	-0.00886	0.0738
Member of religious group	-	-	-0.00786	-0.0132
Member of voluntary service group	-	-	0.068	0.00437
Member of sports club	-	-	-0.0185	-0.0450*
Member of women's group	-	-	-0.398*	-0.00248
Would like own business?	-	-	0.0708**	0.0579
Constant	1.335***	0.675	1.734***	0.696
Observations	2,375	2,408	1,186	1,468
R^2	0.033	0.02	0.076	0.033

Standard errors are omitted for brevity. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE A5: EIV for propensity score matched regressions from Table 35.

	Dep variable: Is in diff industry?		
	Full	Age < 40	Age ≥ 40
Absolute diff in A	-0.00418	0.0156	-0.0286*
	(0.0111)	(0.0151)	(0.0162)
Absolute diff in C	-0.0102	0.00539	0.0265*
	(0.0106)	(0.0140)	(0.0153)
Absolute diff in E	-0.00469	0.00921	0.0116
	(0.00756)	(0.0103)	(0.0108)
Absolute diff in N	0.00220	-0.000584	0.0198**
	(0.00690)	(0.00935)	(0.00971)
Absolute diff in O	0.00129	0.0132	0.00621
	(0.00787)	(0.0107)	(0.0112)
Constant	0.835***	0.789***	0.767***
	(0.0189)	(0.0257)	(0.0259)
Observations	7,017	3,475	3,542
R^2	0.000	0.002	0.005

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

TABLE A6: Mean of 500 bootstrapped OLS coefficients using CEM.

	Dep variable: In different industries?		
	Full sample	Age < 40	Age ≥ 40
Absolute diff in A	0.00412	0.00180	0.00682
Absolute diff in C	0.00438	0.00584	0.00314
Absolute diff in E	0.00164	0.00224	0.00192
Absolute diff in N	-0.00076	-0.00223	0.00040
Absolute diff in O	-0.00080	0.00073	-0.00263
Constant	0.796***	0.798***	0.793***
Observations	3938	1825	2112

*** p<0.01, ** p<0.05, * p<0.1 (Robust s.e.)

BIBLIOGRAPHY

Abdi, Hervé (2007). "The Bonferonni and Šidák corrections for multiple comparisons." *Encyclopedia of Measurement and Statistics*. Vol. 3. Sage, pp. 103–107.

Akerlof, George A. and Rachel E. Kranton (2000). "Economics and identity." *Quarterly Journal of Economics*, pp. 715–753.

Allport, Gordon W. and Henry S. Odbert (1936). "Trait-names: a psycho-lexical study." *Psychological Monographs* 47.1, p. i.

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz (2011). "Personality psychology and economics." *Handbook of The Economics of Education*. Ed. by Stephen Machin Eric A. Hanushek and Ludger Woessmann. Vol. 4. Elsevier. Chap. 1, pp. 1–181.

Angrist, Joshua D. and Jörn-Steffen Pischke (2008). *Mostly harmless econometrics: an empiricist's companion*. Princeton University Press.

Artés, Joaquín, Maria del Mar Salinas-Jiménez, and Javier Salinas-Jiménez (2013). "Small Fish in a Big Pond or Big Fish in a Small Pond? The Effects of Educational Mismatch on Subjective Wellbeing." *Social Indicators Research*, pp. 1–19.

Avery, Christopher, Paul Resnick, and Richard Zeckhauser (1999). "The market for evaluations." *American Economic Review* 89.3, pp. 564–584.

Bauer, Thomas K. (2002). "Educational mismatch and wages: a panel analysis." *Economics of Education Review* 21.3, pp. 221–229.

Becker, Anke, Thomas Deckers, Thomas Dohmen, Armin Falk, and Fabian Kosse (2012). "The relationship between economic preferences and psychological personality measures." *Annual Review of Economics*. Ed. by K.J. Arrow and T.F. Bresnahan. Vol. 4.

Benet-Martínez, Verónica and Oliver P. John (1998). "Los cinco grandes across cultures and ethnic groups: multitrait-multimethod analyses of the Big Five in Spanish and English." *Journal of Personality and Social Psychology* 75.3, p. 729.

Benjamin, Daniel J., Miles S. Kimball, Ori Heffetz, and Alex Rees-Jones (2012). "What do you think would make you happier? What do you think you would choose?" *The American Economic Review* 102.5, p. 2083.

Bertrand, Marianne and Sendhil Mullainathan (2001). "Do people mean what they say? Implications for subjective survey data." *American Economic Review*, pp. 67–72.

Bibby, Peter A and Eamonn Ferguson (2011). "The ability to process emotional information predicts loss aversion." *Personality and Individual Differences* 51.3, pp. 263–266.

Blackwell, Matthew, Stefano M. Iacus, Gary King, and Giuseppe Porro (2009). "cem: Coarsened exact matching in Stata." *Stata Journal* 9.4, pp. 524–546.

Blanchflower, David G. and Andrew J. Oswald (2004a). "Money, sex and happiness: An empirical study." *The Scandinavian Journal of Economics* 106.3, pp. 393–415.

— (2004b). "Well-being over time in Britain and the USA." *Journal of Public Economics* 88.7-8, pp. 1359–1386.

— (2008). "Is well-being U-shaped over the life cycle?" *Social Science & Medicine* 66.8, pp. 1733–1749.

Borghans, Lex, Angela Lee Duckworth, James J. Heckman, and Bas Ter Weel (2008). "The economics and psychology of personality traits." *Journal of Human Resources* 43.4, pp. 972–1059.

Borjas, George J. (1987). "Self-selection and the earnings of immigrants." *American Economic Review* 77.4, pp. 531–553.

Bowles, Samuel, Herbert Gintis, and Melissa Osborne (2001). "The determinants of earnings: a behavioral approach." *Journal of Economic Literature* 39.4, pp. 1137–1176.

Boyce, Christopher J. and Alex M. Wood (2011). "Personality and the marginal utility of income: Personality interacts with increases in household income to determine life satisfaction." *Journal of Economic Behavior & Organization* 78.1, pp. 183–191.

Boyce, Christopher J., Alex M. Wood, and Gordon D. A. Brown (2010). "The dark side of conscientiousness: Conscientious people experience greater drops in life satisfaction following unemployment." *Journal of Research in Personality* 44.4, pp. 535–539.

Boyce, Christopher J., Alex M. Wood, Michael Daly, and Constantine Sedikides (2015). "Personality change following unemployment." *Journal of Applied Psychology* 100.4, p. 991.

Brickman, Philip and Donald T. Campbell (1971). "Hedonic relativism and planning the good society." *Adaption level theory: A symposium*. Ed. by M. H. Appley. New York: Academic Press, pp. 287–302.

Brim, Orville G. (1976). "Theories of the male mid-life crisis." *The Counseling Psychologist*.

Brown, Gordon D. A. (2013). "A model of authenticity." Unpublished.

Budría, Santiago and Ana I. Moro-Egido (2008). "Education, educational mismatch, and wage inequality: Evidence for Spain." *Economics of Education Review* 27.3, pp. 332–341.

Burnham, Kenneth P. and David R. Anderson (2004). "Multimodel inference - understanding AIC and BIC in model selection." *Sociological Methods & Research* 33.2, pp. 261–304.

Caliendo, Marco and Sabine Kopeinig (2008). "Some practical guidance for the implementation of propensity score matching." *Journal of Economic Surveys* 22.1, pp. 31–72.

Carstensen, Laura L. (2006). "The Influence of a Sense of Time on Human Development." *Science* 312.5782, pp. 1913–1915.

Chatman, Jennifer A. (1991). "Matching People and Organizations: Selection and Socialization in Public Accounting Firms." *Administrative Science Quarterly* 36.3, pp. 459–484.

Cheng, Terence Chai, Nattavudh Powdthavee, and Andrew J. Oswald (2015). "Longitudinal evidence for a midlife nadir in human well-being: results from four data sets." *The Economic Journal*.

Clark, Andrew E. (1996). "Job satisfaction in Britain." *British Journal of Industrial Relations* 34.2, pp. 189–217.

Clark, Andrew E. and Andrew J. Oswald (1994). "Unhappiness and Unemployment." *Economic Journal* 104.424, pp. 648–659.

— (1996). "Satisfaction and comparison income." *Journal of Public Economics* 61.3, pp. 359–381.

Clark, John Maurice (1918). "Economics and Modern Psychology: I." *The Journal of Political Economy* 26.1, pp. 1–30.

Cobb-Clark, Deborah A. and Stefanie Schurer (2012). "The stability of Big-Five personality traits." *Economics Letters* 115.1, pp. 11–15.

Coller, Maribeth and Melonie B. Williams (1999). "Eliciting individual discount rates." *Experimental Economics* 2.2, pp. 107–127.

Conti, Gabriella and James J. Heckman (2014). "Understanding conscientiousness across the life course: an economic perspective." *Developmental Psychology* 50.5, p. 1451.

Costa, Paul T. and Robert R. McCrae (1980). "Influence of extraversion and neuroticism on subjective well-being: happy and unhappy people." *Journal of Personality and Social Psychology* 38.4, pp. 668–678.

— (1992). *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO FFI): professional manual.* Psychological Assessment Resources.

Costa, Paul T., Antonio Terracciano, and Robert R. McCrae (2001). "Gender differences in personality traits across cultures: robust and surprising findings." *Journal of Personality and Social Psychology* 81.2, p. 322.

De Fruyt, Filip and Ivan Mervielde (1999). "RIASEC types and Big Five traits as predictors of employment status and nature of employment." *Personnel Psychology* 52.3, pp. 701–727.

DeNeve, Kristina M. (1999). "Happy as an extraverted clam? The role of personality for subjective well-being." *Current Directions in Psychological Science* 8.5, pp. 141–144.

DeNeve, Kristina M. and Harris Cooper (1998). "The Happy Personality: A Meta-Analysis of 137 Personality Traits and Subjective Well-Being." *Psychological Bulletin* 124.2, pp. 197–229.

DeYoung, Colin G., Jacob B. Hirsh, Matthew S. Shane, Xenophon Papademetris, Nallakkandi Rajeevan, and Jeremy R. Gray (2010). "Testing predictions from personality neuroscience brain structure and the big five." *Psychological Science*.

Diener, Ed, Robert A. Emmons, Randy J. Larsen, and Sharon Griffin (1985). "The satisfaction with life scale." *Journal of Personality Assessment* 49.1, pp. 71–75.

Diener, Ed and Richard E. Lucas (1999). "Personality and Subjective Well-Being." *Well-being: Foundations of hedonic psychology*. Ed. by Daniel Kahneman, Ed Diener, and Norbert Schwarz. Russell Sage Foundation, p. 213.

Diener, Ed, Richard E. Lucas, and Christie Napa Scollon (2006). "Beyond the hedonic treadmill: revising the adaptation theory of well-being." *American Psychologist* 61.4, p. 305.

Diener, Ed, Shigehiro Oishi, and Richard E. Lucas (2003). "Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life." *Annual Review of Psychology* 54.1, pp. 403–425.

Dohmen, Thomas (2014). "Behavioral labor economics: Advances and future directions." *Labour Economics* 30, pp. 71–85.

Dolan, Paul, Tessa Peasgood, and Mathew White (2008). "Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being." *Journal of Economic Psychology* 29.1, pp. 94–122.

Dornbusch, Rudiger, Stanley Fischer, and Paul Anthony Samuelson (1977). "Comparative advantage, trade, and payments in a Ricardian model with a continuum of goods." *The American Economic Review*, pp. 823–839.

Easterlin, Richard A. (2005). "Building a better theory of well-being." *Economics and Happiness: Framing the Analysis*. Ed. by Luigino Bruni and Pier Luigi Porta. Oxford University Press. Chap. 1, pp. 29–64.

— (2006). "Life cycle happiness and its sources: Intersections of psychology, economics, and demography." *Journal of Economic Psychology* 27.4, pp. 463–482.

Epps, T.W. and Kenneth J. Singleton (1986). "An omnibus test for the two-sample problem using the empirical characteristic function." *Journal of Statistical Computation and Simulation* 26.3-4, pp. 177–203.

Evans, Anthony M. and William Revelle (2008). "Survey and behavioral measurements of interpersonal trust." *Journal of Research in Personality* 42.6, pp. 1585–1593.

Ferrer-i-Carbonell, Ada (2005). "Income and well-being: an empirical analysis of the comparison income effect." *Journal of Public Economics* 89.5, pp. 997–1019.

Ferrer-i-Carbonell, Ada and Paul Frijters (2004). "How Important is Methodology for the estimates of the determinants of Happiness?" *The Economic Journal* 114.497, pp. 641–659.

Ferrer-i-Carbonell, Ada and John M. Gowdy (2007). "Environmental degradation and happiness." *Ecological Economics* 60.3, pp. 509–516.

Forman, Chris, Anindya Ghose, and Batia Wiesenfeld (2008). "Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets." *Information Systems Research* 19.3, pp. 291–313.

Frey, Bruno S. and Alois Stutzer (2002). "What can economists learn from happiness research?" *Journal of Economic Literature* 40.2, pp. 402–435.

— (2005). "Testing theories of happiness." *Economics and Happiness: Framing the Analysis*. Ed. by Luigino Bruni and Pier Luigi Porta. Oxford University Press. Chap. 4.

Frijters, Paul and Tony Beatton (2012). "The mystery of the U-shaped relationship between happiness and age." *Journal of Economic Behavior & Organization* 82.2, pp. 525–542.

Gardner, William L., Brian J. Reithel, Claudia C. Cogliser, Fred O. Walumbwa, and Richard T. Foley (2012). "Matching personality and organizational culture effects of recruitment strategy and the five-factor model on subjective person-organization fit." *Management Communication Quarterly* 26.4, pp. 585–622.

Gill, David and Victoria L. Prowse (2014). "Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis." *IZA Discussion Paper Series* 8236.

Goerg, Sebastian J. and Johannes Kaiser (2009). "Nonparametric testing of distributions—the Epps–Singleton two-sample test using the empirical characteristic function." *The Stata Journal* 9.3, pp. 454–465.

Goldberg, Lewis R. (1990). "An alternative "description of personality": the Big-Five factor structure." *Journal of Personality and Social Psychology* 59.6, p. 1216.

— (1992). "The development of markers for the Big-Five factor structure." *Psychological Assessment* 4.1, p. 26.

Goldberg, Lewis R., John A. Johnson, Herbert W. Eber, Robert Hogan, Michael C. Ashton, C. Robert Cloninger, and Harrison G. Gough (2006).

"The international personality item pool and the future of public-domain personality measures." *Journal of Research in personality* 40.1, pp. 84–96.

Green, Leonard, Joel Myerson, David Lichtman, Suzanne Rosen, and Astrid Fry (1996). "Temporal discounting in choice between delayed rewards: the role of age and income." *Psychology and Aging* 11.1, p. 79.

Heckman, James J., Lance J. Lochner, and Petra E. Todd (2003). "Fifty years of Mincer earnings regressions." *IZA Discussion paper series, No. 775*.

Heckman, James J. and Guilherme Sedlacek (1985). "Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market." *Journal of Political Economy* 93.6, pp. 1077–1125.

Heckman, James J., Jora Stixrud, and Sergio Urzua (2006). "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor Economics* 24.3, pp. 411–482.

Heineck, Guido and Silke Anger (2010). "The returns to cognitive abilities and personality traits in Germany." *Labour Economics* 17.3, pp. 535–546.

Higgins, E. Tory (1987). "Self-discrepancy: a theory relating self and affect." *Psychological Review* 94.3, p. 319.

Hoeffler, Steve and Dan Ariely (1999). "Constructing stable preferences: A look into dimensions of experience and their impact on preference stability." *Journal of Consumer Psychology* 8.2, pp. 113–139.

Holland, John L. (1997). *Making vocational choices: a theory of vocational personalities and work environments.* Psychological Assessment Resources.

Hu, Nan, Indranil Bose, Noi Sian Koh, and Ling Liu (2012). "Manipulation of online reviews: An analysis of ratings, readability, and sentiments." *Decision Support Systems* 52.3, pp. 674–684.

Hu, Nan, Ling Liu, and Jie Jennifer Zhang (2008). "Do online reviews affect product sales? The role of reviewer characteristics and temporal effects." *Information Technology and Management* 9.3, pp. 201–214.

Hu, Nan, Paul A. Pavlou, and Jennifer Zhang (2006). "Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word-of-mouth communication." *Proceedings of the 7th ACM conference on electronic commerce*, pp. 324–330.

Hu, Nan, Jie Zhang, and Paul A. Pavlou (2009). "Overcoming the J-shaped distribution of product reviews." *Communications of the ACM* 52.10, pp. 144–147.

Hunt, Kate Mathews (2015). "Gaming the system: Fake online reviews v. consumer law." *Computer Law & Security Review* 31.1, pp. 3–25.

Judge, Timothy A. and Daniel M. Cable (1997). "Applicant personality, organizational culture, and organization attraction." *Personnel Psychology* 50.2, pp. 359–394.

Judge, Timothy A., Chad A. Higgins, Carl J. Thoresen, and Murray R. Barrick (1999). "The Big Five personality traits, general mental ability, and career success across the life span." *Personnel Psychology* 52.3, pp. 621–652.

Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin (1997). "Back to Bentham? Explorations of experienced utility." *The Quarterly Journal of Economics*, pp. 375–405.

Kassenboehmer, Sonja C. and John P. Haisken-DeNew (2012). "Heresy or enlightenment? The well-being age U-shape effect is flat." *Economics Letters* 117.1, pp. 235–238.

Kimball, Miles and Robert Willis (2006). "Utility and Happiness."

Koh, Noi Sian, Nan Hu, and Eric K. Clemons (2010). "Do online reviews reflect a product's true perceived quality? An investigation of online movie reviews across cultures." *Electronic Commerce Research and Applications* 9.5, pp. 374–385.

Kremer, Ilan, Yishay Mansour, and Motty Perry (2014). "Implementing the "Wisdom of the Crowd"." *Journal of Political Economy* 122.5, pp. 988–1012.

Kristof-Brown, Amy L., Ryan D. Zimmerman, and Erin C. Johnson (2005). "Consequences of individuals' fit at work: a meta-analysis of person-job, person-organization, person-group, and person-supervisor fit." *Personnel Psychology* 58.2, pp. 281–342.

Larson, Lisa M., Patrick J. Rottinghaus, and Fred H. Borgen (2002). "Meta-analyses of Big Six interests and Big Five personality factors." *Journal of Vocational Behavior* 61.2, pp. 217–239.

Leamer, Edward E. (1995). *The Heckscher-Ohlin model in theory and practice*. International Finance Section, Department of Economics, Princeton University.

Leontief, Wassily (1971). "Theoretical assumptions and nonobserved facts." *American Economic Review* 61.1, pp. 1–7.

Leuven, Edwin and Barbara Sianesi (2014). "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing." *Statistical Software Components*.

Lim, Ee-Peng, Viet-An Nguyen, Nitin Jindal, Bing Liu, and Hady W. Lauw (2010). "Detecting product review spammers using rating behaviors." *Proceedings of the 19th ACM international conference on Information and Knowledge Management*, pp. 939–948.

Liu, Bing and Lei Zhang (2012). "A survey of opinion mining and sentiment analysis." *Mining text data*. Springer, pp. 415–463.

Löckenhoff, Corinna E. (2011). "Age, time, and decision making: from processing speed to global time horizons." *Annals of the New York Academy of Sciences* 1235.1, pp. 44–56.

Löckenhoff, Corinna E. and Joshua L. Rutt (2015). "Age Differences in Time Perception and Their Implications for Decision Making Across the Life Span." *Aging and Decision Making: Empirical and Applied Perspectives*. Ed. by Thomas M. Hess, JoNell Strough, and Corinna E. Löckenhoff. Academic Press, p. 213.

Luca, Michael (2011). "Reviews, reputation, and revenue: The case of Yelp.com." *Harvard Business School Working Paper* 12-016.

Lucas, Richard E. and M. Brent Donnellan (2011). "Personality development across the life span: longitudinal analyses with a national sample from Germany." *Journal of Personality and Social Psychology* 101.4, pp. 847–861.

Luhmann, Maike, Wilhelm Hofmann, Michael Eid, and Richard E. Lucas (2012). "Subjective well-being and adaptation to life events: a meta-analysis." *Journal of Personality and Social Psychology* 102.3, p. 592.

Mason, Winter and Siddharth Suri (2012). "Conducting behavioral research on Amazon's Mechanical Turk." *Behavior Research Methods* 44.1, pp. 1–23.

Mayzlin, Dina, Yaniv Dover, and Judith Chevalier (2014). "Promotional reviews: An empirical investigation of online review manipulation." *American Economic Review* 104.8, pp. 2421–2455.

McAuley, Julian, Rahul Pandey, and Jure Leskovec (2015). "Inferring networks of substitutable and complementary products." *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, pp. 785–794.

McAuley, Julian, Christopher Targett, Qinfeng Shi, and Anton van den Hengel (2015). "Image-based recommendations on styles and substitutes." *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, pp. 43–52.

McCrae, Robert R. and Paul T. Costa (1989). "Reinterpreting the Myers-Briggs Type Indicator From the Perspective of the Five-Factor Model of Personality." *Journal of Personality* 57.1, pp. 17–40.

McCrae, Robert R. and Oliver P. John (1992). "An introduction to the five-factor model and its applications." *Journal of Personality* 60.2, pp. 175–215.

McCrae, Robert R. and Antonio Terracciano (2005). "Universal features of personality traits from the observer's perspective: data from 50 cultures." *Journal of Personality and Social Psychology* 88.3, p. 547.

Mincer, Jacob A. (1974). *Schooling, Experience, and Earnings*. NBER Press.

Morgan, Jennifer Craft, Janette Dill, and Arne L. Kalleberg (2013). "The quality of healthcare jobs: can intrinsic rewards compensate for low extrinsic rewards?" *Work, Employment & Society* 27.5, pp. 802–822.

Mortensen, Dale T. and Christopher A. Pissarides (1994). "Job creation and job destruction in the theory of unemployment." *The Review of Economic Studies* 61.3, pp. 397–415.

Mudambi, Susan M. and David Schuff (2010). "What makes a helpful review? A study of customer reviews on Amazon.com." *MIS Quarterly* 34.1, pp. 185–200.

Mudambi, Susan M., David Schuff, and Zhewei Zhang (2014). "Why Aren't the Stars Aligned? An Analysis of Online Review Content and Star Ratings." *47th Hawaii International Conference on System Science*, pp. 3139–3147.

Mujcic, Redzo and Andrew J. Oswald (2016). "Evolution of Well-Being and Happiness After Increases in Consumption of Fruit and Vegetables." *American Journal of Public Health* 106.8, pp. 1504–1510.

Mukherjee, Arjun, Bing Liu, and Natalie Glance (2012). "Spotting fake reviewer groups in consumer reviews." *Proceedings of the 21st international conference on World Wide Web*. ACM, pp. 191–200.

Na, Jinkyung, Igor Grossmann, Michael E.W. Varnum, Shinobu Kitayama, Richard Gonzalez, and Richard E. Nisbett (2010). "Cultural differences are not always reducible to individual differences." *Proceedings of the National Academy of Sciences* 107.14, pp. 6192–6197.

Nettle, Daniel and Bethany Liddle (2008). "Agreeableness is related to social-cognitive, but not social-perceptual, theory of mind." *European Journal of Personality* 22.4, pp. 323–335.

Nyhus, Ellen K. and Empar Pons (2005). "The effects of personality on earnings." *Journal of Economic Psychology* 26.3, pp. 363–384.

— (2012). "Personality and the gender wage gap." *Applied Economics* 44.1, pp. 105–118.

O'Donnell, Gus and Andrew J. Oswald (2015). "National well-being policy and a weighted approach to human feelings." *Ecological Economics* 120, pp. 59–70.

Office for National Statistics. Social Survey Division (2014). *Annual Population Survey: Subjective Well-Being, April 2013 - March 2014* Colchester, Essex: UK Data Archive [distributor].

Oswald, Andrew J. (1997). "Happiness and economic performance." *The Economic Journal* 107.445, pp. 1815–1831.

Parducci, Allen (1965). "Category judgment: a range-frequency model." *Psychological Review* 72.6, p. 407.

— (1968). "The relativism of absolute judgments." *Scientific American*.

Parducci, Allen and Douglas H. Wedell (1986). "The category effect with rating scales: number of categories, number of stimuli, and method of presentation." *Journal of Experimental Psychology: Human Perception and Performance* 12.4, p. 496.

Park, Andreas and Hamid Sabourian (2011). "Herding and contrarian behavior in financial markets." *Econometrica* 79.4, pp. 973–1026.

Paulus, Martin P., Corianne Rogalsky, Alan Simmons, Justin S. Feinstein, and Murray B. Stein (2003). "Increased activation in the right insula during risk-taking decision making is related to harm avoidance and neuroticism." *Neuroimage* 19.4, pp. 1439–1448.

Pavot, William and Ed Diener (1993). "Review of the satisfaction with life scale." *Psychological Assessment* 5.2, p. 164.

Petrongolo, Barbara and Christopher A. Pissarides (2001). "Looking into the black box: A survey of the matching function." *Journal of Economic Literature*, pp. 390–431.

Polderman, Tinca J.C., Beben Benyamin, Christiaan A. De Leeuw, Patrick F. Sullivan, Arjen Van Bochoven, Peter M. Visscher, and Danielle Posthuma (2015). "Meta-analysis of the heritability of human traits based on fifty years of twin studies." *Nature Genetics*.

Poropat, Arthur E (2009). "A meta-analysis of the five-factor model of personality and academic performance." *Psychological Bulletin* 135.2, p. 322.

R Development Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Reimers, Stian, Elizabeth A. Maylor, Neil Stewart, and Nick Chater (2009). "Associations between a one-shot delay discounting measure and age, income, education and real-world impulsive behavior." *Personality and Individual Differences* 47.8, pp. 973–978.

Ricardo, David (1821). *On the principles of political economy and taxation*. Library of Economics and Liberty.

Riskey, Dwight R., Allen Parducci, and Gary K. Beauchamp (1979). "Effects of context in judgments of sweetness and pleasantness." *Perception & Psychophysics* 26.3, pp. 171–176.

Rosenbaum, Paul R. and Donald B. Rubin (1983). "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70.1, pp. 41–55.

Rothschild, Michael and Joseph E. Stiglitz (1970). "Increasing risk: I. A definition." *Journal of Economic Theory* 2.3, pp. 225–243.

Roy, A. D. (1951). "Some thoughts on the distribution of earnings." *Oxford Economic Papers* 3.2, pp. 135–146.

Rutledge, Robb B., Nikolina Skandali, Peter Dayan, and Raymond J. Dolan (2014). "A computational and neural model of momentary subjective wellbeing." *Proceedings of the National Academy of Sciences*.

Sackett, Paul R. and Philip T. Walmsley (2014). "Which Personality Attributes Are Most Important in the Workplace?" *Perspectives on Psychological Science* 9.5, pp. 538–551.

Schmitt, David P., Jüri Allik, Robert R. McCrae, and Verónica Benet-Martínez (2007). "The geographic distribution of Big Five personality traits patterns and profiles of human self-description across 56 nations." *Journal of Cross-Cultural Psychology* 38.2, pp. 173–212.

Schwandt, Hannes (2016). "Unmet Aspirations as an Explanations for the Age U-shape in Wellbeing." *Journal of Economic Behavior & Organization* 122, pp. 75–87.

Shimer, Robert (2007). "Mismatch." *American Economic Review*, pp. 1074–1101.

Simonson, Itamar and Amos Tversky (1992). "Choice in context: Tradeoff contrast and extremeness aversion." *Journal of marketing research* 29.3, p. 281.

Specht, Jule, Boris Egloff, and Stefan C. Schmukle (2011). "Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the Big Five." *Journal of Personality and Social Psychology* 101.4, pp. 862–882.

Stahl, Dale O. and Paul W. Wilson (1994). "Experimental evidence on players' models of other players." *Journal of Economic Behavior & Organization* 25.3, pp. 309–327.

Steel, Piers, Joseph Schmidt, and Jonas Shultz (2008). "Refining the relationship between personality and subjective well-being." *Psychological Bulletin* 134.1, p. 138.

Stevenson, Betsey and Justin Wolfers (2008). "Economic growth and subjective well-being: Reassessing the Easterlin paradox." *Brookings Papers on Economic Activity*, pp. 1–87.

— (2013). "Subjective well-being and income: Is there any evidence of satiation?" *American Economic Review* 103.3, pp. 598–604.

Terracciano, Antonio, A.M. Abdel-Khalek, N. Adam, L. Adamovova, C-k. Ahn, H-n. Ahn, B.M. Alansari, L. Alcalay, J. Allik, Alois Angleitner, et al. (2005). "National character does not reflect mean personality trait levels in 49 cultures." *Science* 310.5745, pp. 96–100.

Thaler, Richard (2016). "Behavioral Economics: Past, Present, and Future." *American Economic Review* 106.7, pp. 1577–1600.

Tripp, James and Gordon D. A. Brown (2016). "Being Paid Relatively Well Most of the Time: Negatively Skewed Payments are more Satisfying." *Working Paper*.

University of Essex Institute for Social and Economic Research (2010). "British Household Panel Survey: Waves 1-18, 1991-2009 [computer file]. 7th Edition." *UK Data Archive*.

Uysal, Selver Derya and Winfried Pohlmeier (2011). "Unemployment duration and personality." *Journal of Economic Psychology* 32.6, pp. 980–992.

Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp (2007). "The German Socio-Economic Panel study (SOEP)-evolution, scope and enhancements." *Schmollers Jahrbuch* 127.1, pp. 139–169.

Wang, Hongning, Yue Lu, and Chengxiang Zhai (2010). "Latent aspect rating analysis on review text data: a rating regression approach." *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 783–792.

Weiss, Alexander, James E. King, Miho Inoue-Murayama, Tetsuro Matsuzawa, and Andrew J. Oswald (2012). "Evidence for a midlife crisis in great apes consistent with the U-shape in human well-being." *Proceedings of the National Academy of Sciences* 109.49, pp. 19949–19952.

Willis, Robert J and Sherwin Rosen (1979). "Education and self-selection." *Journal of Political Economy* 87.5 pt 2.

Winkelmann, Liliana and Rainer Winkelmann (2008). "Personality, work, and satisfaction: evidence from the German Socio-Economic Panel." *The Journal of Positive Psychology* 3.4, pp. 266–275.

Woods, Stephen A. and Sarah E. Hampson (2010). "Predicting adult occupational environments from gender and childhood personality traits." *Journal of Applied Psychology* 95.6, p. 1045.

Yong, An Gie and Sean Pearce (2013). "A beginner's guide to factor analysis: Focusing on exploratory factor analysis." *Tutorials in Quantitative Methods for Psychology* 9.2, pp. 79–94.