The Scandinavian
Journal of Economics

Scand. J. of Economics 125(4), 823–859, 2023
DOI: 10.1111/sjoe.12538

Experienced versus decision utility: large-scale comparison for income–leisure preferences*

Alpaslan Akay†

University of Gothenburg, SE-40530 Gothenburg, Sweden
alpaslan.akay@gu.se

Olivier B. Bargain‡

Bordeaux School of Economics, FR-33608 Pessac, France
olivier.bargain@u-bordeaux.fr

H. Xavier Jara

London School of Economics and Political Science, London, WC2A 2AE, UK
h.x.jara-tamayo@lse.ac.uk

Abstract

Subjective well-being (SWB) data are increasingly used to perform welfare analysis. Interpreted as "experienced utility", it has recently been compared to "decision utility" using small-scale experiments most often based on stated preferences. We transpose this comparison to the framework of non-experimental and large-scale data commonly used for policy analysis, focusing on the income–leisure domain where redistributive policies operate. Using the British Household Panel Survey, we suggest a "deviation" measure, which is simply the difference between actual working hours and SWB-maximizing hours. We show that about three-quarters of individuals make decisions that are not inconsistent with maximizing their SWB. We discuss the potential channels that explain the lack of optimization when deviations are significantly large. We find proxies for a number of individual and external constraints, and show that constraints alone can explain more than half of the deviations. In our context, deviations partly reflect the inability of the revealed preference approach to account for labor market rigidities, so the actual and SWB-maximizing hours should be used in a complementary manner. The suggested approach based on our deviation metric could help identify labor market frictions.

Keywords: Decision utility; experienced utility; labor supply; subjective well-being

JEL classification: *C*90; *I*31; *J*22

*A. Akay acknowledges visiting financial support from the COV-POP project, funded by the Nouvelle Aquitaine region, which has allowed completing the revision of this paper. We acknowledge financial support from the Chaire BEWELL and the GPR HOPE, funded by the "Investments for the Future" program (PIA) IdEx Université de Bordeaux. We are grateful to discussants and participants at the ASSA meeting, the RES Conference, and various seminars (ISER, DIAL, LISER, AMSE, IZA, BSE) for useful suggestions. Any errors remain our own.
†Also affiliated with Universidad Antonio de Nebrija, Madrid, Spain.
‡Also affiliated with IZA.

1. Introduction

In economics, the standard approach to measure well-being relies on the observation of decisions made by supposedly rational (utility-maximizing) agents. The object derived from the "revealed preference" approach is sometimes referred to as a "decision utility". For more than two decades, some authors have claimed that this decision utility is not always consistent with the well-being associated with different experiences. They recommend developing measures that focus more directly on "experienced utility" (e.g., Dolan and Kahneman, 2008), such as self-reported information on happiness, life satisfaction, or mental health. A growing amount of evidence has shown that such subjective well-being (SWB) information is not pure statistical noise: it reflects some individual heterogeneity that is closely associated with objective measures of well-being and, to some extent, with behavior.[1] Yet, SWB is still seen by many as one argument, among others, in the grand utility function of an individual (Rayo and Becker, 2007; Benjamin et al., 2012; Glaeser et al., 2016).[2] Other studies postulate that SWB answers, commonly provided in survey questionnaires, are consistent with people's revealed preferences (Oswald and Wu, 2010; Decancq et al., 2015).

Whether there is congruence between individual decisions and the SWB derived from these choices is still an open question. This is especially disputable for key economic decisions (such as labor supply), which imply a trade-off between several important dimensions of a good life (e.g., consumption versus leisure). On the one hand, observed choices may reflect heuristics, optimization errors, or the fact that people have imperfect information about what is good for them. Choices are also potentially limited by many personal constraints (e.g., family obligations) and external factors (e.g., market imperfections), the importance of which is difficult to assess in welfare analyses. On the other hand, SWB may not encompass the totality of what humans are trying to achieve when they make decisions. Individual choices may reflect other life goals (e.g., fame) or values (e.g., helping others) that partly differ from, or sometimes conflict with, the pursuit of well-being as we measure it in subjective data. Despite these sources of discrepancy, it seems crucial to test whether there is (at least) minimal consistency between decision utility and experienced utility.

[1] See Krueger and Schkade (2008) and Oswald and Wu (2010), as well as critical reviews in Senik (2005), Clark et al. (2008), Kahneman and Krueger (2006), and Fleurbaey and Blanchet (2013).

[2] Kőszegi and Rabin (2008) argue that both subjective and choice-based measures of well-being contain unique information on a person's true welfare, so that the ideal measure should perhaps combine both types of data.

This paper proposes a tangible approach to address this question in the context of labor supply decisions. Rather than confronting the ordinal preferences consistent with a decision-based versus experience-based welfare metric,[3] we directly compare actual working hours (consistent with decision-utility maximization) and optimal working hours (from the perspective of experienced utility; i.e., hours that maximize income–leisure satisfaction). The comparison is done on a large scale using nationally representative data, from the British Household Panel Survey (BHPS). We necessarily focus on single people, because the joint decision in couples is difficult to apprehend for individual welfare comparisons. Our single-value "deviation" metric is a practical and convenient representation of the potential discrepancies between decision and experienced utilities, which can be used for inference and for exploring the determinants of these discrepancies – here in the context of labor supply or, more generally, in analyses that traditionally rely on the revealed preferences approach.

The suggested procedure goes as follows. We start by calculating the distribution of deviations in the sample. To this end, we combine income and leisure satisfaction domains to construct a proxy of experienced utility in the income–leisure domain. We use this SWB measure to estimate an experienced utility function, adopting a flexible approach borrowed from the labor supply literature. Using the estimated parameters and discretized income–leisure bundles, we numerically search for the amount of working hours that would maximize experienced utility and compare them with actual choices. We find a broad overlap between actual work hours and SWB-maximizing work duration. The average deviation is close to zero (−2.9 hours). The negative sign implies that people "overwork" on average according to SWB maximization, but the deviation is not significantly different from zero for 72 percent of the individuals in the sample. In other words, for a majority of people, actual decisions are not inconsistent with the maximization of their income–leisure satisfaction. We then attempt to describe the large discrepancies observed for specific population subgroups (e.g., by gender, family composition, region of residence, etc.), either characterized as "overworking" (a negative mean deviation) or "underworking" (a positive mean deviation) from an SWB perspective. Results suggest intuitive patterns regarding the direction of the deviations. For instance, those living in London significantly overwork (suggesting social norms or labor market constraints) while those with children tend to work too little (suggesting childcare constraints or labor contracts that are not flexible enough). A detailed analysis by levels of worked hours suggests that significant deviations are primarily due to those at the two ends

[3] This alternative approach is used in Akay et al. (2020) where money metrics are derived from ordinal preferences consistent with either decisions or subjective experience.

of the hours distribution (i.e., those out of work or engaged in overtime). We discuss the broad varieties of factors that can explain discrepancies: constraints, optimization errors, and non-hedonic life goals. The presence of constraints seems to be the prominent explanation in the labor supply context, as suggested by simple regressions of deviations on a broad set of variables associated with individual constraints (e.g., poor health, family obligations) and labor market constraints (e.g., high local unemployment). The proxies for these constraints, as identified from the survey, can explain more than half of the variance in individual deviations. These results are robust to alternative measures of experienced utility, alternative functional forms for experienced utility functions, alternative sample selection (e.g., adding job-seekers and the self-employed), the modeling of individual heterogeneity in SWB levels (either proxied by personality traits or panel data fixed effects), or the addition of heterogeneity in preferences for leisure (alternative sets of "taste-shifters" in work preferences).

The present exercise makes several contributions. First, comparison between experienced and decision utility remains rare in the literature. Small-scale experiments in behavioral economics or psychology have greatly contributed to explain some of the difference between subjective and revealed preferences (e.g., Kahneman and Thaler, 2006), notably in the field of public good valuation (Kahneman and Sugden, 2005). The present work is an original attempt to transpose this comparison in large-scale and non-experimental surveys, which are commonly available and used for policy analysis. In this way, it is very complementary to Benjamin et al. (2012) or earlier experiments (Kahneman et al., 1997). While we cannot experimentally control and manipulate the parameters that possibly explain why people do not maximize SWB, we show how to take advantage of a rich household panel survey to pinpoint a set of factors that could potentially explain discrepancies related to constraints.[4]

Second, our approach is different from the first large-scale comparison suggested by Benjamin et al. (2012), who proxy experienced utility with SWB (as we do) but elicit decision utility using "stated" preferences.[5] In the present study, we consider actual decisions rather than hypothetical life scenarios

[4]Our work also relates to studies that use panel data to check people's expectations regarding the future implications of their current choices or of major life events (Dolan and Kahneman, 2008; Frijters et al., 2009; Odermatt and Stutzer, 2019), or their ability to adjust behavior when reported SWB indicates that actual choices are suboptimal (Clark et al., 1998; Frijters, 2000).

[5]In their application, people are asked to decide between virtual jobs with different work hours–earnings bundles. Other recent studies also use hypothetical situations. For instance, Clark et al. (2015) elicit the relative weights placed by people on their own income versus on others' income. Benjamin et al. (2014a) evaluate the trade-offs between a large set of potential well-being measures.

underlying stated preferences, which bring the decision–experienced utility comparison closer to the context traditionally used for policy and welfare analyses.[6]

Third, we focus on a relevant domain for that purpose. Indeed, even though we focus on two dimensions only, the income–leisure domain is crucial for welfare analyses as this is where second-best redistributive policies operate.[7]

Fourth, our contribution combines different perspectives. Methodologically, we propose a practical way to measure the degree of congruence between decision and experienced utility by means of a deviation metric. Conceptually, we discuss the different channels that can generate large discrepancies, in particular variables related to individual and external constraints. Quantitatively, survey data can explain a reasonable amount of deviations between actual and SWB-maximizing choices.

Finally, our analysis highlights that experienced utility supplies complementary information that can help with investigating the shortcomings of the revealed preference approach, at least in the context of labor supply decisions. We derive implications for future research, notably the fact that deviations derived from SWB could provide a new way to characterize labor market frictions in the labor supply context. Also, the strategy employed here can be extended to investigate deviations in other economic areas that rely on the revealed preference approach (e.g., transportation or consumption studies).

The rest of the paper is organized as follows. In the next section, we present the data, sample selection, and our empirical approach. In Section 3, we present the results in terms of mean deviations, as well as heterogeneity across subgroups, a discussion on the potential channels explaining deviations, and extensive robustness checks. Finally, we conclude in Section 4 by deriving the methodological and welfare policy implications of our results.

[6]Other studies also consider actual choices: Benjamin et al. (2014b) and Glaeser et al. (2016) with residency choices, Fleurbaey and Schwandt (2015) for a whole set of decisions that can potentially affect SWB, and Perez-Truglia (2015) for consumption decisions. Frijters (2000) investigates whether a low satisfaction level in a particular area is correlated with the plan to change current conditions in that area. Clark et al. (1998) find that a lower job satisfaction level (slightly) increases the chances of quitting in the future.

[7]In Akay et al. (2020), we suggest a related approach to discuss the implications of using different types of preference elicitation methods for welfare analysis. We estimate ordinal preferences, consistent with either actual choices or income–leisure satisfaction, in order to compute equivalent incomes from the “fair allocation theory” (Fleurbaey and Maniquet, 2006) in both cases, and we characterize how welfare ranks change when moving from one set of preferences to the other. This analysis is more normative as conclusions depend on ethical priors regarding the degree of individual responsibility upon work aversion.

2. Data and empirical framework

2.1. Data and sample selection

2.1.1. Data. Our analysis is based on data from the BHPS, a large-scale nationally representative survey collected in the United Kingdom between 1991 and 2008. It contains information on labor market status and different domains of satisfaction (overall life satisfaction, income and leisure satisfaction) since 1996. This dataset also provides standard information on individual and household characteristics (gender, age, education, health, psychological traits) as well as regional characteristics that are used in our empirical analysis. As the SWB information is missing for the years 2006–2007, we focus on the period 1996–2005.[8]

2.1.2. Sample selection. In order to compare decision and experienced utilities in a non-experimental context, we necessarily restrict our analysis to single individuals. For individuals living in a couple, comparing their actual working hours to SWB-maximizing hours would be much more complex for several reasons. First, their individual SWB measure, constructed as a combination of income and leisure satisfactions (see below), might be interpreted differently than for singles, especially if, when answering the income satisfaction question, each partner expresses their satisfaction about the household total budget rather than referring to the resources available to them in the household. Second, a person's income–leisure satisfaction would then be estimated on income and leisure variables, but only leisure is individual while income corresponds to total household resources. Indeed, the level of resources accruing to each adult is not observed and is very hard to estimate, as discussed in the literature on collective models of labor supply with nonlinear taxation (see Chiappori and Donni, 2011). Third, the underlying model would be even more complicated as the optimal work duration of a person would depend on their spouse's working hours, so that SWB equations for both spouses should be estimated jointly while accounting for an implicit household optimization mechanism. Finally, the reasons discussed above also mean that the interpretation of SWB-maximizing hours – and thus the interpretation of our "deviation" metric – would be very different than for singles.[9]

We also focus on employed or voluntarily inactive workers in our baseline sample. Indeed, we necessarily apply the same logic as in labor supply models

[8] See https://www.iser.essex.ac.uk/bhps for more detail on the dataset.

[9] Further work should explore ways to include couples in the analysis, addressing each of the challenges outlined above, but it is likely that further progress in modeling collective labor supply is needed first.

(see van Soest, 1995) as we must assume that income–leisure satisfactions result from a trade-off between consumption and free time. People who are not able to arbitrate between these dimensions should show larger deviations between actual and SWB-maximizing hours than the average. Thus, we exclude people who appear as fully rationed from the labor market, using a standard definition of job-seekers,[10] and those not available for work (disabled individuals, full-time students, and pensioners). We retain other inactive people (i.e., those who "voluntarily" choose to be out of work, for example, for childcare or other activities). The self-employed represent a specific population, with labor supply decisions that may considerably differ from those of salaried workers. Also, in their case, information on worked hours and income may be more prone to measurement errors or misreporting.[11] For these reasons, we do not include them in the baseline sample. Yet, we suggest robustness checks where we re-incorporate job-seekers and self-employed in the analysis, increasing the external validity of our demonstration. Finally, we only retain individuals for whom all key characteristics (including socio-demographics) are available for all years. Our selected sample includes 5,501 person × year observations.

2.2. Set-up and measures

2.2.1. Key variables. The key variables for our analysis are leisure (or, equivalently, working hours) and disposable income. Weekly working hours drawn from the data are denoted h_{it} for an individual i at time t. Assuming a maximum working time of 80 hours per week, we normalize leisure time as the residual, namely $l_{it} = 80 - h_{it}$. Disposable income of an individual y_{it} is calculated as

$$y_{it} = G_t(w_{it}h_{it}, \mu_{it}, \zeta_{it}), \tag{1}$$

using reported gross labor income $w_{it}h_{it}$ (hourly wage rates w_{it} × weekly work hours h_{it}), unearned income μ_{it}, and a set of individual characteristics ζ_{it}.[12] Function G_t represents the aggregation of all incomes and the imputation of taxes and benefits, using numerical simulations of tax-and-benefit rules of each period $t = 1, \ldots, T$. The set ζ_{it} represents individual characteristics that matter for tax-and-benefit calculations and are extracted from the data, for

[10] They answer negatively to at least one of the following questions: "have you actively looked for a job within the last four weeks?" and "are you ready to take up a job within the next two weeks?".

[11] For a specific study on entrepreneurs and how their expected life satisfaction deviates from future life satisfaction, see Odermatt et al. (2021).

[12] Unearned income refers to income not derived from labor such as capital income, property income, rents, and private transfers, etc.

instance the presence of children (which conditions the calculation of child benefits, increment of income support, tax credits, etc.).[13]

2.2.2. Measures of experienced utility. In order to predict SWB-maximizing work hours, we must first compute an individual SWB measure focusing on income and leisure dimensions. We denote V_{it}^{E} such an experienced utility of income and leisure for individual i at period t. Our data contain satisfaction on life domains including income and leisure, which can be combined for our purpose (see also van Praag et al., 2003). We use the questions "How dissatisfied or satisfied are you with the income of your household/with the amount of leisure time you have?". The answers, measured on an ordered scale between 1 ("not satisfied at all") and 7 ("completely satisfied"), are denoted S_{it}^{y} for income satisfaction and S_{it}^{l} for leisure satisfaction. To obtain a proxy for the experienced utility V_{it}^{E}, we need to combine these domains of satisfaction into a single measure. Yet the relative weight to be put on each of these domains is unknown. Thus, we use the overall life satisfaction question, with the answer S_{it} recorded on a similar 1–7 scale, to infer these weights. We simply estimate

$$S_{it} = \gamma^{y} S_{it}^{y} + \gamma^{l} S_{it}^{l} + e_{it}, \tag{2}$$

and use the estimated coefficients as weights on each domain to compute the experienced utility $V_{it}^{E} = \widehat{\gamma}^{y} S_{it}^{y} + \widehat{\gamma}^{l} S_{it}^{l}$. It turns out that the two dimensions play a relatively balanced role, as we find that $\widehat{\gamma}^{y}/(\widehat{\gamma}^{y} + \widehat{\gamma}^{l}) = 0.468$. This combined – or concentrated – income–leisure satisfaction measure, extracted from overall life satisfaction, is our baseline proxy for experienced utility, but alternative approaches will be suggested in the robustness checks.

2.3. A structural subjective well-being estimation

To calculate deviations between actual and SWB-maximizing hours, we estimate SWB on income and leisure plus other covariates. Given that this empirical model is then used to predict "optimal" hours in terms of SWB, it must be specified in a relatively more structural way than usual SWB equations

[13] For hourly wage rates, we follow a fairly standard approach: that is, we calculate them as weekly earnings divided by worked hours for workers, then use this information to estimate Heckman-corrected wage equation (instruments are non-labor income and the presence of children aged 0–2) in order to predict a wage rate w_{it} for non-workers. We assume that gross hourly wage rates do not depend on working duration. This assumption is standard (but sometimes relaxed, for instance in Ilmakunnas and Pudney, 1990). In general, when wages are determined by collective bargaining within branches or sectors, discrimination between full-time and part-time workers is less likely to occur.

(i.e., we impose some structure similar to the one used in labor supply models). At the same time, we condition SWB on additional determinants of well-being in order to "clean" the potential noise inherited by subjective measures and following the recommendations in the literature.[14]

2.3.1. Functional form. First, we assume that V^E can be modeled as a deterministic function $U_{it}^E(y_{it}, l_{it}; x_{it})$ of income y_{it} and leisure l_{it}. Several sources of heterogeneity enter the model. The deterministic utility is conditioned on a vector x_{it} of heterogeneity in terms of underlying income–leisure preferences. Additional controls z_{it} and α_i account for individual observed and unobserved heterogeneity in reported levels of well-being. The model is written as

$$V_{it}^E = U_{it}^E\left(y_{it}, l_{it}; x_{it}\right) + \lambda' z_{it} + \alpha_i + \epsilon_{it}. \tag{3}$$

For the deterministic part, $U_{it}^E(y_{it}, l_{it}; x_{it})$, note that relatively simple functional forms are usually employed in the SWB literature, for example, empirical models are usually linear, or log-linear, in income to capture the concave relationship with well-being (cf. Clark et al., 2008). Few empirical studies add leisure (or working hours) as we do.[15] Because our model must come close to the structure of labor supply models, we suggest a relatively flexible functional form for our baseline estimations, namely a quadratic form in income and leisure with an interaction term (Blundell et al., 2000):

$$U_{it}^E(y_{it}, l_{it}; x_{it}) = \beta_{yy} y_{it}^2 + \beta_{ll} l_{it}^2 + \beta_y y_{it} + \beta_l(x_{it}) l_{it} + \beta_{yl} y_{it} l_{it}. \tag{4}$$

Preference heterogeneity is accounted for by linear variation in the leisure coefficient:

$$\beta_l(x_{it}) = \beta_{l,0} + \beta_{l,1}' x_{it}. \tag{5}$$

In the baseline, the vector x_{it} is composed of individual characteristics that possibly influence work preferences. For simplicity, we use binary variables in x_{it} including male, age above 40, presence of children, and living in London. To allow further heterogeneity, we also introduce personality traits. Among the "big five", we select conscientiousness and neuroticism as they are shown to be those that matter the most for labor supply choices (see

[14] Several authors have insisted on the necessity to purge individual SWB measures from idiosyncratic variation in well-being responses and individual-specific circumstances, in order to recover a meaningful preference structure (see Decancq et al., 2015).
[15] An exception is Knabe and Rätzel (2010) who use a log form on income and a linear or quadratic form for leisure, without interaction terms.

Wichert and Pohlmeier, 20).[16] We include dummies indicating above-average conscientiousness and neuroticism. In robustness checks, we will present alternative specifications, for instance using continuous (rather than binary) taste-shifters in x_{it} or including the full set of big five personality traits.

2.3.2. Additive and stochastic terms. Experienced utility based on SWB measures can reflect individual heterogeneity in the way people perceive and/or report levels of leisure and income satisfactions. This makes it more difficult to assume interpersonal comparability in SWB responses when our aim is to extract subjective preferences on income and leisure. To "clean" SWB measures, however, we can model heterogeneity in SWB levels through the additive shift represented by $\lambda' z_{it} + \alpha_i$ in equation (3). The first term z_{it} is a vector of the usual determinants of well-being found in the literature (cf. Clark et al., 2008). The second, α_i, corresponds to time-invariant unobserved heterogeneity. It can be proxied in several ways. In our baseline approach, we rely on the complete set of personality traits (the big five on a 1–4 scale). These traits are usually seen as capturing a large part of the time-invariant unobserved heterogeneity in SWB (Ravallion and Lokshin, 2001; Boyce, 2010). Residuals ϵ_{it} are independent and identically distributed (i.i.d.) and normally distributed error terms so the model can be estimated by standard linear estimation methods on pooled years; in robustness checks, maximum likelihood is used when nonlinear specifications of U^E_{it}, such as Box–Cox, are tried. We will also examine alternative modeling of α_i including quasi-fixed effects following Mundlak (1978) and fixed effects in panel estimations.

2.3.3. Identification. The estimation of the β parameters, interpreted as underlying preferences, may be biased due to omitted variables. This will be the case if actual unobserved heterogeneity in work preferences (e.g., to be morally obliged to work a lot to support the family or, inversely, to stay home to care for a sick parent) is correlated with other unobserved determinants of well-being (e.g., experiencing stress due to moral obligations). Two modeling choices tend to reduce these concerns and support the identification of the model. First, we account for individual heterogeneity – notably in the form of relevant personality traits – both in work preference parameters through x_{it} and in (separately additive) well-being terms z_{it}. Second, as used in the labor supply literature (Blundell et al., 1998), we avail of spatial and temporal

[16]Neuroticism is a fundamental personality trait in the study of psychology characterized by anxiety, fear, moodiness, worry, envy, frustration, jealousy, and loneliness. Conscientiousness is the personality trait of being thorough, careful, or vigilant, implying the desire to do a task well.

variation in net wages due to variation in tax-and-benefit rules in function G. In particular, when pooling different years of data, the same individual may not make the same labor supply choice because they face different work incentives due to different tax-and-benefit schedules (i.e., different functions G_t) over the periods $t = 1, \dots, 10$.[17] These approaches are the best we can do in the present setting but we cannot exclude that some biases remain.

2.4. Construction of the deviation metric

Our approach focuses on a direct comparison between actual hours (consistent with decision utility) and optimal hours (in the perspective of SWB-maximization). The deviation between these measures can be seen as a "projection error" in the sense of Loewenstein et al. (2003) and Loewenstein and Adler (1995), but that would entail a particular interpretation whereby SWB-maximizing errors represent failures of individuals to decide optimally according to their genuine preferences. More generally, deviations cannot be taken *prima facie* as errors if people face some types of constraints (due to health, family, labor market rigidities, social norms, etc.) or pursue other goals than maximizing their short-term SWB (which we can refer to as non-hedonistic objectives, by simplification).

Our statistic of interest is a deviation D_{it} defined, for each individual i at time t, as

$$D_{it} = h_{it} - h^*_{it}, \tag{6}$$

namely the observed actual working hours h_{it} minus the experienced utility maximizing hours h^*_{it} formally defined as

$$h^*_{it} = \arg\max_{h_{it}} U^E_{it}\left(G_t(w_{it}h_{it}, \mu_{it}, \zeta_{it}), 80 - h_{it}; x_{it}\right). \tag{7}$$

In practice, we first estimate the model described by equations (3)–(5) and obtain the parameters of the deterministic part of the experienced utility function U^E_{it}. Thus, we can calculate h^*_{it} by means of numerical optimization of a discrete version of the model.[18] To investigate statistical significance of the

[17] By pooling 10 years of data, we obtain much variation in the UK tax-and-benefit schedules, compounded with spatial variation (e.g., council taxes are specific to England, Scotland, Wales and Northern Ireland). Indeed, the UK system has experienced deep changes over the years under study, notably with the important reforms undertaken by the "New Labour" government regarding income tax, social insurance contributions, council taxes, income support, and tax credits for working poor families; an extensive description of these reforms can be found in Blundell et al. (2000) and Adam and Browne (2010).

[18] First, an agent i at period t is assumed to face J income–labor pairs, denoted (y_{ijt}, h_{ijt}), $j = 1, \dots, J$. In the baseline, we opt for $J = 7$ discrete options corresponding to weekly work

estimated D_{it}, the standard errors are calculated using bootstrap, which goes as follows. We first draw $R = 200$ random bootstrap samples from our overall dataset and estimate the model described by equations (3)–(5) repeatedly. Then, we calculate the bootstrapped standard error of D_{it} for each individual i and period t.

3. Results

We briefly discuss the estimation of the experienced utility function. We then move to the overall distribution of deviations D_{it} and analyze the heterogeneity in deviations with respect to observed individual characteristics and for different levels of working hours. Next, we provide a discussion on the potential explanations for large deviations and attempt to measure the extent to which they are associated with individual and external constraints that might hinder choices. Finally, we present an extensive robustness analysis in terms of sample selection, SWB definition/measure, preference heterogeneity, treatment of the unobserved heterogeneity, and estimation methods.

3.1. Estimation results

Baseline estimations of experienced utility, used to calculate deviations, are presented in model I in Table A1 in the Appendix. We only report the estimates of the deterministic part U_{it}^E because we are mainly interested in the respective roles of income and leisure in the variation of SWB between individuals. As expected, we observe a significant, increasing, and concave effect of income on income–leisure satisfaction. Results for leisure are less clear, and most coefficients are insignificant, but this is because many leisure terms enter the model. If we restrict the deterministic utility to a simple quadratic form without interaction and taste-shifters on leisure, we find that both leisure terms are significant, as shown in model II. Leisure has a positive and concave effect in this case. If we add taste-shifters, in model III, we do not reject the significance of the whole set of leisure terms, that is, the quadratic term and the various linear terms (p-value of 0.022). Turning back to the complete model I, we also see that preference shifters on leisure are broadly insignificant, which is also because these variables enter the model additively through z_{it}

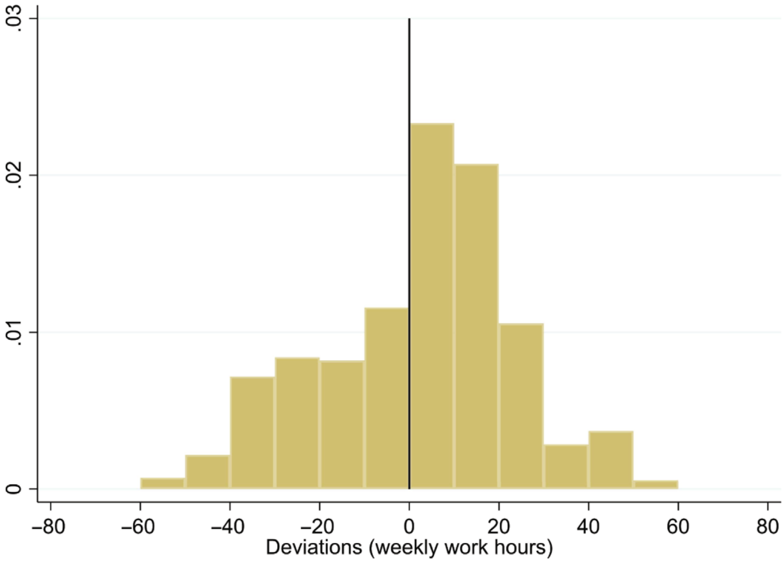
hours h_{ijt} from 0 ($j = 1$) to 60 ($j = 7$) with a step of 10 hours. With total time available for work normalized to 80 hours per week, leisure $l_{ijt} = 80 - h_{ijt}$ ranges from 80 to 20 hours per week. As seen later, our results do not change much when using a thinner grid ($J = 13$). For each hour option j, disposable income $y_{ijt} = G_t(w_{it}h_{ijt}, \mu_{it}, \zeta_{it})$ is easily calculated using gross hourly wage rates w_{it} and discretized values of hours h_{ijt}. Then, we numerically search the option j, hence the hour h_{ijt}, which maximizes U_{it}^E.

(for socio-demographic variables) and α_i (for psychological traits). If we ignore these additive controls, as in model IV, the role of preference-shifters reappears more distinctively. Their effects tend to increase the value of leisure for men, Londoners, or people with high conscientiousness. Inversely, it puts a lower weight on leisure for women and especially single mothers. This result anticipates the characterization that comes next: those who tend to overwork (underwork) value leisure more (less) in their actual income–leisure situation.

3.2. Deviations

3.2.1. Distribution of deviations: overall characterization. We now present deviations between actual and SWB-maximizing hours using the baseline model. We calculate deviations $D_{it} = h_{it} - h^*_{it}$ for every person–time unit of observation. Figure 1 shows their distribution: it is single-peaked, relatively symmetrical, and with a mode close to zero. As reported in the first row of Table 1 (first column), the mean D_{it} is −2.9 weekly hours. That is, on average, individuals work 2.9 hours less than their SWB-maximizing work duration. The bootstrapped standard errors in parentheses indicate that, overall, the mean deviation is not significantly different from zero at conventional

Figure 1. Distribution of individual deviations



Notes: Authors' own calculations from the BHPS. Deviations (horizontal axis) are defined as the distance between observed worked hours and SWB-maximizing hours. The mean deviations is −2.9 hours with a standard error of 5.7 hours. Standard errors are calculated for each individual i and time t using estimates from 200 bootstrapped samples.

levels. This means that, on average, actual labor supply choices – implying maximization of decision utility – are consistent with choices that maximize experienced utility. Note, however, that the mean D_{it} is the result of positive or negative deviations, which differ for each individual and period in the sample. Thus, we also calculate the bootstrapped standard error for each observation in the sample and report in the next columns the frequency of observations for which deviations are significantly different from zero, negative, or positive at the 5 percent level. For the whole sample, the deviations are significantly non-zero in 28 percent of cases, and correspond mostly to significantly negative deviations, which is consistent with the slightly negative average deviation. In other words, for 72 percent of the observations, there is no strong dissonance between actual choices and hours that would maximize SWB.[19]

3.2.2. Comparison with the literature. Despite the non-experimental context, our results are close to the conclusions of controlled experiments. Namely, the bulk of observed choices are consistent with the pursuit of individual satisfaction. In particular, Benjamin et al. (2012) show that most (but not all) individuals are able to predict their SWB at the moment of deciding about (hypothetical) job opportunities. Benjamin et al. (2014b), looking at actual residency choices, show that SWB scores are correlated with the ranking of actual choices (even if the trade-offs between aspects of residency tend to be different). Fleurbaey and Schwandt (2015) ask people if they can think of changes that would increase their SWB score. About 60 percent cannot think of an easy improvement (i.e., feel as if they currently maximize SWB). Clark et al. (2015) also find similar relative concerns in SWB regressions and in hypothetical-choice experiments. Our results are in line with the optimistic view that there is overall congruence between revealed and subjective preferences,[20] but perhaps the most interesting aspect is when there is not – which is what we study below.

[19]Compared with studies of people's views on what would be their best option for maximizing SWB (see Fleurbaey and Schwandt, 2015), we rely on a prediction of this optimal choice using our estimated experienced utility model. This means that some of the deviation may come from prediction errors. We argue that this issue is limited given our rich structure in terms of preference heterogeneity. Also, it is unlikely that unobserved heterogeneity drives the deviation measures upward or downward systematically, either for the whole sample or for broad population groups. Thus, comparing the sign and size of deviations across these groups can still reveal different exposures to the factors that limit the ability to maximize SWB. This is what we check in the following subsections.

[20]This is not always the case. Ferrer-i-Carbonell et al. (2011) compare the estimates on job characteristics in choice equations using vignettes to those on the same characteristics in determining the respondent's own job satisfaction, finding significant differences. Perez-Truglia (2015) shows that real consumption is well predicted by life satisfaction but not by economic satisfaction.

Table 1. Mean deviations: overall and by group

	Mean deviation (working hours)		Proportion of deviations that are		
			non-zero	negative	positive
For the whole sample	−2.9	(5.7)	0.28	0.19	0.10
For specific groups:					
Gender					
Female	−11.1	(7.6)	0.31	0.27	0.04
Male	11.5	(7.2)	0.24	0.04	0.20
Age					
Young	−7.6	(6.2)	0.31	0.25	0.06
Old	3.4	(6.1)	0.25	0.10	0.15
Children					
No	6.3	(6.1)	0.19	0.05	0.14
Yes	−20.5**	(9.3)	0.46	0.45	0.02
London					
No	−4.6	(6.1)	0.27	0.20	0.07
Yes	19.9***	(7.4)	0.53	0.02	0.51
Conscientiousness					
Low	0.9	(5.9)	0.29	0.16	0.13
High	−7.7	(6.4)	0.28	0.22	0.06
Neuroticism					
Low	−1.8	(5.8)	0.27	0.17	0.10
High	−4.0	(6.2)	0.30	0.21	0.09

Notes: Authors' own calculations from the BHPS. Deviations are defined as the distance between observed worked hours and SWB-maximizing hours. Standard errors in parentheses are calculated for each individual i and time t using estimates from 200 bootstrapped samples. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

3.3. Discrepancies and suggestive explanations

3.3.1. Observed heterogeneity. Deviations are small, on average, and infrequent. Yet, larger discrepancies appear for specific groups, as illustrated in Table 1 from the second row onwards. We observe that the average deviation is positive and large for men (11.5 hours, s.e. 7.2) and Londoners (19.9 hours, s.e. 7.4). This can be interpreted as if these two groups of individuals were working too much from am SWB-maximization perspective. Inversely, women and single parents seem to work too little as their mean deviation is negative, on average. The fraction of statistically significant deviations ranges from 24 percent to 53 percent and is consistent across groups: large proportions of significant discrepancies are seen when the mean deviations is large in absolute terms (e.g., positively for Londoners or negatively for single parents). The last two columns confirm that the sign is right. For instance, the very large mean deviation for Londoners coincides with almost all of the significant deviations being positive.

Table 2. Deviations: by discrete hour level

Hours of work	Mean deviation	s.e.	Proportion of deviations that are significantly non-zero	negative	positive
All					
0	−34.6***	(7.8)	0.78	0.78	0.00
10	−16.9**	(7.3)	0.31	0.31	0.01
20	−14.4*	(8.1)	0.30	0.27	0.03
30	−4.7	(7.1)	0.09	0.06	0.03
40	7.9	(5.5)	0.10	0.00	0.10
50	19.8***	(6.1)	0.41	0.00	0.41
60	31.9***	(6.0)	0.83	0.00	0.83
Male					
0	−27.2***	(7.6)	0.66	0.66	0.00
10	−4.2	(7.2)	0.21	0.16	0.05
20	−0.8	(7.6)	0.14	0.08	0.06
30	3.9	(6.3)	0.16	0.00	0.16
40	12.8	(7.2)	0.14	0.00	0.14
50	23.6***	(7.6)	0.58	0.00	0.58
60	35.4***	(7.9)	0.96	0.00	0.96
Female					
0	−35.6***	(8.7)	0.80	0.80	0.00
10	−18.3**	(8.1)	0.33	0.33	0.00
20	−15.4*	(8.6)	0.31	0.28	0.02
30	−6.2	(8.0)	0.08	0.07	0.01
40	2.6	(7.6)	0.06	0.00	0.06
50	13.8	(8.7)	0.13	0.00	0.13
60	25.5***	(8.2)	0.61	0.00	0.61

Notes: Authors' own calculations from the BHPS. Deviations are defined as the distance between observed worked hours and SWB-maximizing hours. Standard errors are calculated for each individual i and time t using estimates from 200 bootstrap samples. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

These results are also consistent with the simple intuitions from the SWB estimates above, which already revealed the overworked or underworked groups to some extent. We provide more extensive interpretations on the nature of these discrepancies below. Beforehand, Table 2 reports the distribution of deviations by actual work duration (expressed by discretized weekly hours). People working a standard full-time (30 or 40 hours per week) show small average deviations – and a low rates of significant deviations – compared with those at the extremes of the hour distribution (0–20 and 50–60 hours). As expected, those at zero hours tend to work too little and those at 50–60 hours per week appear to work too much from an SWB perspective. The remaining columns of Table 2 show the result by gender, which we comment on later. We will also discuss the fact that zero or reduced work hours might largely reflect labor market constraints. Note that we have excluded job-seekers from our baseline sample, who are possibly rationed out of the labor market because of Keynesian or classic unemployment. With this interpretation, the extent of underworked situations would be even larger if we included them, which we do in robustness checks. That said, there might also be a fair amount of rationing in our baseline sample, namely among inactive people who declare that they are not looking for a job. This could be the case of discouraged workers (people who have given up searching for a job because of labor market conditions) or of those financially disincentivized to work (due to low productivity and/or high childcare costs).

3.3.2. Broad factors explaining deviations. Large deviations can be explained by three broad types of mechanisms: constraints, mistakes, and alternative life goals. First, the presence of constraints that prevent first-best choices pertains to individual factors (e.g., family obligations) or external factors (such as market imperfections for credit, labor, or housing markets). This explanation is very likely in our context, especially the role of labor market constraints, as shown below.[21] Constraints might explain, at least partly, the contrasted pattern observed for men versus women in Table 1. The fact that women work too little from an SWB perspective might result from underemployment due to labor market rationing, discrimination (Petrongolo, 2004) and sticky floor, or low financial gains from work for low-skilled women and those facing high childcare costs (Blundell et al., 2000, 2008; Viitanen, 2005). More generally, constraints seem a good explanation for the pattern in Table 2, whereby large discrepancies are concentrated at extreme hours. Large negative (positive) deviations and a high frequency of people reporting underwork (overwork) situations are found for people with no or small activity (long working weeks) and especially for women (men).

The second type of factor explaining deviations pertains to optimization errors from an SWB perspective.[22] In our context, people may fail to predict

[21] See also the evidence based on desired hour information (e.g., Ilmakunnas and Pudney, 1990; Böheim and Taylor, 2004; Bryan, 2007). Yet, note that when information on desired hours is available, it is difficult to make sure that individuals' answers to the preferred hours question only reflect preferences (and are not themselves affected by some constraints).

[22] This aspect is extensively investigated through numerous experiments in the behavioral economics literature, exploring different dimensions of suboptimality (such as projection errors, as in Loewenstein et al., 2003, and Loewenstein and Adler, 1995), excessive aspirations, heuristics, or "focusing illusions" (Kahneman et al., 2006).

the future satisfaction levels resulting from their choices when they had to make a labor supply decision (see also Odermatt and Stutzer, 2019; Odermatt et al., 2021). They may work too much due to peer pressure or to a focusing illusion on the importance of income for instance.[23] This might explain some of the differences between Londoners and the rest of the UK, if there are regional differences in aspirations and positional concerns (e.g., local norms may generate adaptive preferences leading to workaholism; cf. Golden and Altman, 2008). Gender differences in career orientation or concern for status might also explain that men suffer from doing more excessive overtime, as illustrated in Table 2; for example, Frijters (2000) consistently find that men are more likely to find their job important, indicating a higher level of ambition. Note that concepts are not mutually exclusive, which makes interpretations even more difficult. For instance, suboptimal behavior (e.g., excessive overtime or workaholism) can be due to a combination of ambition, status concerns, and psychological biases (e.g., the need for recognition, etc.) and/or normative constraints or associated beliefs (e.g., demanding job rhythm due to social pressure on the high-skilled, Londoners, etc.).[24]

The third mechanism is of a somewhat opposite nature: actual decisions might be more relevant than SWB if they reveal other life goals than the pursuit of short-term personal satisfaction (as we measure it). Life goals might be different because of altruism (e.g., working hard to provide for one's children, to leave a bequest, etc.), intertemporal optimization (e.g., working hard to save for later, to achieve fame, etc.), or alternative objectives that diverge from SWB (e.g., moral objectives, honor, religious motives, recognition, etc.). It is more difficult to see how this type of factor could explain observed differences between men and women, Londoners and others in our results.[25] Moreover, experienced utility in our baseline is a "concentrated" measure of income–leisure preferences, which is relatively specific and possibly distant from some other life goals.[26]

[23]People might focus on one aspect (income) while ignoring the effect of hedonic adaptations to a certain level of wealth (Kahneman and Thaler, 2006; Di Tella et al., 2010). Kahneman et al. (2006) state that "despite the weak relation between income and global life satisfaction or experienced happiness, many people are highly motivated to increase their income. In some cases, this focusing illusion may lead to a misallocation of time."

[24]See Farzin (2009) and Hamermesh and Slemrod (2008) on beliefs and norms, and see Loewenstein et al. (2003) on projection biases that can create a tendency to repeatedly increase labor and decrease leisure relative to earlier plans.

[25]We are going to investigate below further subgroups, including caring for an elderly person at home ("Family care" in Table 3), which might relate to this explanation more closely.

[26]Yet, our discussion is predicated on the idea that an individual's response to SWB questionnaires is about maximizing personal immediate gratification while even our income–leisure satisfaction might reflect some of the other life goals or values (for instance, if people internalize the future benefits of working hard in the present, the satisfaction of spending time caring for someone else, etc.).

3.3.3. A focus on optimization constraints. Explaining discrepancies between experienced and decision utility for a particular group is a daunting task. First, it can be difficult to disentangle the three sets of factors outlined above. Interpretations of the role of specific factors might not be mutually exclusive. For instance, underemployment due to the care of an elderly parent can be seen as an alternative life objective or as a constraint (altruistic goal versus moral obligations). Second, it is certainly impossible to find variables that would comprehensively capture these three groups of factors. Non-hedonistic life goals and irrational behavior are especially hard to proxy with the information available in standard surveys such as the BHPS. Consequently, we suggest a simple exercise mainly focusing on constraints. We extract from the BHPS a number of proxies that potentially relate to different barriers on a person's ability to choose their desired working time. We distinguish between external constraints (e.g., pertaining to labor market conditions) and individual constraints (such as family obligations or health conditions). Results are reported in Table 3.

We first use variation in local unemployment rates across 12 regions × 10 periods to capture high versus low tension in the labor market. Recall that we exclude job-seekers so that, in our sample, the proportion of underwork by those voluntarily inactive or in small part-time work is not very different across regions with high versus low unemployment. However, Table 3 shows that 20 percent of our observations correspond to people who tend to overwork when there is high unemployment. They might refrain from changing jobs (i.e., to adjust their working time to improve SWB) due to high local employment insecurity. This is consistent with past evidence for the UK using information on desired hours of work; Stewart and Swaffield (1997) show that many workers would prefer to work less than they do when there is relative scarcity of alternative job opportunities. Next, we see that ethnic minorities also seem to face high pressure to work more than would be in line with income–leisure satisfaction. This seems to prevail over any form of discrimination in terms of access to jobs for the period under study.

We then exploit variation in individual constraints. Individuals' health status might be an important factor as we observe that those in poor health tend to work too little from an SWB perspective. This is also the case of those who have experienced long unemployment spells in the past, which might reflect scaring effects or selection, and those with low education. Regarding the family, we consider a broader concept than just the presence of children (as some children may be old enough not to require care time). A "family care" dummy accounts more explicitly for the fact that a person must take time to care for a person (e.g., an elderly relative) who is not necessarily living in the household. In Table 3, this situation is associated with extremely large deviation denoting underwork (note that the fraction of negative deviations is even higher than when we use a dummy for the presence of children in

Table 3. Deviations by proxies for potential explanatory factors: optimization constraints

Explanatory factor	Mean deviation		Proportion of deviations that are significantly		
		s.e.	non-zero	negative	positive
Labor market conditions					
Regional unemployment rate					
High	1.8	(5.3)	0.36	0.16	0.20
Low	−4.1	(6.0)	0.26	0.19	0.07
Ethnicity					
Non-White	9.4*	(5.4)	0.39	0.11	0.28
White	−3.2	(5.8)	0.28	0.19	0.09
Personal circumstances					
Health					
Poor	−9.5	(6.2)	0.41	0.32	0.08
Good	−2.3	(5.7)	0.27	0.18	0.10
Previous unemployment spells					
Long	−26.3***	(6.0)	0.62	0.60	0.02
Short/none	−2.0	(5.7)	0.27	0.17	0.10
Education					
Low	−8.0	(6.5)	0.34	0.27	0.07
High	2.1	(5.2)	0.23	0.11	0.12
Family care					
Yes	−35.8***	(9.1)	0.81	0.81	0.00
No	1.9	(5.6)	0.21	0.10	0.11
Commuting					
High	7.6	(4.9)	0.22	0.02	0.20
Low	−5.0	(6.0)	0.30	0.22	0.08

Notes: Authors' own calculations from the BHPS. Deviations are defined as the distance between observed worked hours and SWB-maximizing hours. Standard errors are calculated for each individual i and time t using estimates from 200 bootstrap samples. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

Table 1). Admittedly, it could also be interpreted as other life goals (taking care of loved ones); yet Fleurbaey and Schwandt (2015) show that "family obligations" are among the factors reported as most important for what prevents individuals from achieving greater SWB. Finally, we observe that long commutes entail the feeling of working too much while living far from one's work might be due to housing market constraints.[27]

[27] In Stutzer and Frey (2008), long commuting is indeed negatively correlated with SWB even after controlling for the endogenous sorting of individuals into location choice. Yet our result is also consistent with suboptimal decisions, if people who choose faraway jobs are not able to correctly guess well-being implications (see Kimball and Willis, 2006). The consequences of a focusing illusion on work and money can include both overtime and lengthy commutes.

Table 4. Explaining deviations using proxies for constraints

	(1)	(2)	(3)	(4)	(5)
High regional unemployment	5.256***	6.207***	6.327***	5.897***	5.652***
	(1.134)	(1.024)	(0.841)	(0.793)	(0.776)
Non-White ethnic origin	9.863***	11.04***	10.21***	11.36***	11.07***
	(2.998)	(3.060)	(2.300)	(2.259)	(2.218)
Poor health	−4.573***	−3.025**	−0.781	−0.770	−0.852
	(1.616)	(1.336)	(1.001)	(0.908)	(0.882)
Long unemployment spells	−21.52***	−23.39***	−28.46***	−27.53***	−27.03***
	(1.676)	(1.534)	(1.437)	(1.404)	(1.398)
Low education	−8.739***	−5.609***	−0.901	0.228	0.474
	(1.055)	(0.864)	(0.705)	(0.646)	(0.646)
Female		−22.10***	−16.87***	−12.54***	−12.50***
		(0.764)	(0.673)	(0.710)	(0.707)
Family care			−31.64***	−26.56***	−26.14***
			(0.912)	(0.892)	(0.893)
One child				−13.72***	−13.54***
				(0.904)	(0.897)
Two children				−10.81***	−10.62***
				(0.939)	(0.940)
Three children				−9.944***	−9.699***
				(1.420)	(1.409)
Four or more children				−3.788	−3.529
				(2.641)	(2.618)
High commuting					3.357***
					(0.698)
Constant	1.142*	13.36***	11.77***	11.89***	11.12***
	(0.627)	(0.630)	(0.557)	(0.540)	(0.568)
R^2	0.106	0.340	0.544	0.593	0.596
Number of observations	5,501	5,501	5,501	5,501	5,501

Notes: Authors' own calculations from the BHPS. The dependent variable is the deviations. It is defined as the distance between observed worked hours and SWB-maximizing hours. The models are estimated using OLS with robust standard errors. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

Although the set of explanations in terms of implicit constraints, as mobilized in the previous analysis, might not be exhaustive of all the constraints faced by British workers, we wish to test whether they already explain a substantial part of the observed variation in deviations. We regress D_{it} on these variables in a stepwise way and report the results in Table 4. Column 1 includes labor market conditions and individual factors related to health, past unemployment, and education. The signs are in line with previous interpretations: high tensions on the labor market or being from an ethnic minority group contribute to an upward pressure on work duration, health contributes to underwork situations, and so do the scaring effects from past unemployment or being low-skilled. Column 2 isolates the role of gender, which may pertain, to a large extent, to differences in labor market constraints between men and women. It partly correlates with the low-skill effect (the associated coefficient decreases) but it has a strong independent contribution to underemployment (R^2 increases by 23 points).[28] Further, Column 3 adds family care: those in charge are compelled to work less than desired (the effect is substantial, as R^2 increases by another 20 points). Column 4 refines the picture by adding detailed information on the number of children, which correlates with child age.[29] Column 5 adds high commuting as a potential constraint, which correlates with overwork. Interestingly, this set of constraint variables alone explains in total more than half of the variation in individual deviations (final $R^2 = 0.596$).

3.4. Sensitivity checks

Finally, we provide an extensive sensitivity analysis of our results. Our findings are summarized in Table 5 (the first row reproduces our baseline results) and discussed below. Detailed results are reported in the Appendix.

3.4.1. Alternative measures of experienced utility. Our baseline proxy for experienced utility was a concentrated income–leisure satisfaction measure $V_{it}^E = \widehat{\gamma}^y S_{it}^y + \widehat{\gamma}^l S_{it}^l$, with weights obtained from a regression of life satisfaction S_{it} on income satisfaction S_{it}^y and leisure satisfaction S_{it}^l. In Table 5 (rows 2–5), we suggest alternative proxies for experienced utility. We first employ a more flexible specification of the first-stage estimation (row 2), namely quadratic with an interaction term and heterogeneous coefficients (using the same variables x_{it} as in taste-shifters for the experienced utility estimation:

$$S_{it} = \gamma_1^y(x_{it})S_{it}^y + \gamma_1^l(x_{it})S_{it}^l + \gamma_2^y S_{it}^{2,y} + \gamma_2^l S_{it}^{2,l} + \gamma^{y,l} S_{it}^y S_{it}^l + e_{it}. \quad (8)$$

We also extend the concentrated measure to other domains of satisfaction (row 3), which might be correlated with the appreciation of one's income and time, namely satisfaction with health (he) and housing (ho):

$$S_{it} = \gamma^y S_{it}^y + \gamma^l S_{it}^l + \gamma^{he} S_{it}^{he} + \gamma^{ho} S_{it}^{ho} + e_{it}. \quad (9)$$

[28]We run additional multinomial logit estimations with three alternatives: negative deviation, positive deviation, and insignificant deviation (the reference category). We find that being female both increases the probability of negative deviations and reduces the probability of positive deviations.

[29]Underemployment appears to be a stronger concern for those with only one child while it has a less depressing effect for larger families (i.e., probably when some children are older and can possibly care for their siblings).

Table 5. Robustness checks

	Mean deviation		Proportion of significant deviations
(1) Baseline	−2.9	(5.7)	0.28
Alternative definitions of SWB			
(2) Quadratic in income and leisure satisfactions, with demographic shifters	−3.5	(4.7)	0.25
(3) Linear in income, leisure and additional satisfaction domains	−6.7	(5.0)	0.30
(4) PCA income–leisure satisfaction	−5.1	(5.9)	0.28
(5) Overall life satisfaction	−16.7***	(6.4)	0.48
Alternative functional forms			
(6) Quadratic no interaction	−3.2	(5.8)	0.29
(7) Cubic	−1.9	(6.3)	0.24
(8) Log-linear	−6.0	(7.4)	0.44
(9) Box–Cox	−7.5	(10.1)	0.36
(10) Quadratic with alternative discretization (13 income–labor pairs)	−5.2	(5.7)	0.28
Alternative treatments of additive heterogeneity			
(11) Fixed-effects	−7.3	(5.6)	0.33
(12) Random-effects	−4.7	(5.4)	0.28
(13) Quasi-fixed-effects	−1.6	(5.1)	0.28
(14) No additive observed heterogeneity	−12.0***	(4.2)	0.55
Alternative specifications of taste-shifters			
(15) Continuous age and personality scores	−4.1	(4.3)	0.39
(16) Baseline with all big five	−3.1	(5.7)	0.27
(17) Baseline with all big five and all other explanatory variables	−2.0	(6.1)	0.33
Additional checks: estimators and sample selection			
(18) Cross-sectional ordered probit model	−8.6	(7.5)	0.27
(19) Including job-seekers	−7.0	(5.1)	0.31
(20) Including self-employed	−6.9	(4.2)	0.40

Notes: Authors' own calculations from the BHPS. The deviation is defined as the distance between observed worked hours and SWB-maximizing hours. Standard errors are calculated for each individual i and time t using estimates from 200 bootstrap samples. Detailed results for the subgroups are presented in Tables A2–A6 in the Appendix. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

We then suggest a measure based on a principal component analysis (PCA) of income and leisure satisfactions (row 4). In all these cases, results are very close to the baseline, with salient groups affected by underemployment (women, those with poor health, or those with previous experiences of long unemployment spells) or excessive overtime (Londoners). Detailed results are shown in Table A2 in the Appendix (Columns 2–5). Note that a last variant (in row 5 of Table 5) takes overall life satisfaction S_{it} as a measure of experienced utility V_{it}^{E}. Overall satisfaction is noisy, mixes many life dimensions, and absorbs much individual heterogeneity, so results are different and point to large deviations. Nonetheless, the aforementioned differences between groups (e.g., with or without family care) are still visible qualitatively (see Column 5 of Table A2). Yet the use of overall satisfaction is not very informative.

3.4.2. Functional forms and hour discretization. We also check the sensitivity of our results to alternative parametric forms for the deterministic part of experienced utility $U^E(y_{it}, l_{it}; x_{it})$. Results are given in Table 5 (rows 6–9), and detailed estimates are shown in Table A3 (Columns 6–9). We first use a less flexible quadratic form whereby separability between income and leisure is imposed (row 6):

$$U_{it}^{E}(y_{it}, l_{it}; x_{it}) = \beta_{yy} y_{it}^2 + \beta_{ll} l_{it}^2 + \beta_y y_{it} + \beta_l(x_{it}) l_{it}. \tag{10}$$

That is, there is no income × leisure interaction term, as in Knabe and Rätzel (2010) or the alternative specifications discussed in Section 3.1. Inversely, we suggest a more flexible polynomial form, namely a cubic specification including all possible interaction terms between income and leisure (row 7). Two other functional forms are popular in welfare economics, namely the log-linear utility (row 8):

$$U_{it}^{E}(y_{it}, l_{it}; x_{it}) = \beta_y \ln y_{it} + \beta_l(x_{it}) \ln l_{it}, \tag{11}$$

often used in SWB studies (e.g. Clark et al., 2008) and capturing some nonlinearity in income and leisure, and the Box–Cox utility (row 9)

$$U_{it}^{E}(y_{it}, l_{it}; x_{it}) = \beta_y \left(\frac{y_{it}^{\lambda_y} - 1}{\lambda_y}\right) + \beta_l(x_{it}) \left(\frac{l_{it}^{\lambda_l} - 1}{\lambda_l}\right) \tag{12}$$

used in numerous empirical studies (e.g., Decoster and Haan, 2015). All these models include taste-shifters on the composite leisure term $\beta_l(x_{it})$. In all these cases, results are very similar to the baseline. There are small variations, especially in the log-linear case, which is arguably more restrictive. Yet, our conclusions are broadly robust to the choice of the functional form imposed on the deterministic part of the experienced utility function.

Regarding the discretization used to compute deviations, we account for $J = 7$ different income–leisure pairs (y_{ijt}, h_{ijt}) in the baseline, corresponding to weekly work hours from 0 ($j = 1$) to 60 ($j = 7$) with a step of 10 hours. This grid seems precise enough to accommodate any actual choices. However, the approximation might lead to measurement error when assessing the degree of deviations. Thus, we experiment with a thinner grid, namely $J = 13$ points with a step of five hours. Results show that the mean deviation increases slightly (row 10), but not the share of individuals with a significant error. The heterogeneity across groups is very similar to the baseline (see the last column of Table A3).

3.4.3. Treatment of additive individual heterogeneity. In the model described by equation (3), the part of the utility not related to income and leisure is supposed to capture individual heterogeneity in how people perceive and report their well-being. For that purpose, we have included observed individual characteristics as additive shifters z_{it} and a time-invariant individual effect α_i based on key psychological traits, as sometimes done in the literature (see Boyce, 2010). Alternatively, we can use panel estimations of the experienced utility function with α_i modeled as fixed effects (FE), random effects (RE), or quasi-fixed effects (QFE) following Mundlak (1978). Relying only on within variation, QFE following Mundlak are modeled as RE plus the time average of relevant time-varying controls in the estimation (time-variant variables in the auxiliary distribution of unobserved heterogeneity are health status, number of children, and region).[30] Estimates of the FE, RE, and QFE models are reported in Table 5 (rows 11–13). Reassuringly, results are relatively close to the baseline. A specification without additive terms $\lambda' z_{it} + \alpha_i$ shows extremely noisy results and confirms the point made by Decancq et al. (2015) that an attempt to recover a meaningful preference structure needs to clean SWB from individual heterogeneity. Detailed results are reported in Table A4 in the Appendix.

3.4.4. Preference heterogeneity (taste-shifters). We test the sensitivity of our results with respect to the specification of preference shifters x_{it} used in the deterministic part of experienced utility function. In our baseline, the coefficient of leisure varied linearly with the set x_{it}. For the ease of exposition of heterogeneous results across population groups, these shifters were defined as binary variables (male, age above 40, presence of children, living in London, above-average conscientiousness, and average neuroticism). In Table 5, we present additional results (rows 15 and 16), starting with

[30]Indeed, between variation can attenuate differences (as it captures long-term trends possibly smoothed by adaptation) while within variation can lead to different estimates (in particular, subjective appreciation of transition in or out of work might be stronger for those who experience these changes over the course of the survey). See Fleurbaey and Blanchet (2013) for a discussion of SWB estimations in the context of panel data.

the same set of variables but using intensive form of age (in years) and personality traits (on a scale of 1–4), then expanding shifters to the whole set of personality traits. In both cases, results are similar to the baseline. Finally, we extend the set of shifters by including various variables used in our previous characterization of the potential factors explaining deviations (all characteristics appearing in Tables 1 and 3). Many of these variables pertain to the demand side of the labor market or other sources of constraints, rather than preferences, so that this specification can be seen as the reduced form of a more complete model. With some exceptions, results are close to the baseline (row 17), which means that basic taste-shifters – that comply more with a labor supply interpretation – also captured much of these other dimensions. Detailed results are presented in Table A5 in the Appendix.

3.4.5. Ordered probit estimation and inclusion of job-seekers and self-employed. We suggest three last sensitivity checks. The first is the use of an alternative estimation method. The concentrated satisfaction measure has been treated as a continuous variable for linear estimations. Yet, the satisfaction measures are observed on an ordinal scale and we aim to investigate whether the results are sensitive to the choice of estimator. Having calculated the concentrated measure of experienced utility, we transform the variable back to its original ordinal state (i.e., the nearest integer to reconstitute a 1–7 scale as for the original income and leisure satisfaction answers). Doing so, we then estimate an ordered probit model (instead of OLS) using the discretized concentrated SWB measure. Results are close to our baseline (row 18 in Table 5 and the second column in Table A6 in the Appendix).

Our baseline sample has excluded job-seekers and the self-employed. We now add these groups of individuals into our analysis for a better external validity. To be able to include job-seekers without biasing our main results, we suggest an alternative estimation method based on a double hurdle model (Blundell et al., 2000). Table 5 shows that the intensity of negative deviations increases (row 19), with a mean deviation of −7 hours. This is expected as job-seekers are constrained, by definition, and contribute to our characterization of underwork. Yet, they represent only a small percentage (3 percent) of the initial sample, which explains why the share of significant deviations increases only slightly (from 28 percent to 31 percent). Heterogeneous effects, described in Table A6, vary a little but do not lead to different conclusions. Among exceptions, we see that the mean error for men is now negative, which translates the fact that job-seekers are mainly men rationed out of the labor market.

Finally, we add the self-employed to our baseline sample (the resulting sample is 6,088 observations with 9.6 percent of self-employed). The inclusion of self-employed workers yields a larger mean deviation (−6.9 hours), which remains statistically insignificant for the whole sample (row 20 of Table 5). However, for this group, the mean deviation is statistically significant and large (last rows of Table A6). This is not surprising considering that working hours of the self-employed vary with several other factors potentially related to individual life goals (e.g., autonomy, personal ambition, among many others). This is consistent with the literature suggesting that the self-employed might suffer from mispredicting their well-being in relation to their actual working hours (e.g., Odermatt and Stutzer, 2019; Odermatt et al., 2021).

4. Concluding discussion

This paper compares decision and experienced utility using a large household survey. We focus on labor supply decisions, motivated by the fact that income–leisure domains crucially matter for welfare analysis and the design of redistributive policies. To this end, we estimate a series of experienced utility functions, with a structure similar to that of labor supply models, and we derive for each individual the deviation between the actual choice (consistent with decision utility) and the choice that would maximize their experienced utility. We find a high proportion of insignificant deviations, indicating a broad congruence between actual hours of work and SWB-maximizing decisions. However, deviations can be very large in some groups and explained by a variety of factors. Nonetheless, our analysis provides suggestive evidence that personal constraints (family obligations) and labor market constraints explain the bulk of these discrepancies.

In the particular context of labor supply and policy analysis, the methodological implication of our work is that there should be ways to improve our modeling of employment decisions by combining information on actual choices and the self-reported well-being derived from individual situations. Our deviation metric could be used as an original way to elicit labor market frictions and could be compared with other attempts to account for restrictions in labor supply models (e.g., Altonji and Paxson, 1982; Ilmakunnas and Pudney, 1990; Dickens and Lundberg, 1993; van Soest, 1995; Aaberge et al., 1999; Dagsvik and Strøm, 2006; Bloemen, 2008; Beffy et al., 2019).[31] A more systematic characterization of how deviations vary across countries/regions and, above all, with business cycles might help to

[31] Recent approaches characterize labor market frictions by comparing long- and short-term adjustments, assuming people are less constrained in the long run.

validate this measure, with larger deviations expected when frictions appear in places/times of strong demand-side constraints.

Many extensions and improvements can be suggested. First, our implicit comparison of decision and experienced utility in the context of non-experimental data could easily be extended to other areas in economics, such as transportation choices or savings (for consumption decisions, see Perez-Truglia, 2015). Second, deviations could be better explained – at least regarding observed heterogeneity – using longer and richer household surveys. Third, our models are static and do not consider the intertemporal decisions and the dynamic nature of repeated occurrences of decision and experience. Modeling intertemporal decisions would require additional information, including actual consumption at each period (e.g., Haan et al., 2008). More generally, further research should account for the potential time discrepancy and causal link between the observed decision (possibly made in the past) and the resulting income–leisure satisfaction. It could combine our approach with the panel dimension in order to check if people showing large deviations at one point in time are more likely to change job/contract in the future to adjust their working time – along the lines of Frijters (2000), Benjamin et al. (2012), Odermatt et al. (2021), and Odermatt and Stutzer (2019).

Appendix. Additional empirical results

Table A1. SWB estimations

	I	II	III	IV
Income^2	−4.61e−07***	−4.66e−07***	−4.94e−07***	−4.24e−07***
	(1.37e−07)	(1.36e−07)	(1.36e−07)	(1.43e−07)
Income	0.000762**	0.00117***	0.00123***	0.000825**
	(0.000374)	(0.000184)	−0.000185	(0.000417)
Income × Leisure	9.30e−06			7.32e−06
	(6.17e−06)			(6.91e−06)
Leisure^2	−4.22e−05	−0.000101**	−8.17e−05	−5.32e−05
	(5.82e−05)	(4.97e−05)	(5.17e−05)	(6.33e−05)
Leisure	0.00552	0.0140**	0.0116**	0.00788
	(0.00722)	(0.00561)	(0.00585)	(0.00812)
× male	0.00208		0.00197	0.00157**
	(0.00238)		(0.00238)	(0.000790)
× age	0.00111		0.00117	6.10e−05
	(0.000823)		(0.000825)	(0.000609)
× child	−0.00167		−0.00112	−0.00582***
	(0.00244)		(0.00244)	(0.000763)
× London	0.00571*		0.00581*	0.00869***
	(0.00300)		(0.00304)	(0.00301)
× high conscientiousness	−0.00146*		−0.00148*	0.00140**
	(0.000883)		(0.000886)	(0.000617)
× high neuroticism	0.000319		0.000321	−0.00542***
	(0.000834)		(0.000835)	(0.000607)
Additive controls z_{it} and α_i	Yes	Yes	Yes	No
Region and year dummies	Yes	Yes	Yes	Yes
R^2	0.243	0.239	0.242	0.148
Number of observations	5,501	5,501	5,501	5,501

Notes: Authors' estimations of subjective well-being (i.e., income–leisure satisfaction) using the BHPS. In baseline model I and some of the variants, the subjective well-being equation includes additively separable controls z_i (the same variables as in leisure interaction terms plus age squared, family size, health status, homeownership) and α_i (all personality traits). Standard errors are given in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

Table A2. Robustness checks: alternative measures of SWB

		Concentrated life satisfaction			PCA income–leisure satisf.	Life satisf.
		Baseline	Quadratic	Linear		
Corresponding rows in Table 5		1	2	3	4	5
Mean deviation		−2.9	−3.5	−6.7	−5.1	-16.7^{***}
		(5.7)	(4.7)	(5.0)	(5.9)	(6.4)
Proportion of significant deviations		0.28	0.25	0.30	0.28	0.48
Gender	Female	−11.1	−5.5	-13.5^{**}	-13.1^{*}	-17.1^{*}
		(7.6)	(5.6)	(6.1)	(7.6)	(8.8)
	Male	11.5	−0.2	5.1	8.8	-16.1^{**}
		(7.2)	(5.9)	(6.1)	(7.3)	(6.3)
London	No	−4.6	−4.7	−7.7	−6.9	-19.0^{***}
		(6.1)	(4.9)	(5.2)	(6.3)	(6.7)
	Yes	19.9^{***}	11.2	5.5	18.2^{**}	13.3
		(7.4)	(7.4)	(7.1)	(7.8)	(13.2)
Health	Poor	−9.5	-11.4^{**}	-13.8^{***}	-11.7^{*}	-22.5^{***}
		(6.2)	(5.3)	(5.4)	(6.3)	(7.3)
	Good	−2.3	−2.9	−6.2	−4.6	-16.3^{***}
		(5.7)	(4.7)	(5.0)	(5.9)	(6.3)
Regional unemployment	High	1.8	−0.5	−5.6	−0.2	−10.4
		(5.3)	(4.8)	(5.0)	(5.5)	(6.8)
	Low	−4.1	−4.4	−7.0	−6.4	-18.5^{***}
		(6.0)	(4.8)	(5.1)	(6.2)	(6.6)
Ethnicity	Non-White	9.4^{*}	7.9	1.2	7.0	2.4
		(5.4)	(4.8)	(4.9)	(5.5)	(7.5)
	White	−3.2	−3.8	−6.9	−5.4	-17.2^{***}
		(5.8)	(4.7)	(5.0)	(5.9)	(6.4)
Previous unemployment spells	Long	-26.3^{***}	-29.5^{***}	-31.6^{***}	-28.0^{***}	-41.5^{***}
		(6.0)	(5.3)	(5.3)	(6.3)	(7.0)
	Short/none	−2.0	−2.6	−5.8	−4.3	-15.8^{**}
		(5.7)	(4.7)	(5.0)	(5.9)	(6.4)
Education	Low	−8.0	−6.5	-12.7^{**}	−10.3	-19.9^{**}
		(6.5)	(5.2)	(5.7)	(6.7)	(7.7)
	High	2.1	−0.7	−1.0	−0.1	-13.7^{**}
		(5.2)	(4.5)	(4.6)	(5.4)	(5.6)
Family care	Yes	-35.8^{***}	-24.9^{***}	-34.5^{***}	-37.8^{***}	-30.6^{**}
		(9.1)	(6.9)	(7.0)	(9.0)	(13.1)
	No	1.9	−0.4	−2.7	−0.4	-14.7^{**}
		(5.6)	(4.7)	(5.0)	(5.8)	(5.8)
Commuting	High	7.6	2.2	3.3	5.5	−8.6
		(4.9)	(4.5)	(4.4)	(5.0)	(5.5)
	Low	−5.0	−4.7	-8.8^{*}	−7.2	-18.4^{***}
		(6.0)	(4.8)	(5.2)	(6.1)	(6.8)

Notes: Authors' own calculations from the BHPS. Standard errors are calculated for each individual i and time t using estimates from 200 bootstrap samples. Baseline is linear in income and leisure satisfactions, no heterogeneity; quadratic in income and leisure satisfactions, with demographic heterogeneity; linear in income, leisure and additional satisfaction dimensions. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

Table A3. Robustness checks: alternative functional forms and hour discretization

		Baseline	Quadratic (no interaction)	Cubic	Log-linear	Box–Cox	Quadratic (alternative discretization)
Corresponding rows in Table 5		1	6	7	8	9	10
Number of discretized hours		7	7	7	7	7	13
Mean deviation		−2.9	−3.2	−1.9	−6.0	−7.5	−5.2
		(5.7)	(5.8)	(6.3)	(7.4)	(10.1)	(5.7)
Proportion of significant deviations		0.28	0.29	0.24	0.44	0.36	0.28
Gender	Female	−11.1	−11.1	−10.7	−17.3	−18.8*	−13.1*
		(7.6)	(7.4)	(8.0)	(10.6)	(9.8)	(7.6)
	Male	11.5	10.5	13.5*	13.7*	12.1	8.4
		(7.2)	(7.5)	(7.3)	(7.3)	(13.8)	(7.2)
London	No	−4.6	−5.0	−3.5	−6.9	−10.1	−6.9
		(6.1)	(6.1)	(6.7)	(7.6)	(10.0)	(6.1)
	Yes	19.9***	20.2**	20.0***	5.9	26.4*	17.0**
		(7.4)	(8.1)	(7.0)	(8.9)	(15.5)	(7.4)
Health	Poor	−9.5	−8.9	−8.7	−15.7*	−14.4	−11.2*
		(6.2)	(6.1)	(6.8)	(8.6)	(10.3)	(6.2)
	Good	−2.3	−2.8	−1.3	−5.3	−7.0	−4.8
		(5.7)	(5.8)	(6.3)	(7.3)	(10.1)	(5.7)
Regional unemployment	High	1.8	2.9	2.1	−3.6	0.9	−0.5
		(5.3)	(5.3)	(5.5)	(7.4)	(11.1)	(5.3)
	Low	−4.1	−4.9	−3.0	−6.7	−9.8	−6.5
		(6.0)	(6.1)	(6.6)	(7.5)	(9.9)	(6.0)
Ethnicity	Non-White	9.4*	8.9	10.8*	−1.6	12.8	5.9
		(5.4)	(5.7)	(5.6)	(7.3)	(12.6)	(5.4)
	White	−3.2	−3.5	−2.2	−6.1	−8.0	−5.5
		(5.8)	(5.8)	(6.3)	(7.4)	(10.1)	(5.8)
Previous unemployment spells	Long	−26.3***	−25.5***	−24.6***	−30.6***	−27.1**	−26.0***
		(6.0)	(6.0)	(6.7)	(7.7)	(11.2)	(6.1)
	Short/none	−2.0	−2.4	−1.0	−5.1	−6.8	−4.5
		(5.7)	(5.8)	(6.3)	(7.4)	(10.1)	(5.7)
Education	Low	−8.0	−6.8	−7.3	−14.9*	−13.1	−9.7
		(6.5)	(6.1)	(7.1)	(8.7)	(10.3)	(6.5)
	High	2.1	0.3	3.4	2.5	−2.1	−0.9
		(5.2)	(5.6)	(5.9)	(6.5)	(10.1)	(5.2)
Family care	Yes	−35.8***	−35.5***	−35.2***	−56.0***	−49.8***	−35.7***
		(9.1)	(9.0)	(10.1)	(14.5)	(10.4)	(9.1)
	No	1.9	1.5	3.0	1.3	−1.4	−0.8
		(5.6)	(5.7)	(6.1)	(6.9)	(10.4)	(5.6)
Commuting	High	7.6	6.3	9.2	8.0	5.6	4.8
		(4.9)	(5.3)	(5.5)	(6.2)	(10.7)	(4.9)
	Low	−5.0	−5.1	−4.1	−8.8	−10.2	−7.2
		(6.0)	(5.9)	(6.5)	(7.7)	(10.0)	(6.0)

Notes: Authors' own calculations from the BHPS. Standard errors are calculated using 200 bootstrap samples for each individual and then averages for the mean deviation. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

Table A4. Robustness checks: alternative treatment of additive heterogeneity

		Baseline	Fixed effects	Random effects	Quasi-fixed effects	No additive observed heterogeneity
Corresponding rows in Table 5		1	11	12	13	14
Mean deviation		−2.9	−7.3	−4.7	−1.6	−12.0***
		(5.7)	(5.6)	(5.4)	(5.1)	(4.2)
Proportion of significant deviations		0.28	0.33	0.28	0.28	0.55
Gender	Female	−11.1	−8.4	−8.9	−5.2	−22.8***
		(7.6)	(6.8)	(6.5)	(6.1)	(4.3)
	Male	11.5	−5.3	2.7	4.7	6.9
		(7.2)	(8.3)	(7.1)	(7.3)	(4.6)
London	No	−4.6	−9.4	−6.7	−3.5	−11.8***
		(6.1)	(5.9)	(5.8)	(5.4)	(4.2)
	Yes	19.9***	20.1	22.2**	22.8**	−14.2***
		(7.4)	(14.1)	(9.0)	(8.9)	(4.6)
Health	Poor	−9.5	−10.9*	−10.9**	−7.7	−24.4***
		(6.2)	(5.9)	(5.5)	(5.1)	(4.2)
	Good	−2.3	−7.0	−4.2	−1.1	−11.0***
		(5.7)	(5.6)	(5.4)	(5.1)	(4.2)
Regional unemployment	High	1.8	1.8	3.0	5.4	−14.5***
		(5.3)	(5.9)	(4.8)	(4.6)	(4.3)
	Low	−4.1	−9.7	−6.7	−3.5	−11.3***
		(6.0)	(5.9)	(5.8)	(5.4)	(4.1)
Ethnicity	Non-White	9.4*	3.5	8.0	10.3*	−8.9**
		(5.4)	(9.4)	(6.1)	(6.1)	(4.3)
	White	−3.2	−7.5	−5.0	−1.9	−12.1***
		(5.8)	(5.6)	(5.4)	(5.1)	(4.2)
Previous unemployment spells	Long	−26.3***	−30.1***	−28.7***	−25.7***	−39.9***
		(6.0)	(5.7)	(5.4)	(5.1)	(4.3)
	Short/none	−2.0	−6.4	−3.8	−0.7	−11.0***
		(5.7)	(5.6)	(5.4)	(5.1)	(4.2)
Education	Low	−8.0	−11.2**	−8.8	−5.7	−19.0***
		(6.5)	(5.7)	(5.5)	(5.2)	(4.1)
	High	2.1	−3.4	−0.7	2.4	−5.2
		(5.2)	(5.8)	(5.5)	(5.3)	(4.3)
Family care	Yes	−35.8***	−43.0***	−37.8***	−35.1***	−53.2***
		(9.1)	(9.1)	(8.6)	(8.5)	(3.7)
	No	1.9	−2.1	0.2	3.3	−6.0
		(5.6)	(5.6)	(5.3)	(5.1)	(4.3)
Commuting	High	7.6	3.9	5.5	8.4	−2.7
		(4.9)	(5.9)	(5.2)	(5.1)	(4.3)
	Low	−5.0	−9.5*	−6.7	−3.6	−13.8***
		(6.0)	(5.7)	(5.5)	(5.2)	(4.2)

Notes: Authors' own calculations from the BHPS. Standard errors are calculated for each individual i and time t using estimates from 200 bootstrap samples. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

Table A5. Robustness checks: alternative specification of preference heterogeneity

		Baseline	Continuous age and personality scores	Baseline with all big 5	Baseline with all big 5 and all other explanatory variables
Corresponding rows in Table 5		1	15	16	17
Mean deviation		−2.9	−4.1	−3.1	−2.0
		(5.7)	(4.3)	(5.7)	(6.1)
Proportion of significant deviations		0.28	0.39	0.27	0.33
Gender	Female	−11.1	−10.3*	−11.3	−11.8
		(7.6)	(5.3)	(7.6)	(7.9)
	Male	11.5	6.7	11.3	15.0*
		(7.2)	(5.4)	(7.2)	(8.1)
London	No	−4.6	−5.6	−4.8	−3.6
		(6.1)	(4.5)	(6.1)	(6.5)
	Yes	19.9**	15.9**	19.9**	19.1**
		(7.4)	(6.8)	(7.5)	(6.7)
Health	Poor	−9.5	−8.9*	−10.4	−32.9**
		(6.2)	(4.6)	(6.3)	(7.4)
	Good	−2.3	−3.7	−2.5	0.4
		(5.7)	(4.3)	(5.7)	(6.4)
Regional unemployment	High	1.8	0.2	1.9	3.5
		(5.3)	(4.3)	(5.4)	(6.1)
	Low	−4.1	−5.2	−4.4	−3.5
		(6.0)	(4.5)	(6.0)	(6.5)
Ethnicity	Non-White	9.4*	6.9	8.0	−19.4
		(5.4)	(4.5)	(5.5)	(12.6)
	White	−3.2	−4.3	−3.3	−1.6
		(5.8)	(4.4)	(5.8)	(6.3)
Previous unemployment spells	Long	−26.3**	−23.1**	−26.4**	−30.5**
		(6.0)	(4.2)	(6.1)	(8.1)
	Short/none	−2.0	−3.4	−2.2	−1.0
		(5.7)	(4.3)	(5.7)	(6.2)
Education	Low	−8.0	−7.3	−8.2	−15.8*
		(6.5)	(4.6)	(6.6)	(8.5)
	High	2.1	−1.0	1.9	11.3*
		(5.2)	(4.3)	(5.2)	(6.7)
Family care	Yes	−35.8**	−36.2**	−36.2**	−42.9**
		(9.1)	(6.6)	(9.1)	(9.4)
	No	1.9	0.6	1.8	3.9
		(5.6)	(4.2)	(5.6)	(6.2)
Commuting	High	7.6	4.5	7.5	8.8
		(4.9)	(4.1)	(4.9)	(5.7)
	Low	−5.0	−5.8	−5.2	−4.2
		(6.0)	(4.4)	(6.0)	(6.4)

Notes: Authors' own calculations from the BHPS. Standard errors are calculated for each individual i and time t using estimates from 200 bootstrap samples. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

Table A6. Robustness checks: alternative estimator and samples selections

		Baseline	Ordered probit model	Job-seekers included	Self-employed included
Corresponding rows in Table 5		1	18	19	20
Mean deviation		−2.9	−8.6	−7.0	−6.9
		(5.7)	(7.5)	(5.1)	(4.2)
Proportion of significant deviations		0.28	0.27	0.31	0.40
Gender	Female	−11.1	−16.1*	−11.6*	−6.9
		(7.6)	(8.8)	(6.1)	(4.3)
	Male	11.5	4.6	0.7	−6.9
		(7.2)	(9.0)	(5.9)	(5.5)
London	No	−4.6	−11.1	−8.7	−8.8*
		(6.1)	(7.9)	(5.5)	(4.6)
	Yes	19.9***	24.3***	14.8**	15.4**
		(7.4)	(7.6)	(6.6)	(6.1)
Health	Poor	−9.5	−14.7*	−13.0**	−12.0***
		(6.2)	(8.2)	(5.4)	(4.2)
	Good	−2.3	−8.1	−6.6	−6.5
		(5.7)	(7.4)	(5.1)	(4.2)
Regional unemployment	High	1.8	−1.7	−2.7	−0.1
		(5.3)	(6.7)	(4.5)	(3.5)
	Low	−4.1	−10.4	−8.2	−8.8*
		(6.0)	(7.8)	(5.4)	(4.5)
Ethnicity	Non-White	9.4*	9.4	3.2	2.8
		(5.4)	(6.6)	(4.5)	(3.8)
	White	−3.2	−9.0	−7.3	−7.2*
		(5.8)	(7.5)	(5.2)	(4.2)
Previous unemployment spells	Long	−26.3***	−32.1***	−32.0***	−27.0***
		(6.0)	(8.2)	(5.4)	(4.3)
	Short/none	−2.0	−7.7	−5.8	−6.2
		(5.7)	(7.4)	(5.1)	(4.2)
Education	Low	−8.0	−13.9	−11.9**	−9.5**
		(6.5)	(8.5)	(5.8)	(4.4)
	High	2.1	−3.5	−2.2	−4.6
		(5.2)	(6.7)	(4.7)	(4.3)
Family care	Yes	−35.8***	−39.0***	−34.3***	−31.0***
		(9.1)	(11.2)	(7.4)	(5.7)
	No	1.9	−4.2	−3.0	−3.8
		(5.6)	(7.2)	(5.0)	(4.2)
Commuting	High	7.6	3.2	3.6	5.3
		(4.9)	(6.1)	(4.3)	(3.9)
	Low	−5.0	−11.0	−9.1*	−9.1**
		(6.0)	(7.8)	(5.4)	(4.3)
Employment status	Employee				2.3
					(4.2)
	Self-employed				−37.3***
					(4.7)
Number of observations		5,501	5,501	5,689	6,088

Notes: Authors' own calculations from the BHPS. Standard errors are calculated for each individual i and time t using estimates from 200 bootstrap samples. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

Supporting information

Additional supporting information can be found online in the supporting information section at the end of the article.

Replication files

References

Aaberge, R., Colombino, U., and Strøm, S. (1999), Labour supply in Italy: an empirical analysis of joint decisions, with taxes and quantity constraints, *Journal of Applied Econometrics 14*, 403–422.

Adam, S. and Browne, J. (2010), Redistribution, work incentives and thirty years of UK tax and benefit reform, Insitute for Fiscal Studies Working Paper 10/24, https://ifs.org.uk/publications/redistribution-work-incentives-and-thirty-years-uk-tax-and-benefit-reform.

Akay, A., Bargain, O., and Jara, H. X. (2020), "Fair" welfare comparisons with heterogeneous tastes: subjective versus revealed preferences, *Social Choice and Welfare 55*, 51–84.

Altonji, J. G. and Paxson, C. H. (1982), Labor supply preferences, hours constraints, and hours–wage trade-offs, *Journal of Labor Economics 6*, 254–276.

Beffy, M., Blundell, R., Bozio, A., Laroque, G., and Tô, M. (2016), Labour supply and taxation with restricted choices, *Journal of Econometrics 211*, 16–46.

Benjamin, D., Heffetz, O., Kimball, M., and Rees-Jones, A. (2012), What do you think would make you happier? What do you think you would choose?, *American Economic Review 102* (5), 2083–2110.

Benjamin, D., Heffetz, O., Kimball, M., and Szembrot, N. (2014a), Beyond happiness and satisfaction: toward well-being indices based on stated preference, *American Economic Review 104* (9), 2698–2735.

Benjamin, D., Heffetz, O., Kimball, M., and Rees-Jones, A. (2014b), Can marginal rates of substitution be inferred from happiness data? Evidence from residency choices, *American Economic Review 104* (11), 3498–3528.

Benjamin, D., Cooper, K., Heffetz, O., and Kimball, M. (2020), Self-reported wellbeing indicators are a valuable complement to traditional economic indicators but aren't yet ready to compete with them, *Behavioural Public Policy 4* (Special Issue 2), 198–209.

Bloemen, H. G. (2008), Job search, hours restrictions, and desired hours of work, *Journal of Labor Economics 26*, 137–179.

Blundell, R., Duncan, A., and Meghir, C. (1998), Estimating labor supply responses using tax reforms, *Econometrica 66*, 827–861.

Blundell, R., Duncan, A., McCrae, J., and Meghir, C. (2000), The labour market impact of the working families' tax credit, *Fiscal Studies 21*, 75–103.

Blundell, R., Brewer, M., and Francesconi, M. (2008), Job changes and hours changes: understanding the path of labor supply adjustment, *Journal of Labor Economics 26*, 421–445.

Böheim, R. and Taylor, M. (2004), Actual and preferred working hours, *British Journal of Industrial Relations 42*, 149–166.

Boyce, C. J. (2010), Understanding fixed effects in human well-being, *Journal of Economic Psychology 31*, 1–16.

Bryan, M. (2007), Free to choose? Differences in the hours determination of constrained and unconstrained workers, *Oxford Economic Papers 59*, 226–252.

Chiappori, P. A. and Donni, O. (2011), Non-unitary models of household behavior: a survey of the literature, in A. Molina (ed.), *Household Economic Behaviors*, Springer, Berlin.

Clark, A., Georgellis, Y., and Sanfey, P. (1998), Job satisfaction, wage changes and quits: evidence from Germany, *Research in Labor Economics 17*, 95–121.
Clark, A. E., Frijters, P., and Shields, M. (2008), Relative income, happiness and utility: an explanation for the Easterlin paradox and other puzzles, *Journal of Economic Literature 46*, 95–144.
Clark, A., Senik, C., and Yamada, K. (2015), When experienced and decision utility concur: the case of income comparisons, IZA Discussion Paper 9189.
Dagsvik, J. K. and Strøm, S. (2006), Sectoral labour supply, choice restrictions and functional form, *Journal of Applied Econometrics 21*, 803–826.
Decancq, K., Fleurbaey, M., and Schokkaert, E. (2015), Happiness, equivalent incomes, and respect for individual preferences, *Economica 82*, 1082–1106.
Decoster, A. and Haan, P. (2015), Empirical welfare analysis with preference heterogeneity, *International Tax and Public Finance 22*, 224–251.
Dickens, W. and Lundberg, S. (1993), Hours restrictions and labor supply, *International Econonomic Review 34*, 169–191.
Di Tella, R., Haisken-De New, J., and MacCulloch, R. (2010), Happiness adaptation to income and to status in an individual panel, *Journal of Economic Behavior and Organization 76*, 834–852.
Dolan, P. and Kahneman, D. (2008), Interpretations of utility and their implications for the valuation of health, *Economic Journal 118*, 215–234.
Farzin, Y. H. (2009), The effect of non-pecuniary motivations on labor supply, *Quarterly Review of Economics and Finance 49*, 1236–1259.
Ferrer-i-Carbonell, A., van Praag, B. M. S., and Theodossiou, I. (2011), Vignette equivalence and response consistency: the case of job satisfaction, IZA Discussion Paper 6174.
Fleurbaey, M. and Blanchet, D. (2013), *Measuring Welfare and Assessing Sustainability*, Oxford University Press, Oxford.
Fleurbaey M. and Maniquet, F. (2006), Fair income tax, *Review of Economic Studies 73*, 55–83.
Fleurbaey, M. and Schwandt, H. (2015), Do people seek to maximize their subjective well-being?, IZA Discussion Paper 9450.
Frijters, P. (2000), Do individuals try to maximize satisfaction with life as a whole, *Journal of Economic Psychology 21*, 281–304.
Frijters, P., Greenwell, H., Shields, M. A., and Haisken-De New, J. P. (2009), How rational were expectations in East Germany after the falling of the wall?, *Canadian Journal of Economics 42*, 1326–1346.
Glaeser, E. L., Gottlieb, J. D., and Ziv, O. (2016), Unhappy cities, *Journal of Labor Economics 34* (S2), S129–S182.
Golden, L. and Altman, M. (2008), Why do people overwork? Oversupply of hours of labour, labor market forces and adaptive preferences, in R. Burke and C. Cooper (eds), *The Long Work Hours Culture: Causes, Consequences and Choices*, Emerald Publishing, Bingley, 61–83.
Haan, P., Prowse, V. L., and Uhlendorff, A. (2008), Employment effects of welfare reforms: evidence from a dynamic structural life-cycle model, IZA Discussion Paper 3480.
Hamermesh, D. S. and Slemrod, J. B. (2008), The economics of workaholism: we should not have worked on this paper, *B.E. Journal of Economic Analysis & Policy 8* (1).
Ilmakunnas, S. and Pudney, S. (1990), A model of female labor supply in the presence of hours restrictions, *Journal of Public Economics 41*, 183–210.
Kahneman, D. and Krueger, A. B. (2006), Developments in the measurement of subjective well-being, *Journal of Economic Perspectives 20* (1), 3–24.
Kahneman, D. and Sugden, R. (2005), Experienced utility as a standard of policy evaluation, *Environmental and Resource Economics 32*, 161–181.
Kahneman, D. and Thaler, R. (2006), Utility maximization and experienced utility, *Journal of Economic Perspectives 20* (1), 221–234.

Kahneman, D., Wakker, P., and Sarin, R. (1997), Back to Bentham? Explorations of experienced utility, *Quarterly Journal of Economics 112*, 375–405.

Kahneman D., Krueger, A., Schkade, D. A., Schwarz, N., and Stone, A. (2006), Would you be happier if you were richer? A focusing illusion, *Science 312*, 1908–1910.

Kimball, M. and Willis, R. (2006), Utility and happiness, University of Michigan, Working Paper.

Knabe, A. and Rätzel, S. (2010), Income, happiness, and the disutility of labour, *Economics Letters 107*, 77–79.

Köszegi, B. and Rabin, M. (2008), Choices, situations, and happiness, *Journal of Public Economics 92*, 1821–1832.

Krueger A. B. and Schkade, D. A. (2008), The reliability of subjective well-being measures, *Journal of Public Economics 92*, 1833–1845.

Loewenstein, G. and Adler, D. (1995), A bias in the prediction of tastes, *Economic Journal 105*, 929–937.

Loewenstein, G., O'Donoghue, T., and Rabin, M. (2003), Projection bias in predicting future utility, *Quarterly Journal of Economics 118*, 1209–1248.

Mundlak, Y. (1978), On the pooling of cross section and time series data, *Econometrica 46*, 69–85.

Odermatt, R. and Stutzer, A. (2019), (Mis-)Predicted subjective well-being following life events, *Journal of the European Economic Association 17*, 245–283.

Odermatt, R., Powdthavee, N., and Stutzer, A. (2021), Are newly self-employed overly optimistic about their future well-being?, *Journal of Behavioral and Experimental Economics 95*, 101779.

Oswald, A. and Wu, S. (2010), Objective confirmation of subjective measures of human well-being: evidence from the USA, *Science 327*, 576–579.

Perez-Truglia, R. (2015), A Samuelsonian validation test for happiness data, *Journal of Economic Psychology 49*, 74–83.

Petrongolo, B. (2004), Gender segregation in employment contracts, *Journal of the European Economic Association 2*, 331–345.

Ravallion, M. and Lokshin, M. (2001), Identifying welfare effect from subjective questions, *Economica 68*, 335–357.

Rayo L. and Becker, G. S. (2007), Evolutionary efficiency and happiness, *Journal of Political Economy 115*, 302–337.

Senik, C. (2005), Income distribution and well-being: what can we learn from subjective data?, *Journal of Economic Surveys 19*, 43–63.

Stewart, M. and Swaffield, J. (1997), Constraints on the desired hours of work of British men, *Economic Journal 107*, 520–535.

Stutzer, A. and Frey, B. S. (2008), Stress that doesn't pay: the commuting paradox, *Scandinavian Journal of Economics 110*, 339–336.

van Praag, B. M. S., Frijters, P., and Ferrer-i-Carbonell, A. (2003), The anatomy of subjective well-being, *Journal of Economic Behavior and Organization 51*, 29–49.

van Soest, A. (1995), Structural models of family labor supply: a discrete choice approach, *Journal of Human Resources 30*, 63–88.

Viitanen, T. (2005), Cost of childcare and female employment in the UK, *Labour 19* (Special Issue), 149–179.

Wichert, L. and Pohlmeier, W. (2010), Female labor force participation and the big five, ZEW Discussion Paper 10-003.

First version submitted July 2017;
final version received June 2023.