Survey Evidence on Habit Formation

Jiannan Zhou*

December 25, 2019

(Click here for the latest version)

Abstract

Models with habit formation have been used to explain many important economic phenomena, such as the equity premium puzzle, excess sensitivity and smoothness of consumption, and causal effect of high growth on high saving. The literature, however, disagrees or is uninformative regarding the micro evidence of habit formation, which has led to controversies over its existence, specification, and implications. To address this gap, I designed and fielded a survey eliciting ten preference parameters of habit formation. My estimates show that both internal and external habits exist, with the latter accounting for about 17% of habit. Adjustment and cognition costs do not explain habit formation. Habit depreciates by around two thirds per year. Habit formation affects us about as much as does keeping up with the Joneses. Proposing and implementing four tests for the preference of habit formation, I find that both the additive and multiplicative specifications, ubiquitous in the literature, are rejected. Finally, I show that habit formation, when combined with keeping up with the Joneses, could explain the Easterlin paradox. The explanation implies that even if happiness eventually stops growing with income, continued income growth is still necessary to *maintain* happiness.

JEL: E21, G12, D60, I31.

Keywords: habit formation, micro evidence, keeping up with the Joneses, Easterlin paradox.

*Department of Economics, University of Colorado Boulder, 256 UCB, Boulder, CO 80309 (email: jiannan.zhou@colorado.edu). I am indebted to Professors Alessandro Peri, Xingtan Zhang, Martin Boileau, and especially Miles Kimball for their advice and support. I also thank Professors Richard Mansfield, Carlos Martins-Filho, Adam McCloskey, Sanjai Bhagat, Scott Savage, Terra McKinnish, Tony Cookson, Nathalie Moyen, Scott Schuh, Jin-Hyuk Kim, Nicholas Flores, Charles de Bartolomé, and Shuang Zhang, as well as numerous seminar and conference participants for helpful comments.

1 Introduction

Habit formation[1] has helped explain important phenomena in many areas of economics, such as asset pricing, business cycles, and economic growth.[2] However, the literature is uncertain about its micro evidence (Cochrane, 2017). As a result, controversies arise around the existence, specification, and implications of habit formation, leaving open the foundational question of whether, how, and when to use habit formation in economics. To answer this question, I use direct survey measurement of preferences (Barsky et al., 1997) to provide a new and extensive set of micro evidence for habit formation, as follows.

First, habit formation exists, as a phenomenon distinct from adjustment and cognition costs. Depending on the source of habit, habit formation can be categorized into internal habit formation (habit based on oneself's past behavior) and external habit formation (habit based on other people's past behavior). Literature investigating evidence of habit formation has mostly focused on the existence of internal habit formation and has drawn mixed conclusions (see column 1 of Table 1). Studies based on macrodata tend to find significant evidence of internal habit formation, whereas those based on microdata find it to be relatively less significant. Existence of external habit formation has received much less attention (column 2 of Table 1), although numerous theoretical papers assume its existence (see, e.g., Abel, 1990; Campbell and Cochrane, 1999; Smets and Wouters, 2007; Dou et al., 2017). Without a consensus on the existence of habit formation, researchers have begun investigating whether habit formation is capturing other phenomena. Flavin and Nakagawa (2008) proposed adjustment costs as a deeper explanation for habit formation. Matyskova et al. (2019) argued for habit formation as a way to economize cognition costs. In this paper, I provide evidence for the existence of both internal and external habit formation and evidence that habit formation cannot be explained by adjustment and cognition

[1] Throughout this paper, habit refers to total real consumption habit. There is a literature on deep habit where people form habits of individual varieties of consumption, which is not the focus of this paper. It is also worth noting that habit formation, as modeled by current economics and finance literature, captures the phenomenon of response decrement to repetitive stimulation, which differs from the day-to-day notion of the cue-routine-reward habit. Because it is already an accepted term in the literature, I will continue referring to this phenomenon as habit formation. This definition of habit formation sets it apart from path dependence, with which it is sometimes confused. It is also worth noting that habit formation differs from desensitization, which would imply reduced response to small changes. Habit formation increases response to small changes.

[2] For example, equity premium puzzle and stock market behavior (Constantinides, 1990; Campbell and Cochrane, 1999); excess smoothness and excess sensitivity of consumption (Fuhrer, 2000); causal effect of high growth inducing high saving (Carroll, Overland and Weil, 2000).

costs.[3] Few authors have studied the composition of internal and external habit formation (column 3 of Table 1). Allowing for habit to be formed both internally and externally as per Grishchenko (2010), I estimate that external habit accounts for a small portion (about 17%) of habit.

Second, habit depreciates by about two thirds per year. Most specifications of habit formation depend on two important parameters: habit depreciation rate[4] and habit intensity. Existing research has focused predominantly on estimating the habit intensity parameter (Havranek, Rusnák and Sokolova, 2017; column 4 of Table 1) while treating the habit depreciation rate largely as a free parameter. Two potential reasons for the current state of the literature are a lack of recognition of the importance of the habit depreciation rate parameter and the massive disagreement between its few existing micro and macro estimates (column 5 of Table 1). To illustrate this parameter's importance, I show that simply changing its value can significantly affect the performance of habit formation models. I also provide a micro estimate of this parameter aggregated to macro level, which is much closer to existing macro estimates than present micro estimates.

Third, neither the additive nor multiplicative habit preference is consistent with people's behavior. Almost all current habit formation models in the literature assume either one of these two habit utility functions (column 6 of Table 1). The conclusions drawn from these studies are, therefore, joint tests and estimates with these specifications of unknown validity. In a general utility function naturally nesting these two formulations, I propose and implement four tests of the micro validity of the preference specifications. I find that both types of habit utility function fail the tests by large margins, casting doubt on the validity of the predictions from existing habit formation models. Even though these two common specifications are rejected, estimates of the signs of utility derivatives in the general habit formation preference are consistent with the idea of habit formation,[5] suggesting that specifications of habit formation preferences that are consistent with the survey evidence could be found.[6]

Fourth, the effect of habit formation on utility is about as strong as that of keeping up with the Joneses. As two important interdependent preferences, keeping up with

[3]Biologists have found evidence of habit formation in humans and all animals, including single-celled Amoebas that do not have a nervous system (Folger, 1926). My result implies that people form habits for reasons beyond adjustment and cognition costs. For example, Rayo and Becker (2007) argued from the evolutionary perspective that habituating living standards increases our motivation to strive for more, which potentially helps with survival.

[4]I focus on depreciation rate rather than catch-up rate because the latter varies under different normalizations of habit whereas the former is invariant to such normalizations.

[5]That is, $u_{CH} > 0$ and $u_{HH} < 0$.

[6]I explore this possibility in another paper.

the Joneses allows interpersonal dependence, while habit formation allows intertemporal dependence (in internal and external habits) as well as interpersonal dependence (in external habit). Previous researchers have found a strong effect of keeping up with the Joneses (Luttmer, 2005; Lewbel et al., 2018; De Giorgi, Frederiksen and Pistaferri, 2019) but disagreed on the relative strength of these two phenomena. Ravina (2007) found internal habit formation to be about 70% stronger than keeping up with the Joneses, while Alvarez-Cuadrado, Casado and Labeaga (2015) estimated internal habit formation to be as strong as keeping up with the Joneses. Allowing both internal and external habit formation, I provide an estimate of the relative strength of the two phenomena without taking a potentially problematic stance on the functional forms of the utility function.

Fifth, habit formation combined with keeping up with the Joneses could generate the happiness–income pattern of the Easterlin paradox. Easterlin (1973, 1974) highlighted the tension between cross-sectional positive correlation and time-series zero correlation between happiness and income and proposed keeping up with the Joneses as an explanation for its effect on averaging happiness. As happiness data accumulate over time, the literature discovers that the zero time-series correlation tends to hold only in the long run while the short-run correlation is generally positive (Stevenson and Wolfers, 2008; Sacks, Stevenson and Wolfers, 2012; Easterlin, 2017).[7] Habit formation has been proposed as a potential explanation for the original version of the paradox[8] (Easterlin, 1995) and, in particular, for the relatively newly discovered temporal heterogeneity of the correlation (Clark, Frijters and Shields, 2008; Clark, 2016). To the best of my knowledge, evidence on whether habit formation can actually fit this particular shoe is absent from the literature. Utilizing my aforementioned extensive set of evidence on habit formation, I conducted a semistructural simulation and found that, when coupled with keeping up with the Joneses, habit formation can explain the observed patterns of income and happiness across all the dimensions: cross-section, short-run, and long-run.

To illustrate the intuition of the explanation, I propose an analogy that I call "running against an escalator." Imagine running, with a uniform speed, against an escalator that initially is still but once you step onto it will gradually accelerate to

[7] I am aware of the controversy around the existence of the paradox. My purpose here is not to defend the paradox, as I do not provide new evidence on happiness measures, but to see how habit formation (and keeping up with the Joneses) affect the relationship between income (or consumption) and happiness.

[8] Kimball and Willis (2006) proposed as a potential explanation of the paradox that happiness is time-intensive. They also survey several other potential explanations. In this project, I focus on the effects of habit formation and keeping up with the Joneses.

your running speed. The total stairs you run and the elevation you reach represent your income and happiness, respectively, while the escalator symbolizes the happiness effect of habit formation and keeping up with the Joneses. For a while after you step onto the escalator, your elevation changes in the same direction as that of the vertical component of your running velocity, implying a positive correlation between total stairs run and elevation, just like the positive happiness–income gradient in the short run. After the escalator catches up with you, your elevation stops changing even though you keep running, implying a zero correlation between total stairs run and elevation, just like the long-run nil (or low) happiness–income gradient. People who run faster plateau at higher elevations, implying a positive correlation between total stairs run and elevation, just like the cross-section positive happiness–income gradient. This analysis implies that even if happiness eventually stops growing with income, continued income growth is still necessary to maintain happiness. In the language of the analogy, keeping running is necessary to maintain elevation.

Providing this new and extensive set of micro evidence for habit formation involves confronting two main challenges. First, the framework needs to be general enough to nest all current habit formation models that are heterogeneous along many dimensions while remaining agnostic about the existence, specification, and implications of habit formation. To tackle this difficulty, I took a general habit formation model with assumptions of minimal interference with the evidence I am providing. In the general habit formation model, I identified ten preference parameters[9] that are instrumental to the above set of evidence regarding habit formation. The second challenge is that the estimation of all the preference parameters of interest requires variations that, to the best of my knowledge, do not exist in reality and are very hard, if not impossible, to approximate using conventional methods and data under the required generality of the framework.[10] To deal with this challenge, I designed thought experiments to elicit the preference parameters, implemented the thought experiments in a survey, and fielded two waves of the survey on Amazon Mechanical Turk. The application of this method in economics can be traced back at least to Thurstone (1931) and has been widely used in eliciting reduced-form preference parameters in, among others, environmental economics, experimental economics, and health economics (Johnston et al., 2017; Ameriks et al., 2019; Jha and Shayo, 2019). The use of this method in

[9]They are habit depreciation rate, time discount rate, external habit mixture coefficient, all ratios of utility derivatives up to the second order, a measure of the relative strength of habit formation and keeping up with the Joneses, and two quantities concerning the existence of internal and external habit formation.

[10]The required variation for estimating each parameter is different from that of another parameter. See section 4 for the required variations.

eliciting structural preference parameters started at least with Barsky et al. (1997) and has been commonly seen in the literature (see, e.g., Kapteyn and Teppa, 2003; Sahm, 2007; Kimball, Sahm and Shapiro, 2008; Kimball and Shapiro, 2008; Kimball, Sahm and Shapiro, 2009; Benjamin et al., 2014; Kimball et al., 2015; Benjamin et al., 2019).

Though subject to potential response biases and errors, the method of direct survey measurement of preferences is designed to overcome the problems of the conventional method—such as weak and other identification issues, and measurement error and other data problems (Kimball and Shapiro, 2008). This method is especially helpful in the context of habit formation where a lack of required variations in conventional data has confined current literature to mostly studying three, barely touching two other, and completely ignoring the rest half of the ten preference parameters this paper estimates, all of which are crucial in addressing the aforementioned controversies over habit formation. Even if future conventional method allows estimation of all the preference parameters, the survey evidence will remain important, because the limitations of the conventional method tend to be orthogonal to the limitations of the survey method, implying that the evidence from either method complements that of the other. Response biases and errors can be and have been carefully studied and dealt with through defining, measuring, and finding ways to address them. In this paper, I deal with potential response biases and errors through the design and implementation of the thought experiments, survey, estimation, and robustness checks.

For accuracy, I jointly estimated all the parameters while addressing potential response biases and errors. To deal with the resulting high dimensionality of the estimation, I use a Bayesian method that bypasses the computational burden of direct optimization of maximum likelihood estimation. In particular, I employ Hamiltonian Monte Carlo, a Markov Chain Monte Carlo method that enjoys state-of-the-art sampling efficiency in high dimensions. My implementation of the estimation establishes direct equivalence between one of my point estimators and the maximum likelihood estimator.

In addition to providing the new and extensive set of micro evidence on habit formation, this paper also contributes methodologically to the literature. The majority of the current research elicits structural preference parameters in preferences with fully specified functional forms. To be immune to misspecification errors, some authors have dispensed with the functional form assumptions and used first-order approximations in eliciting parameters in preferences of unspecified forms (see, e.g., Benjamin et al., 2014). In this paper, I show that this approach is a special case of a general approach where one could use any, even the infiniteth, order of approximation. This allows for drastic reductions in approximation errors and improvements in the accuracy of the elicitation of preference parameters.

I present the general habit formation model and the survey instrument in section 2. I then summarize the data and the statistical model that is applied to the data (section 3). Section 4 contains the elicitation and estimate of each preference parameter of interest and the implications of the estimates. I explore the explanation of the Easterlin paradox in section 5 and check the robustness of my results in section 6. Finally, section 7 concludes.

2 Methodology

2.1 Habit Formation Model

The breadth of the questions this paper intends to answer requires the habit formation model to be as general as possible so that it nests all specifications of interest. Generality means as few assumptions as possible. Still, enough structure must be put in place so that the model is useful for the purposes of this paper.

The agent maximizes

$$\mathbb{E}_0 \int_0^\infty e^{-\rho t} u\left(C\left(t\right), H\left(t\right)\right) dt$$

where C is individual spending, H habit, and ρ time discount rate. Henceforth the time index will be omitted when there is no ambiguity in doing so. I maintain expected utility and exponential time discounting.[11] As is ubiquitous in the literature, I require the utility function to be analytic and to satisfy positive monotonicity of consumption ($u_C > 0$) and diminishing marginal utility of consumption ($u_{CC} < 0$). These assumptions aid identification and estimation without interfering with the evidence this paper is providing. For example, I leave open whether habit affects the utility and the sign and magnitude of any derivative of the utility function with respect to habit. I allow the respondent's utility to depend on other variables (e.g., labor), but because they will be kept constant in the survey, abstracting from explicitly listing them as the arguments of the utility function results in no loss of generality. In the discussion of some survey questions where I do vary things other than self spending and habit (e.g., other people's spending), I will explicitly show the additional variables

[11]Nearly all current habit formation models make these two assumptions.

Table 1: Estimates of Habit Parameters in Selected Literature

Preference parameters in the table are from specializations of the following habit formation model:

$$u\left(C, H\right) = \begin{cases} v\left(C - \alpha H\right) & \text{Additive Habit} \\ v\left(C / H^{\alpha}\right) & \text{Multiplicative Habit} \end{cases} \quad \text{s.t. } \dot{H} = \theta\left(\left(1 - \omega\right) C + \omega C_{others} - H\right)$$

where α is habit intensity, θ habit depreciation rate, and ω external habit mixture coefficient.

Study	Internal Habit (1)	External Habit (2)	ω (3)	α (4)	θ[3] (5)	Additive or Multiplicative[5] (6)
A. Microdata						
Naik, Moore (1996)	Y	*(N)*	*(0)*	0.081^{***}	*(Y)*	*(A)*
Dynan (2000)	N	*(N)*	*(0)*	-0.038	*(Y)*	*(A)*
Guariglia, Rossi (2002)	N	*(N)*	*(0)*	-0.272^{***}	*(Y)*	*(A)*
Lupton (2002)	Y	*(N)*	*(0)*	0.225^{***}	9.2%/Y***	*(A)*
Kapteyn, Teppa (2003)	Y	*(N)*	*(0)*	0.777^{***}	[4]	*(M)*
Rhee (2004)	Y, N[1]	*(N)*	*(0)*	0.6	*(Y)*	*(A)*
Ravina (2007)	Y	N	0.025[2]	0.503^{***}	*(Q)*	*(A, M)*
Browning, Collado (2007)	Y, N[1]	*(N)*	*(0)*	0.01-0.14	*(Q)*	*(A)*
Alessie, Teppa (2010)	Y	*(N)*	*(0)*	0.211^{**}	*(Y)*	*(A)*
Iwamoto (2013)	N	*(N)*	*(0)*	-0.3787^{***}	*(Y)*	*(A)*
Khanal et al. (2018)	Y	*(N)*	*(0)*	0.545^{***}	*(Y)*	*(A)*
⋮ ($\geq$ 10 studies)	⋮	⋮	⋮	⋮	⋮	⋮
B. Macrodata						
Ferson, Constantinides (1991)	Y	*(N)*	*(0)*	$0.64\text{-}0.97^{***}$	*(M, Q, Y)*	*(A)*
Fuhrer (2000)	Y	*(N)*	*(0)*	0.80^{***}	99.9%/Q***	*(M)*
Stock, Wright (2000)	Y, N[1]	*(N)*	*(0)*	-	*(M, Y)*	*(A)*
Smets, Wouters (2003)	*(N)*	Y	*(1)*	0.573^{***}	*(Q)*	*(A)*
Lubik, Schorfheide (2004)	Y	*(N)*	*(0)*	0.57^{***}	*(Q)*	*(M)*
Christiano et al. (2005)	Y	*(N)*	*(0)*	0.65^{***}	*(Q)*	*(A)*
Smets, Wouters (2007)	*(N)*	Y	*(1)*	0.71^{***}	*(Q)*	*(A)*
Adolfson et al. (2007)	Y	*(N)*	*(0)*	0.650^{***}	*(Q)*	*(A)*
Korniotis (2010)	N	Y	0.79[2]	0.33[2]	*(Y)*	*(A)*
Grishchenko (2010)	Y	N	0.000^{***}	0.895[2]	70.7%/Q***	*(A)*
Altig et al. (2011)	Y	*(N)*	*(1)*	0.76^{***}	*(Q)*	*(A)*
⋮ ($\geq$ 65 studies)	⋮	⋮	⋮	⋮	⋮	⋮

Notes: The studies are selected based on citation, publication year, and number of parameters estimated. Characters in parenthesis and italics are assumed parameter values of the studies. 1: depends on goods or time horizon; 2: implied estimates; 3: M/Q/Y—habit depreciates fully at the end of a month/quarter/year; 4: geometric habit evolution speed: 0.071 (0.007); 5: A/M—additive/multiplicative habit. Significance levels: ** 5%, *** 1%.

in the utility function. For clarity, I will save the details for relevant contexts.

I assume that habit evolves according to

$$\dot{H} = \theta (C - H)$$

where θ is the habit depreciate rate. This formulation has been in the literature since at least Houthakker and Taylor (1970) and is the most commonly used habit evolution equation in the literature. Researchers have used slightly different formulations of the evolution. However, the difference consists in either simply rescaling of the unit of habit (e.g., Constantinides, 1990) or disappears in steady state (e.g., Campbell and Cochrane, 1999). For general habit evolutions that are potentially nonlinear, I show that they are observationally equivalent to this linear habit evolution under reasonable conditions.[12] I also choose this habit evolution for its natural unit which is the same as that of consumption. For example, if a person has been consuming $5,000 per month for as long as she can remember, then her current habit is also $5,000 per month.

To answer the motivating questions, I needed to see whether the utility function depends on habit. If habit exists, I also needed to know the values of the preference parameters, θ, ρ, ratios of utility derivatives up to the second order ($-\frac{u_H}{u_C}$, $\frac{Hu_{HH}}{u_H}$, $\frac{u_{CH}}{u_{HH}}$, and $\frac{u_{CC}}{u_{HH}}$), and, in the extended models,[13] external habit mixture coefficient and strength of habit formation relative to keeping up with the Joneses.

2.2 Survey Overview

To elicit the preference parameters, I specified simple thought experiments that identify the preference parameters of interest while controlling for potential confounding factors and response biases and errors. I implemented these thought experiments in a survey.

My survey starts with a preamble module that specifies the basic hypothetical situation and instructs the respondents on the survey's format. Nine core modules follow to elicit the preference parameters of interest.

The hypothetical situation is designed to be as simple as possible while still allowing the elicitation of my objects of interest and avoiding potential confounding factors that plague traditional data. The key is not realism but simplicity. If the situation is

[12] See section B of the appendix for the proof. The basic intuition of the proof is that the nonlinearity can be "transported" to the utility function that is agnostic of how habit affects it.

[13] See sections 4.4 and 4.5.

simple enough, although hypothetical, people would be able to understand it. In particular, I wanted to free respondents from worrying about changes to the purchasing power of their money, about durable goods, and about changes in preferences. The exact wording of the hypothetical situation in the survey is the following.

> Please answer all survey questions under the following hypothetical situation:
>
> - There is no inflation, and prices of everything stay the same over time.
> - You rent the durable goods you consume, including residence, furniture, car, etc.
> - Things you want don't change over time.
> - People not mentioned in questions always spend $5,000 per month.
> - Everything else unmentioned in the questions is and stays the same.

I tested respondents' understanding of this hypothetical situation. Only those who passed the test were able to proceed to the remaining modules in the survey.

The respondents did not know that they were answering a survey about habit formation. I told the respondents that the survey was about spending behavior. I did this for two reasons, the first of which was to avoid potential confusion. More likely than not, a typical respondent would not know what the consumption habit is, as we economists call it. The second reason was to avoid potential biases. I could not prime respondents with habit formation while attempting to test its very existence.

Because I am interested in how habit affects people's well-being and responds to spending changes, I varied people's past and future spending profiles to elicit the preference parameters of interest. I created variations in both past and future spending paths because habit reveals itself in the way that the past affects the future. Because income affects utility through spending, I did not specify the income process except to tell the respondents that they could afford the spending profiles in the survey.

To make a representation of a spending profile intuitive and to simplify comparison across several such spending profiles, I drew a spending path in, as I call it, a monthly spending graph (Figure 1). In the figure, time is on the horizontal axis: past on the left, now in the middle, and future on the right. The bars on the graph represent monthly spending and are drawn to scale and colored differently to help distinguish time horizons. In the example in Figure 1, the spending path represents spending

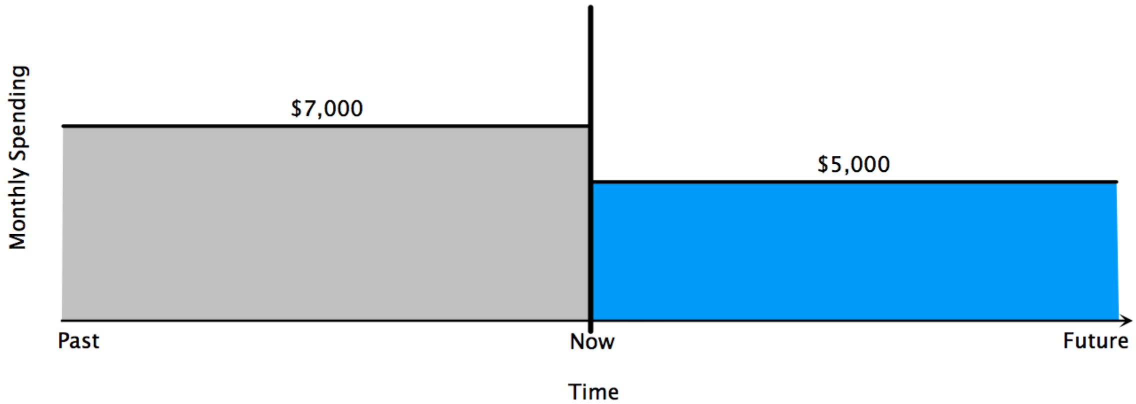


Figure 1: A typical monthly spending graph.

$7,000 per month in the past until now and $5,000 per month in the future starting now. The respondents underwent instruction and were tested on reading the monthly spending graphs before answering questions in the core modules.[14]

To alleviate the concern that in reality each person has only one past spending path, I invoked the metaphor of parallel universes, in which everything is the same except for the spending profiles. I then asked the respondents which universe brought them a better future experience (i.e., how they feel in the future, starting now).

Despite their advantages, surveys can be subject to response errors and biases. Nevertheless, it is always good to have more perspectives on the same question. This is especially valuable when traditional methods suffer from major limitations (e.g., alternative explanations) and when papers using them result in conflicting conclusions. I built into the survey various details to minimize response errors and biases (see Figures 2 and 3 for screenshots of a typical survey question). For example, to reinforce the idea that the only variation between universes is in spending paths, I emphasized it at the start of every set of questions.[15] To help the respondents compare different spending paths, I told them in what time horizon the paths differed. To help them distinguish past experience from future experience, I allowed them to express views on both experiences. I also reiterated definitions of the experiences of interest and highlighted the key words—past and future—to further remind the respondents of which experience the question asked about. To avoid respondents accidentally clicking an option different from the one they wanted to choose, I integrated the monthly

[14] For details on this instruction, see section E.1 of the appendix.

[15] A set of questions corresponds to one or two preference parameters of interest and usually consists of 2 to 3 subquestions.

spending graphs into the clickable choices. To help them confirm that they answered as they intended, I slightly darkened the background of an option when their mouses hovered over it and completely darkened the background of the option they finally selected (Figure 4).

To make sure that my respondents were paying attention, I used a set of attention checks, ranging from the explicit, like the quiz on the basic hypothetical situation at the start of the survey, to the implicit, such as time spent on the survey, a quiz on the basic hypothetical situation at the end of the survey, and cross-wave consistency of their answers to the demographic questions. To encourage effort, I told the respondents about the existence of such attention checks but did not tell them where they were or how to identify them. In addition, to encourage more effort, I told them in the survey's introduction that respondents whose responses were of high quality would be entered in a small ($1) lottery with the winning odds of 1 in 100.[16]

To further minimize potential response errors and biases, I conducted two waves of the survey, changing the sequence of core modules and reversing the order of choices for questions in the second wave. I dealt with remaining response errors and biases through a statistical model and robustness checks.[17]

The survey responses do not tell me the values of the preference parameters, but give me information with which I can uncover the values of the preference parameters. In essence, the uncovering process is a mathematical proof that translates survey responses into values of the preference parameters of interest. Based on the accuracy of the uncovering process, I distinguish different orders of elicitation. For example, first-order elicitation means that the approximation error of the uncovering process will be up to the remainder of a first-order Taylor expansion, while exact elicitation means no approximation error. In eliciting some of the preference parameters under the general habit model, approximation errors are inevitable. However, my estimates show that even first-order elicited preference parameters can be very close to those estimated using more parametric methods.[18] The elicitation varies from one preference parameter to another.

[16]Of the 515 responses I collected, six respondents won this small lottery.

[17]See sections 3 and 6 for details.

[18]See section 4.5 for details.

- Imagine that Universes One and Two are identical except your monthly spending in the 'past'.
- Remember that past experience is how you felt about the 'past' until 'now'.

Which universe gave you a better **past** experience?

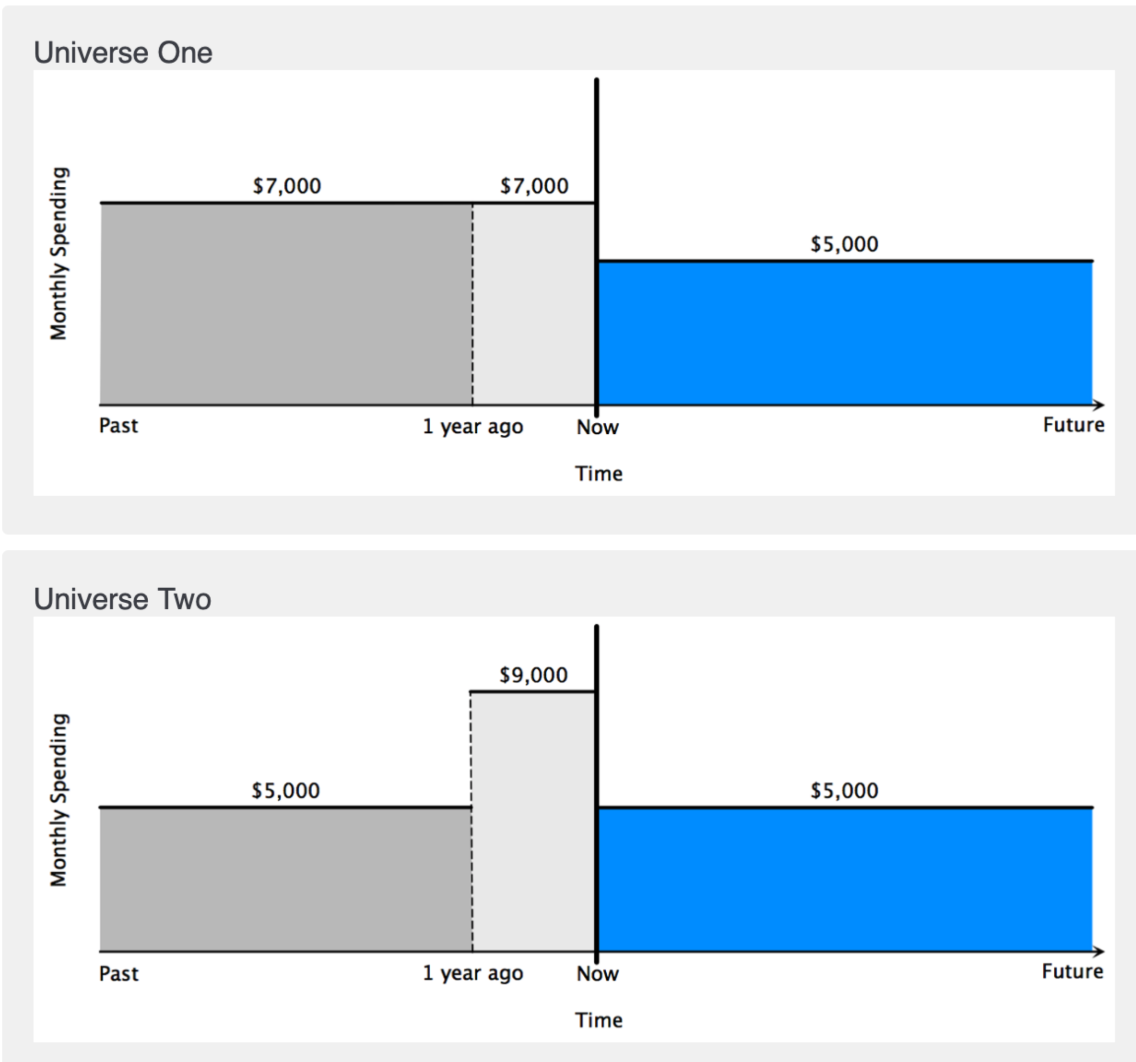
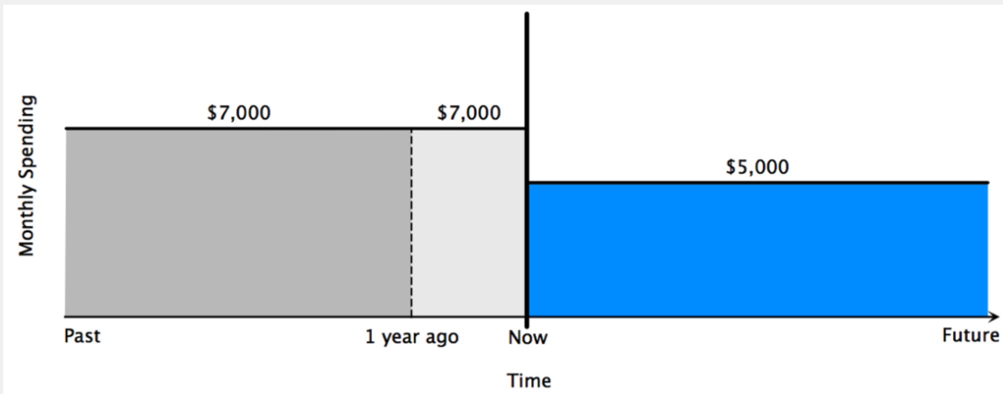


Figure 2: A typical survey question (part 1 of 2).

- Remember that future experience is how you feel about the 'future' starting 'now'.

Which universe will give you a better **future** experience?

Universe One



Universe Two

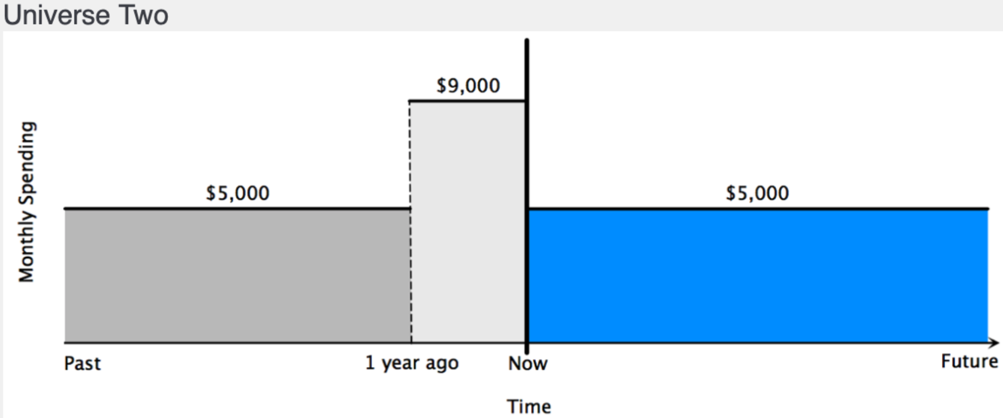


Figure 3: A typical survey question (part 2 of 2).

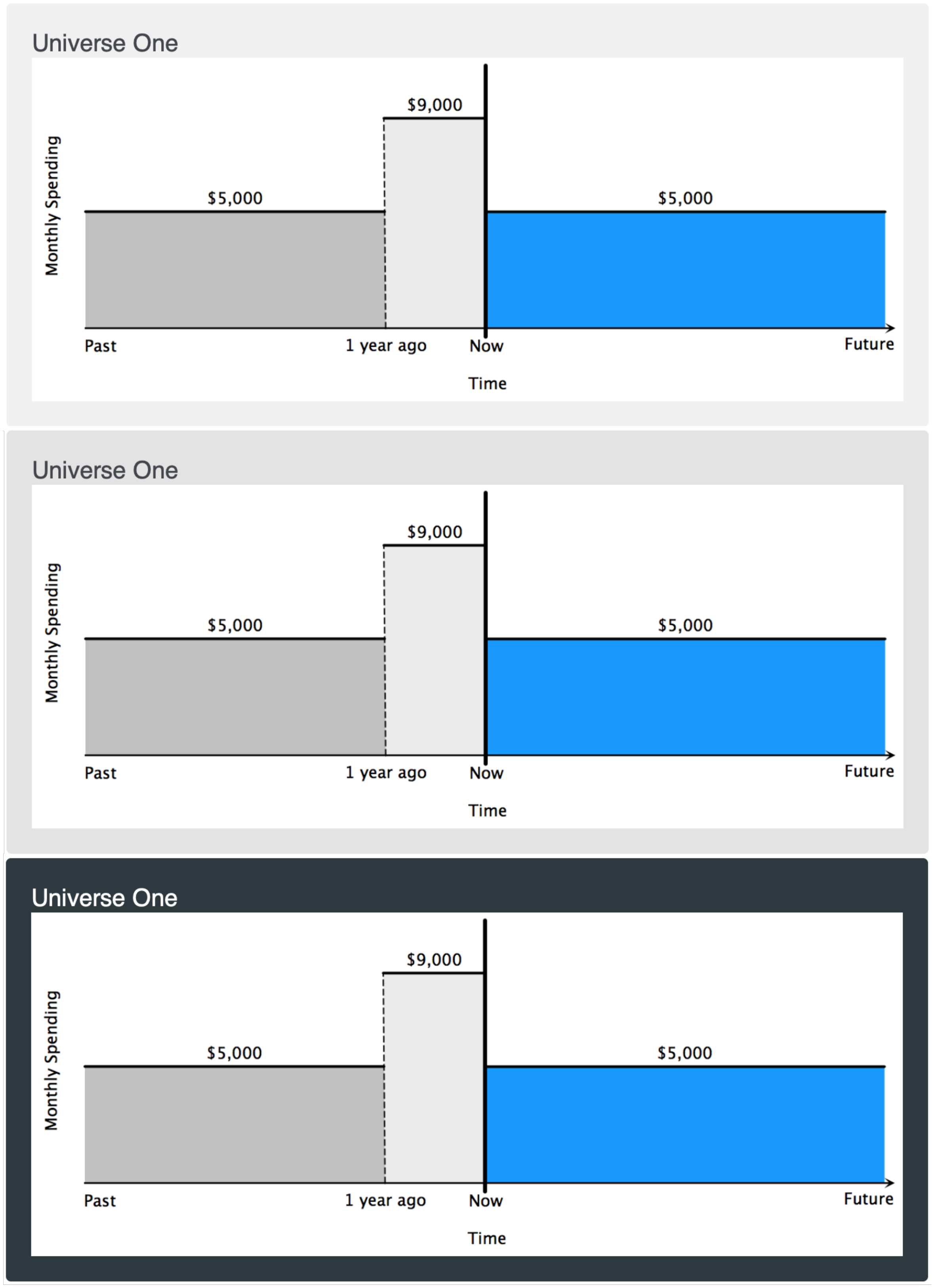


Figure 4: Option apperances (top to bottom: initial, hover, click).

3 Data and Statistical Model

3.1 Data

I conducted two waves of the survey on Amazon Mechanical Turk.[19] Although this sample is less representative of the U.S. population than national probability samples such as the Health and Retirement Study and Panel Study of Income Dynamics, it is more representative than lab samples (Berinsky, Huber and Lenz, 2012). Johnson and Ryan (2018) showed that this sample can provide consistent and economically meaningful data. Many economics papers have used this sample (see, e.g., Oster, Shoulson and Dorsey, 2013; Kuziemko et al., 2015; Bordalo et al., 2016; Cavallo, 2017; Martínez-Marquina, Niederle and Vespa, 2019; Benjamin et al., 2019).

I restricted my respondents to U.S. residents. From the 301 first-wave respondents who expressed a willingness to participate in future studies, I randomly invited 200 to participate in the second wave and got a response rate of about 75%. After excluding respondents located outside of the United States who submitted duplicate responses, or who sped through the survey, I was left with 348 and 140 valid responses from the two waves, respectively, from respondents across the United States (Figure 5).

In the survey, I collected demographic information on age, gender, household income, and household size. These sample characteristics are very close to their national counterparts (Table 2). At the time of the survey, a typical respondent was about 38 years old, lived with another person or two people in a household of an annual income of the range \$50,000 to \$60,000, was slightly more likely to be female if participating in only the first wave and slightly more likely to be male if participating in both waves, and spent less than half an hour on the survey.

The survey has nine core modules, each corresponding to one or two preference parameters of interest. Eight of the ten elicited preference parameters are identifiable to scale and estimated jointly for accuracy. As a result, response frequencies of individual parameters alone are not particularly informative and are, therefore, relegated to the appendix (Table 25). The response frequencies of parameters identifiable only to sign are reported in section 4.

3.2 Statistical Model

The preference parameters are estimated jointly using Hamiltonian Monte Carlo. The statistical model underlying the estimation addresses response errors not addressed

[19]The first wave was conducted on July 23, 2018, and the second wave on August 11, 2018.

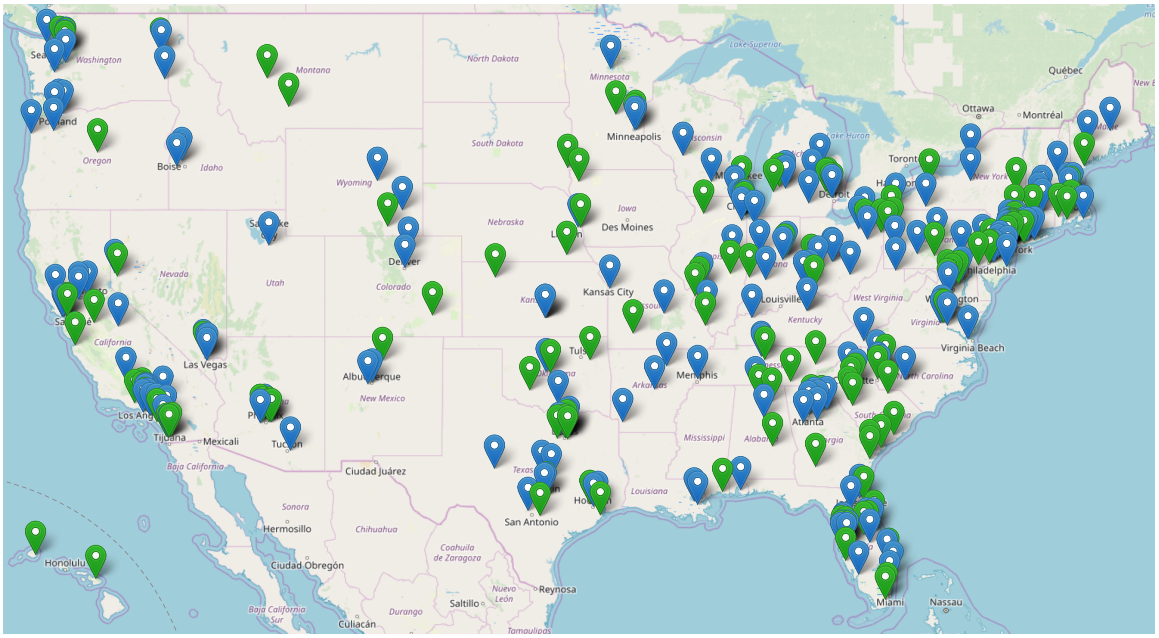


Figure 5: Locations of respondents. Green (lighter shaded in black and white) markers indicate locations of respondents participating in both waves, whereas blue (darker shaded in black and white) markers indicate locations of respondents participating only in the first wave.

Table 2: Demographic Statistics

	First Wave (348 obs.)	Second Wave (140 obs.)	United States
Age, median	38	37	38
Household income, median	\$50,000–\$60,000	\$50,000–\$60,000	\$57,652
Female percentage	53.2%	47.9%	50.8%
Household size, mean	2.66	2.70	2.63
Time on survey, mean	28’6”	25’32”	

Note: Household income is annual.
Source: For the last column, U.S. Census Bureau—2018 Population Estimates (for age and female percentage), 2017 American Community Survey, and 2017 Puerto Rico Community Survey (for household income and size).

by the survey design or by eliminating invalid responses from the sample. Potential remaining response biases were dealt with through robustness checks (section 6)

I model an observed response for preference parameter x of individual i in wave w, $X_{i,w}$, as

$$X_{i,w} = \sum_k k \cdot 1\left(T_{k,\tilde{x}} \leq \tilde{x}_{i,w} \leq T_{k+1,\tilde{x}}\right)$$

where the unobserved parameter value

$$\tilde{x}_{i,w} = x_i + \varepsilon_{i,x,w},$$

and $T_{\{k\},\tilde{x}}$ is the sequence of known thresholds informed by the elicitation of the parameter. The true parameter value for individual i, x_i, is drawn from $\mathcal{N}(\mu_x, \sigma_x^2)$. The individual-parameter-wave-specific response error $\varepsilon_{i,x,w}$ is drawn from $\mathcal{N}(0, \varsigma_{x,w}^2)$ independently of the true parameter value. In the robustness section, I allow the mean of the response errors to be nonzero and to vary across waves. Estimates of the means of the more general response errors are indistinguishable from zero, and the estimates of the preference parameters are not significantly different from those under the zero-mean response error specification here. I assume that the parameters are drawn independently within a respondent. Because my respondents were spread across the United States (Figure 5) and most likely did not know each other, I assume that responses are independent across respondents.

Allowing the response error to be persistent across waves (i.e., $Cov\left(\varepsilon_{i,x,1}, \varepsilon_{i,x,2}\right) = \sigma_{\varepsilon_x}^2 > 0$ and $\varsigma_{x,w}^2 = \sigma_{\varepsilon_x}^2 + \sigma_{\varepsilon_{x,w}}^2$), I arrive at the joint distribution of respondent i's parameter x in the two waves of the survey:

$$\begin{bmatrix} \tilde{x}_{i,1} \\ \tilde{x}_{i,2} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_x \end{bmatrix}, \begin{bmatrix} \sigma_x^2 + \sigma_{\varepsilon_x}^2 + \sigma_{\varepsilon_{x,1}}^2 & \sigma_x^2 + \sigma_{\varepsilon_x}^2 \\ \sigma_x^2 + \sigma_{\varepsilon_x}^2 & \sigma_x^2 + \sigma_{\varepsilon_x}^2 + \sigma_{\varepsilon_{x,2}}^2 \end{bmatrix} \right)$$

.

In section A of the appendix, I prove that the individuals' parameter values aggregate to the mean for the representative agent. That is,

$$x_R = \frac{1}{N} \sum_i x_i,$$

where x_R is the value of the representative agent's parameter x. Given that almost all current habit formation models assume a representative agent, I will focus on the implication of my estimates for a representative agent model with habit formation. Because $x_R = \mu_x$, the estimate of interest is that of μ_x.

Regarding the implementation of Hamiltonian Monte Carlo, I chose uniform distributions as priors, not only to let data speak as much as possible but also to establish the equivalence between my maximum a posteriori (MAP) estimates and maximum likelihood estimates. I ran ten Markov chains initialized from random diffuse starting points and collected 15,000 iterations of warmup and 25,000 draws of sampling. I report all three Bayesian point estimators[20] and the highest posterior density or mass interval (HPDI or HPMI).

4 Elicitation and Estimation

In this section, I show how the design of the survey and the estimates of preference parameters answer the questions on the micro evidence of habit formation. I break discussion of the existence of internal habit formation and external habit formation into two parts, because the latter involves model augmentation that is best done after discussing survey questions that do not require it. The later model augmentation does not affect the results preceding it.[21] First, let us look at the existence of internal habit formation, habit depreciation speed, and additive versus multiplicative habits. Then, we will turn to the existence of external habit formation, composition of the two habits, and relative strength of habit formation and keeping up with the Joneses.

4.1 Existence of Internal Habit Formation

The fundamental characteristic of habit formation is the diminishing response to a repetitive stimulus. In the case of internal habit formation, the higher a person's past consumption (stimulation), the lower her future utility (response). As a measure of the intensity and persistence of past stimulation, habit increases with past consumption. That is, habit formation is consistent with $Q_H \equiv u(C, H + \Delta h) - u(C, H) < 0$ but not with $Q_H \geq 0$, for $\Delta h > 0$.

To elicit the sign of Q_H, I varied the respondent's past spending profile so that the respondent's current habit varied across the parallel universes. I chose a common level of future spending for all the universes so that the only difference between the universes was the current level of habit regarding future experience. It is worth emphasizing again that I did not prime the respondents with habit formation in the survey and that made no assumption about the signs of derivatives of the utility function with respect to habit.

[20] MAP together with posterior mean and median.

[21] For more discussion on this, see section 4.4.

Table 3: Response Frequency (Percentage) of Existence of Internal Habit Formation

Universe	1	2	3	4	5
1st Wave	56	3	9	1	30
2nd Wave	60	2	6	1	30

Table 4: Estimate of Sign of Q_H

	MAP	Mean	Median	99% HPMI
$\mathrm{sgn}\,(Q_H)$	-1.00	-1.00	-1.00	[-1.00, -1.00]

The survey question asked the respondents to pick the monthly spending graph that gave them a better future experience (Figure 6). In this context, preferring a universe with less past spending over one with higher past spending implies $Q_H < 0$. The responses to this question show that the average respondent chose Universe One (Table 3), consistent with the existence of habit formation for the representative agent. My estimate confirms this (Table 4).

By monotonicity of the utility function,[22] $Q_H < 0$ implies $u_H < 0$.

This survey question also has also enabled me to distinguish between adjustment costs and internal habit formation. Flavin and Nakagawa (2008) argued that the adjustment cost of consumption commitments matches data better and raises the possibility that adjustment cost is the structural explanation of habit formation. Chetty and Szeidl (2016) showed that the difference between the two models matters for welfare and policy because consumption commitment can be abandoned quickly, whereas habit cannot. To distinguish the two phenomena, imagine that only adjustment cost exists. As is currently modeled, one pays adjustment cost whenever she changes her level of consumption. It follows that Universe Three would give the best future experience because it involves the least change in consumption levels and thus the lowest adjustment cost. If the respondents think that there are no consumption commitments—perhaps because they rent durable goods—and therefore, that they pay no adjustment cost, their choice of one of the five universes would be uniformly distributed. Neither of these two patterns are supported by the responses in which most respondents chose Universe One; therefore, adjustment cost cannot explain habit formation.

Cognition costs also cannot generate the response pattern. If one models cognition costs as associated with coming up with the optimal spending path, the respondents should choose the five universes uniformly because the consumption path is already

[22] This is implied by the analyticity of the utility function (Debreu, 1972).

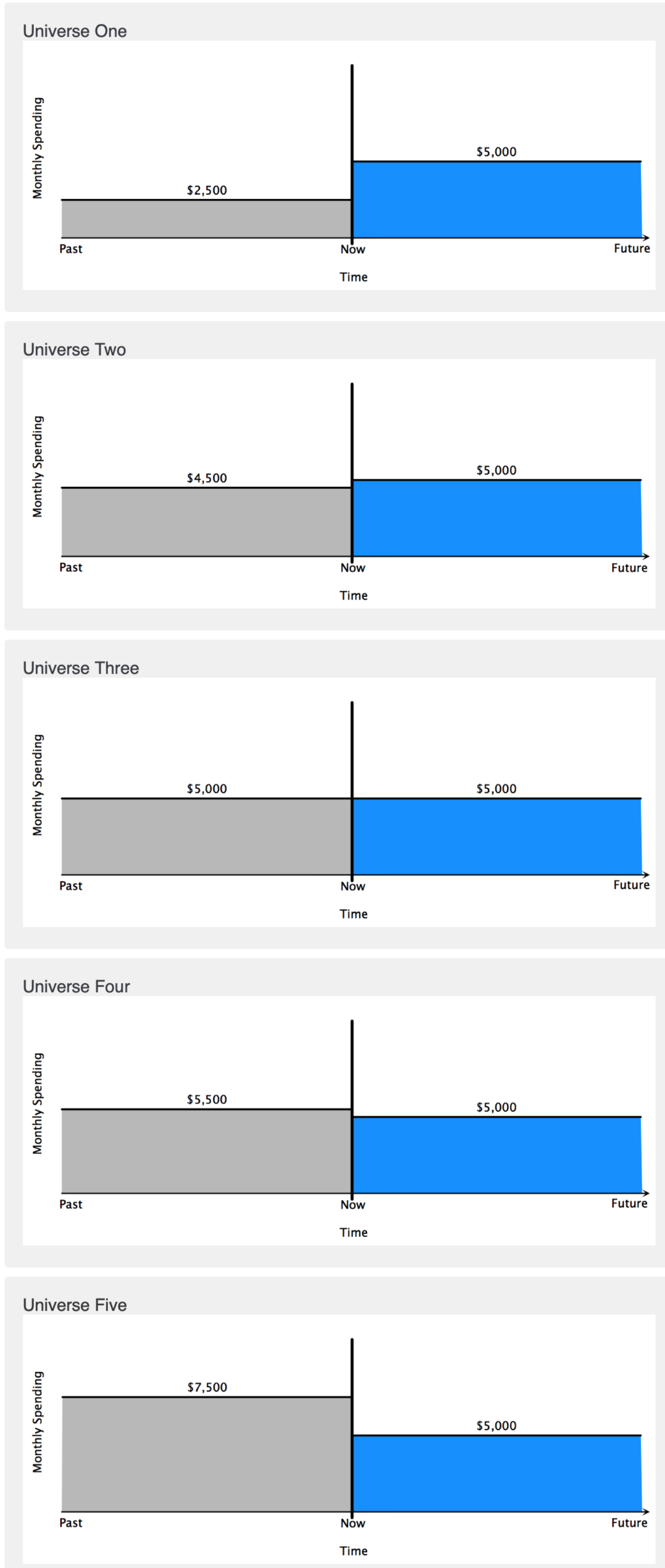


Figure 6: Monthly spending graphs: existence of habit formation.

specified in each universe and therefore requires no cost to determine it. If one models cognition costs as the cost associated with changing consumption path,[23] the respondents should choose Universe Three over the others because it involves the least change of spending and, therefore, the least level of cognition costs. Whichever way one models cognition costs, the responses would be dramatically different from mine.

As a clarification of the above discussion, the fact that my evidence supports the existence of habit formation being separate from adjustment and cognition costs does not reject the existence of adjustment and cognition costs. What is shown is that habit formation exists separately from adjustment and cognition costs and, consequently, that the latter cannot be deeper causes of the former.

4.2 Habit Depreciation Speed

To find out how fast habit depreciates, I elicit and estimate the parameter θ as in $\dot{H} = \theta\,(C - H)$. In the survey question (Figure 3), I varied the persistence and level of past spending so that the resulting difference in current levels of habit enabled the elicitation of θ (Proposition 1).

Proposition 1. *Under exact elicitation, $\theta > -\ln\left(1 - \frac{\Delta C_{U1}}{\Delta C_{U2}}\right)$ if the respondent chooses Universe One over Universe Two for a better future experience in the habit depreciation rate question.*

All proofs are in section F of the appendix. See section F.2 for the proof of this proposition.

In Proposition 1, ΔC_{U1} (ΔC_{U2}) is the difference between the monthly spending in Universe One (Universe Two) and the baseline monthly spending, \$5,000 per month. In the example of Figure 3, $\Delta C_{U1} = \$2,000$ and $\Delta C_{U2} = \$4,000$. Thus, according to Proposition 1, this survey question separates the values of θ into two complementary intervals: $\theta > \ln 2$ and $\theta < \ln 2$.[24] The intuition of the proposition is that choosing the spending path with more persistent past spending for a better future experience means a faster habit depreciation speed.

I use unfolding brackets to further pin down a finer range of θ for each possible response. Each respondent answered one to two follow-up questions that put her answer into one of the six brackets in Figure 7.

[23] E.g., one may need to spend some cognition efforts to adjust her spending to a level different from her previous level.

[24] I abstract from $\theta = \ln 2$ because θ has a probability of 0 to be exactly equal to $\ln 2$. The threshold of $\ln 2$ in continuous time corresponds to a threshold of 0.5 in discrete time at the annual frequency.

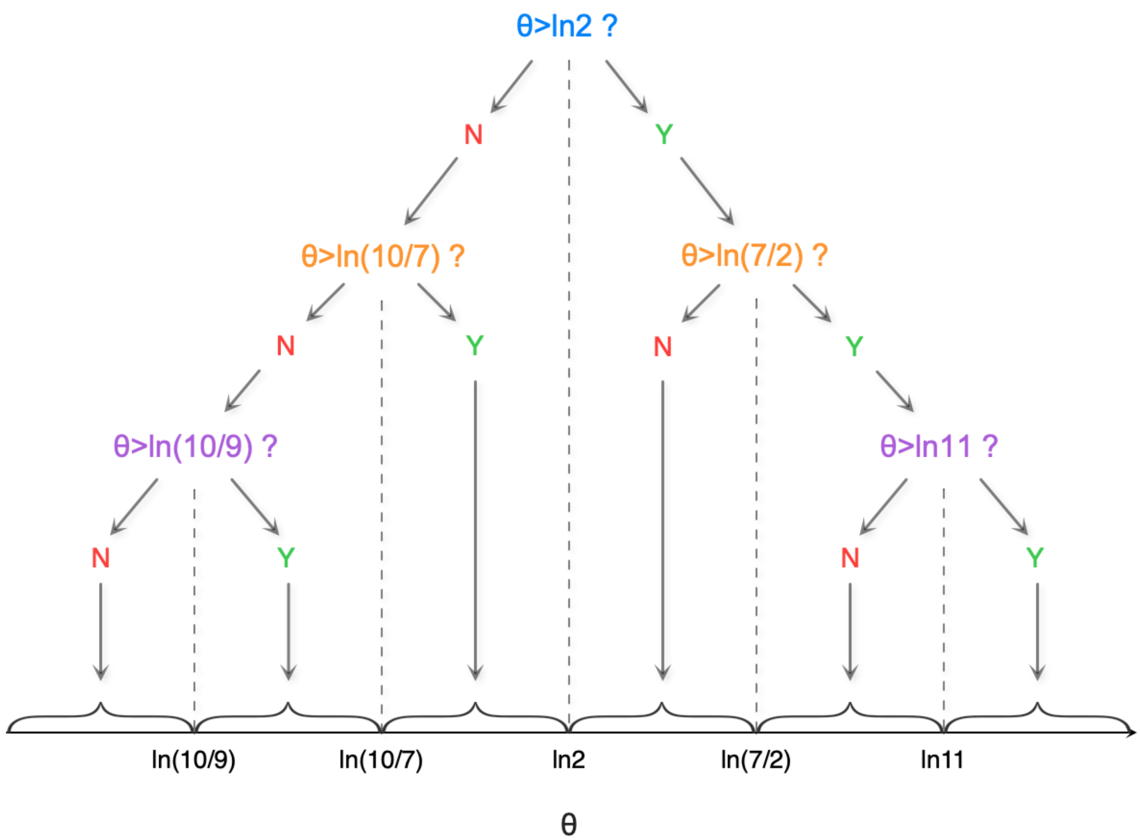


Figure 7: Unfolding brackets.

Table 5: Estimates of Habit Depreciation Speed

	MAP	Mean	Median	95% HPDI
Habit Depreciation Rate	1.11	1.11	1.11	[0.89, 1.31]
Habit Depreciation Factor	0.67	0.67	0.67	[0.60, 0.74]

Note: The habit depreciation factor is annual.

The estimate for habit depreciation rate is 1.11, which implies that habit depreciates by 67% annually (Table 5). The annual depreciation rate of 67% implies that habit will have changed about 90% after two years, which is remarkably close to the findings in the psychology literature that income adaptation takes about two years.[25]

The speed at which habit depreciates is important. In a typical additive habit formation model, the faster habit depreciates, the less risk averse agents of the model become because habit adjusts faster with consumption. The most cited paper on habit formation, Campbell and Cochrane (1999),[26] also employed the additive habit.[27] In their model, however, the faster habit depreciates, the more risk averse people are. The reason is that in their model, the implied steady-state habit intensity[28] is not constant but increases with the habit depreciation rate. The higher the habit intensity, the likelier it is that a fluctuation of consumption causes consumption to fall below the habit-intensity-adjusted level of habit, and thus the more risk averse the agent becomes. The net effect of a higher habit depreciation rate in their model is the sum of these two effects, which ultimately makes the agent in the model more risk averse.

Plugging my estimate of the habit depreciation speed into Campbell and Cochrane (1999) causes the agents to become so risk averse that they require a higher equity premium than has not been observed historically (Table 6). The time discount factor also must be unrealistically low, 0.35 per year, to match the mean historical risk-free rate (column 3). When a more realistic annual time discount factor of 0.89 is used, the level Campbell and Cochrane (1999) chose, people become even more risk-averse. They require a still higher expected return and are willing to accept a hugely negative interest rate, -92.19% per year, to be able to save (column 4). The intuition is that when the higher time discount factor makes people care more about the future, future risk matters more to them, and, as a result, they become yet more risk-averse. The

[25]See Clark, Frijters and Shields (2008) for a review.

[26]Google Scholar reports that this paper has been cited 4,960 times as of November 14, 2019.

[27]They specified a nonlinear evolution for habit or surplus consumption ratio, to be precise. It coincides with the linear habit evolution specified here in steady state.

[28]Under additive habit, the instantaneous utility is $u(C - \alpha H)$ where α is the habit intensity parameter. In Campbell and Cochrane (1999)'s notation $X_t = \alpha_t H_t$. The implied steady-state habit intensity is, therefore, X_t/C_t.

Table 6: Effect of Habit Depreciation Speed in Campbell and Cochrane (1999): Equity Premium

	Postwar Sample	Model Sample				
	(1)	(2)	(3)	(4)	(5)	(6)
Habit Depreciation Factor	-	0.11	**0.67**	**0.67**	**0.60**	**0.30**
Time Discount Factor	-	0.89	0.35	**0.89**	0.42	0.71
Expected Excess log Return	6.69%	6.71%	44.08%	101.56%	37.44%	16.51%
Std of Excess log Return	15.20%	15.64%	31.08%	96.70%	29.62%	22.01%
Sharpe Ratio	0.43	0.43	1.42	1.05	1.26	0.75
Mean of Risk Free Rate	0.94%	0.94%	0.94%	-92.19%	0.94%	0.94%

Notes: All annualized values. Bold numbers are my changes to Campbell and Cochrane (1999)'s calibration. Column 1 is based on postwar (1947–95) value-weighted New York Stock Exchange stock index returns and 3-month Treasury bill rate; the column 2 is based on model sample under Campbell and Cochrane (1999)'s calibration (0.11 is the annual habit depreciation factor implied by Campbell and Cochrane (1999)'s calibration of the persistence coefficient, ϕ, of the surplus consumption ratio in their model); the column 3 is based on model sample under my estimate of habit depreciation factor; the column 4 is based on model sample under my estimate of habit depreciation factor and the time discount factor of 0.89; the column 5 is based on model sample under the lower bound of the 95% HPDI of my estimate of habit depreciation factor; and the column 6 is based on model sample under a habit depreciation factor far smaller than the lower bound of the 99% HPDI of my estimate of it. The power coefficient of the constant relative risk aversion utility function is 2, as in Campbell and Cochrane (1999).

higher risk aversion drives up the motive of precautionary saving. This motive is so strong that people are willing to pay more than 90% of the principal to be able to transfer the remaining less than 10% of it to the next year. When one lowers the time discount factor or the habit depreciation factor, the model moments are closer to reality, but the percentage differences are still at least 40% (columns 5 and 6), even when habit depreciates by only 30% each year, which is far from the 99% HPDI of the habit depreciation factor.

Table 7 shows the effect of my estimated depreciation factor on the predictability of dividend yield on expected return. With my estimated habit depreciation factor, the regression coefficients and R^2s are larger than their postwar counterparts. Interestingly, R^2s start to decrease after the second year. Figure 8 replicates Figure 9 of Campbell and Cochrane (1999), with an additional curve showing model-predicted price-dividend ratio with a habit depreciation factor of 0.67. Even though the fluctuations retain about the same pattern, a higher habit depreciation speed makes the price-dividend ratio much smaller. This observation is consistent with the fact that

Table 7: Effect of Habit Depreciation Speed in Campbell and Cochrane (1999): Long-run Regression

			$\theta = 0.11$				$\theta = 0.67$			
	Postwar Sample		Consumption Claim		Dividend Claim		Consumption Claim		Dividend Claim	
Horizon (Years)	10 × Coefficient	R^2	10 × Coefficient	R^2	10 × Coefficient	R^2	10 × Coefficient	R^2	10 × Coefficient	R^2
1	-2.6	.18	-2.0	.13	-1.9	.08	-9.6	.60	-9.7	.53
2	-4.3	.27	-3.7	.23	-3.6	.14	-12.9	.71	-13.0	.60
3	-5.4	.37	-5.1	.32	-5.0	.19	-14.0	.67	-14.0	.56
5	-9.0	.55	-7.5	.46	-7.3	.26	-14.7	.55	-14.7	.44
7	-12.1	.65	-9.4	.55	-9.2	.30	-14.4	.44	-14.3	.34

Notes: Long-horizon regressions of log excess stock returns on the log price/dividend ratio in historical and simulated data. θ is the annual habit depreciation factor. $\theta = 0.11$ is Campbell and Cochrane (1999)'s calibration, and $\theta = 0.67$ is my estimate. This table replicates part of Table 5 of Campbell and Cochrane (1999).

a higher value of θ induces agents to be more risk-averse and, therefore, to require a higher return to take on the same level of consumption risk.

My respondents might not be representative of marginal investors who price the assets. It is, however, unnecessary for my respondents to represent the marginal investors in every way possible. The above discussion stands as long as the typical habit depreciation speed of my respondents is the same or close to that of the typical marginal investor, which would be the case if this parameter is a deep preference parameter that does not vary significantly across demographics. In section 6, I present such evidence: Habit depreciation rate does not vary with age, gender, household size, and household income. One potential explanation could be that the speed at which people's habit adjusts is determined genetically.

One of my goals for this study is to uncover the true capability of habit formation models. The literature has done a great job in exploring the models' potential. What is missing, however, is whether habit formation is *indeed* that powerful. I have found one inconsistency with reality in a popular habit formation model when I plugged in my micro-based estimate of habit depreciation factor. One, however, must take extra caution in interpreting this inconsistency. Given that my evidence supports the existence of habit formation, it is more likely that the way habit formation is modeled needs improvement (this will be elaborated in the following sections) than that modeling habit formation is the wrong route. Just as we need features beyond diminishing marginal utility—no matter how realistic and fundamental it is—to be able to explain reality better, we might also need features *in addition to* habit formation to

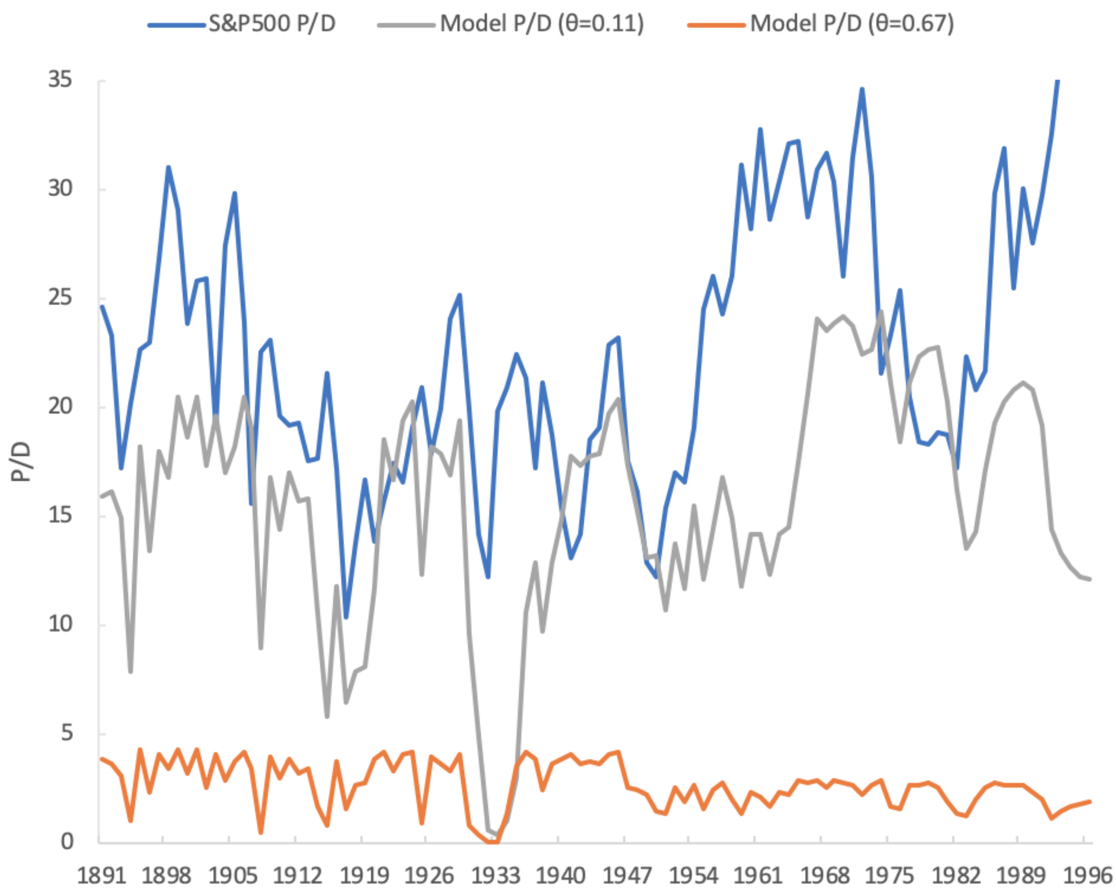


Figure 8: Historical price-dividend ratio and model predictions based on consumption history.

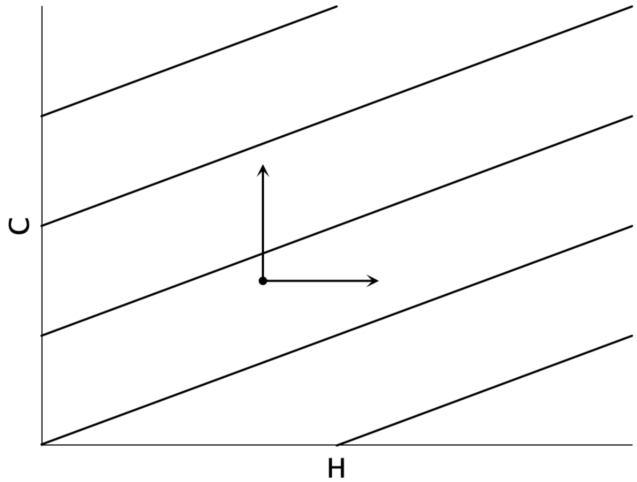


Figure 9: Additive habit indifference map.

fully explain asset pricing and other phenomena.

4.3 Additive and Multiplicative Habits

Additive and multiplicative habits constitute most, if not all, of habit specifications that have been taken to data. Additive habit is adopted relatively more often in the literature for its time-varying risk aversion, whereas multiplicative habit, at least in its simplest form, is not able to generate such pattern of risk aversion. Does survey evidence support this theoretical choice? I answer this question here, starting with four tests that distinguish the two formulations.

Proposition 2. *Additive habit, $u(C,H) \equiv v(C-\alpha H)$ with $\alpha \in \mathbb{R}^+$, implies $\frac{u_{CH}}{u_{HH}}\frac{u_H}{u_C} = 1$ and $\frac{u_{CH}}{u_{CC}}\frac{u_C}{u_H} = 1$.*[29]

The intuition for this set of tests is that under additive habit, the indifference curves are parallel straight lines so that moving in any direction in the indifference map will not change the slope of the indifference curves. The two tests are the two bases spanning all such movements: increase H alone and increase C alone (Figure 9).

Proposition 3. *Multiplicative habit, $u(C,H) \equiv v(C/H^\alpha)$ with $\alpha \in \mathbb{R}^+$, implies $\frac{Hu_H u_{CH}}{u_C u_H + Hu_C u_{HH}} = 1$ and $\frac{Cu_C u_{CH}}{u_C u_H + Cu_H u_{CC}} = 1$.*[30]

[29]See section F.3 of the appendix for proof.

[30]One can derive a class of tests based on the homotheticity of multiplicative habit. The two tests here imply this class of tests. See section F.4 of the appendix for proof.

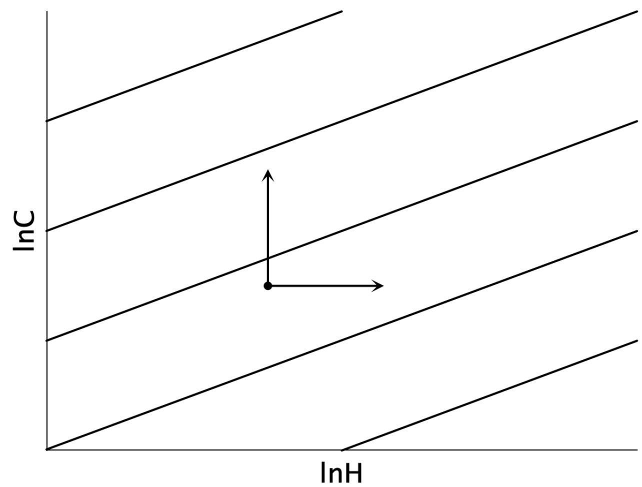


Figure 10: Multiplicative habit indifference map in $(\ln C, \ln H)$ space.

In the space of $(\ln C, \ln H)$, the two tests of multiplicative habit have similar intuition as those for the additive habit (Figure 10).

Because the left-hand sides of the four equations are functions of $-\frac{u_H}{u_C}$, $\frac{Hu_{HH}}{u_H}$, $\frac{u_{CH}}{u_{HH}}$, and $\frac{u_{CC}}{u_{HH}}$, implementing the tests requires their elicitation. Because of the generality of the utility function, elicitation of these preference parameters will be up to third order.

$-\frac{u_H}{u_C}$ is the slope (or inverse of the slope) of the indifference curve. To elicit it, I changed both future and past spending in the same direction to move along an indifference curve. The resulting survey question has monthly spending graphs as in Figure 11.

Proposition 4. *Under second-order elicitation, $-\frac{u_H}{u_C} < \frac{(\rho+\theta)\Delta f}{\rho\Delta e+\theta\Delta f}$ if the respondent chooses Universe One over Universe Two for a better future experience in the slope of the indifference curve.*[31]

My estimate for the slope of the indifference curve is 0.62 (Table 8). The implied positive sign of u_C is consistent with my assumption of the positive monotonicity of consumption. The magnitude of this estimate implies that about 62% of utility gain from a higher level of consumption gradually disappears because of habit formation. This number is remarkably close to the finding that about 60% of the effect of income on happiness is lost across time (Van Praag, 1971; Van Praag and Frijters, 1999).

[31] $\Delta e = \$2,000$ and $\Delta f = \$200$ in the example of Figure 11. Knowledge about the time discount rate, ρ, is required to estimate the slope of indifference curve and other preference parameters, the elicitation of which is relegated to section C of the appendix because of this indirect interest in it. See section F.5 of the appendix for proof.

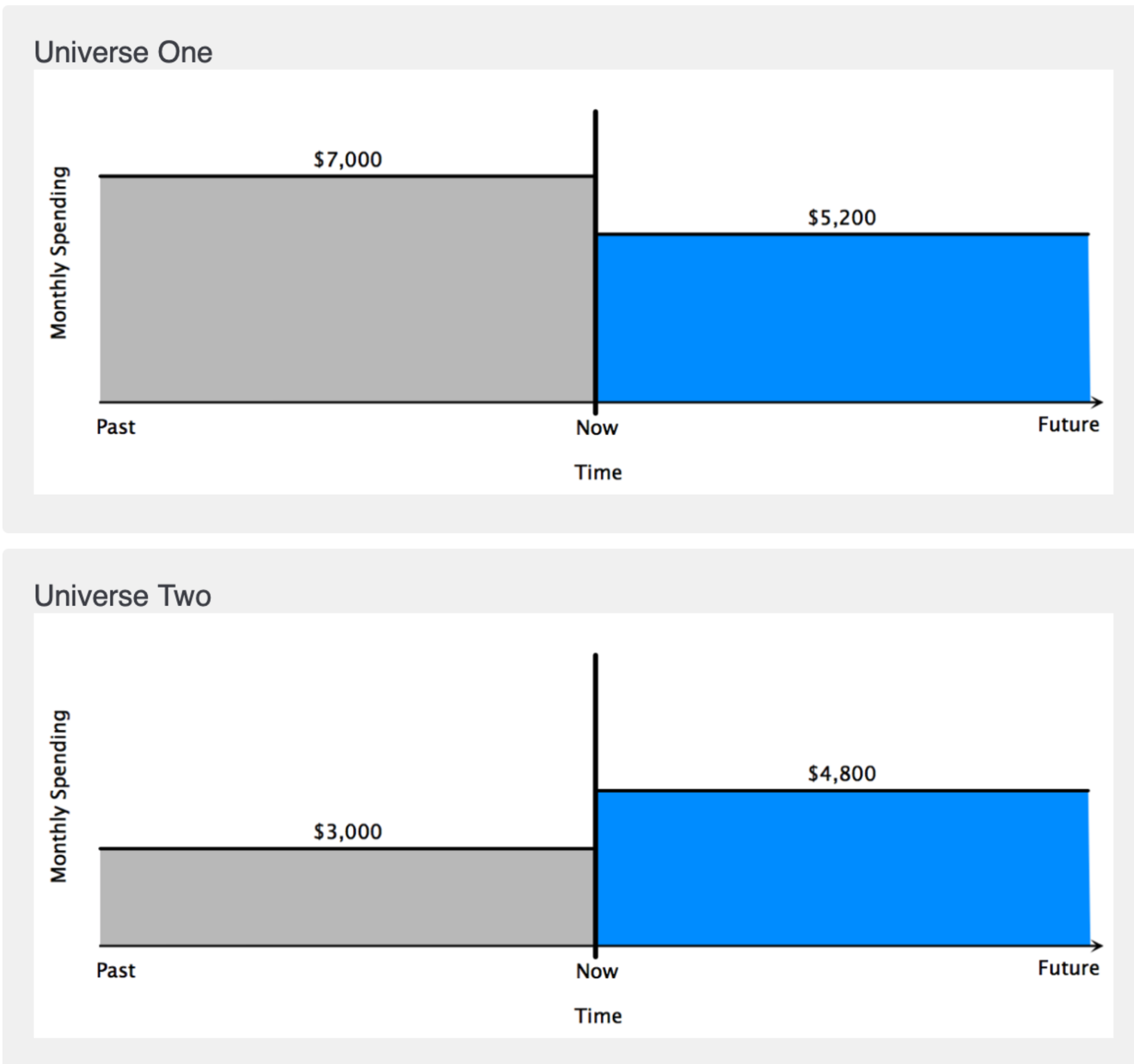


Figure 11: Monthly spending graphs—slope of indifference curve.

Table 8: Estimates of Ratios of Utility Derivatives

	MAP	Mean	Median	95% HPDI
$-u_H/u_C$	0.62	0.62	0.62	[0.51, 0.73]
Hu_{HH}/u_H	7.71	7.78	7.76	[6.84, 8.78]
u_{CH}/u_{HH}	-0.86	-0.85	-0.86	[-1.00, -0.70]
u_{CC}/u_{HH}	3.72	3.72	3.72	[3.01, 4.41]

Because the habit intensity parameter, α, equals the slope of indifference curve in both additive habit (everywhere) and multiplicative habit (in steady state), my estimate of 0.62 is essentially the same as the macro estimates of about 0.6 in the literature (Havranek, Rusnák and Sokolova, 2017). Havranek, Rusnák and Sokolova (2017) also found that the estimates using microdata tend to be lower, 0.1, than those estimated using macrodata, 0.6. Interestingly, my estimate using microdata is indistinguishable from those using macrodata.

To elicit $\frac{Hu_{HH}}{u_H}$, I presented the respondents with a trade-off between past spending and fluctuation of past spending (Figure 12).

Proposition 5. *Under second-order elicitation,* $\frac{Hu_{HH}}{u_H} < \frac{2(\rho+2\theta)}{\rho+\theta}\frac{\Delta f/\Delta e-1}{(\Delta f/\Delta e)^2+1}\frac{H}{\Delta e}$ *if the respondent chooses Universe One over Universe Two for a better future experience in the* $\frac{Hu_{HH}}{u_H}$ *question.*[32]

I estimate $\frac{Hu_{HH}}{u_H}$ to be 7.71 (Table 8). Because I already estimated that $u_H < 0$, this implies that $u_{HH} < 0$.

The elicitation of $\frac{u_{CH}}{u_{HH}}$ rests on inducing fluctuations in both future and past spending at the same time. The monthly spending graphs for the survey question are in Figure 13.

Proposition 6. *Under third-order elicitation,* $\frac{u_{CH}}{u_{HH}} < -\frac{(\rho+\theta)\Delta e+2\theta\Delta f}{2(\rho+2\theta)\Delta f}$ *if the respondent chooses Universe One over Universe Two for a better future experience in the* $\frac{u_{CH}}{u_{HH}}$ *question.*[33]

My estimate of $\frac{u_{CH}}{u_{HH}}$ is -0.86 (Table 8). Given that I estimated that $u_{HH} < 0$, this tells me that $u_{CH} > 0$, which is consistent with the idea of habit formation: The higher habit is, ceteris paribus, the more valuable an additional unit of consumption is.

[32] Δe = \$500 and Δf = \$650 in the example of Figure 12. See section F.6 of the appendix for proof.

[33] Δe = \$1000 and Δf = \$200 in the example of Figure 13. See section F.7 of the appendix for proof.

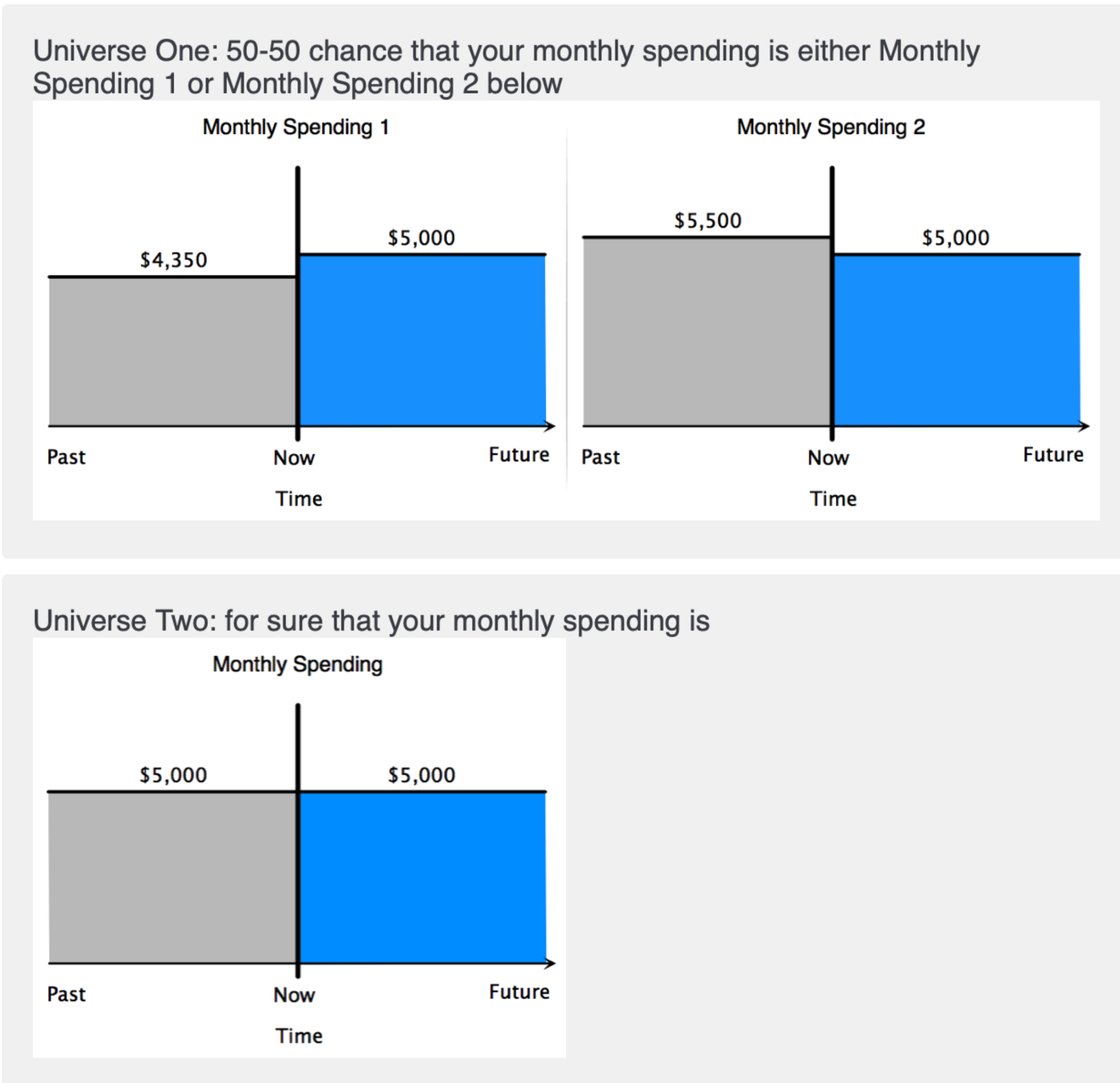


Figure 12: Monthly spending graphs—Hu_{HH}/u_H.

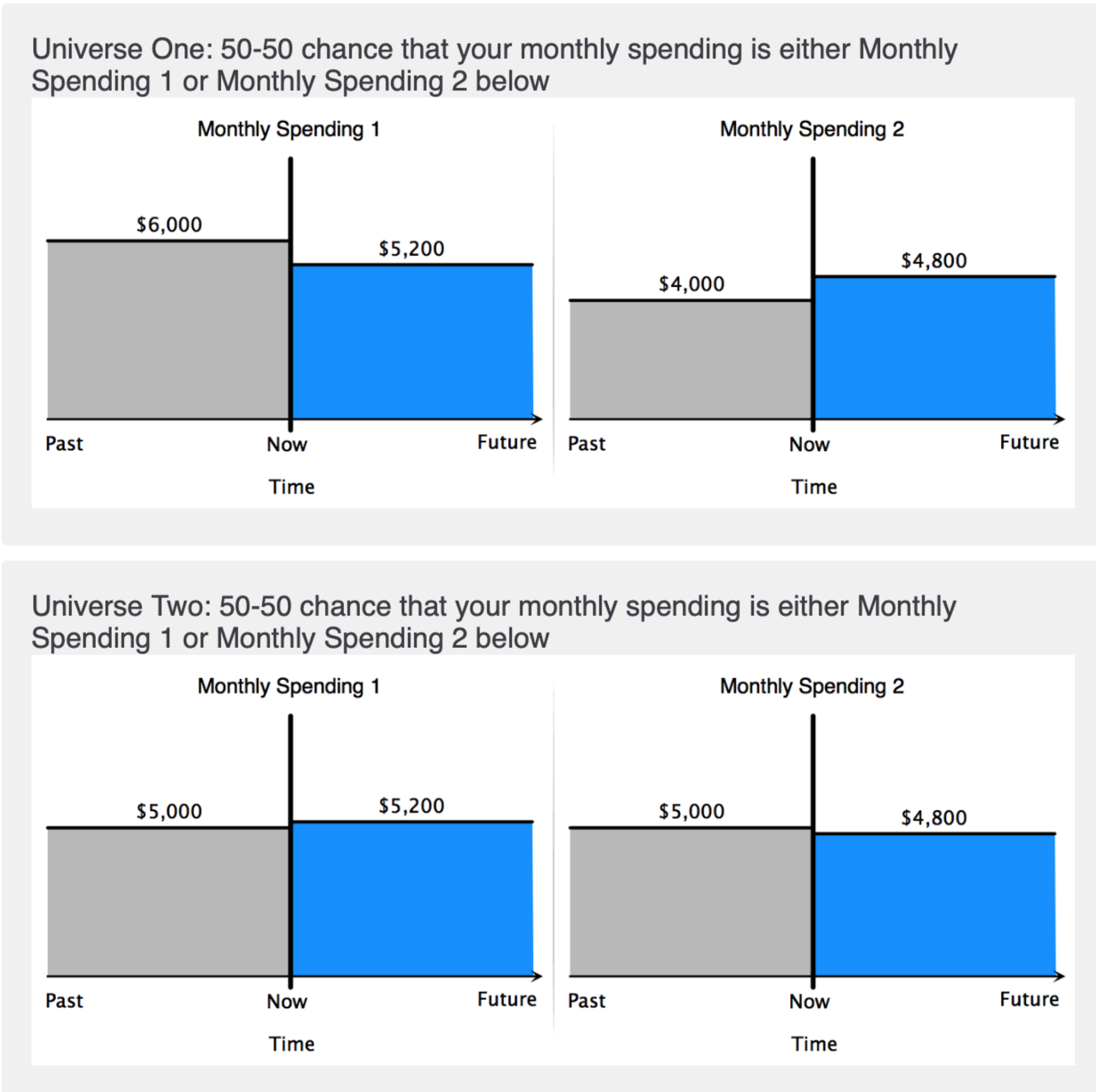


Figure 13: Monthly spending graphs—u_{CH}/u_{HH}.

Table 9: Estimates for Testing Additive and Multiplicative Habits

	MAP	Mean	Median	99% HPDI
$\frac{u_{CH}u_H}{u_{HH}u_C}$	0.53	0.53	0.53	[0.36, 0.71]
$\frac{u_{CH}u_C}{u_{CC}u_H}$	0.37	0.38	0.37	[0.23, 0.56]
$\frac{Hu_H u_{CH}}{u_C(u_H+Hu_{HH})}$	0.46	0.47	0.47	[0.32, 0.63]
$\frac{Cu_C u_{CH}}{u_H(u_C+Cu_{CC})}$	-0.24	-0.25	-0.24	[-0.34, -0.16]

$\frac{u_{CC}}{u_{HH}}$ is about the trade-off between two sources of fluctuations, one from future spending and the other from past spending, which leads to the monthly spending graphs in Figure 14.

Proposition 7. *Under third-order elicitation,* $\frac{u_{CC}}{u_{HH}} < \frac{\rho}{\rho+2\theta}\left(\frac{\Delta e}{\Delta f}\right)^2 - \frac{2\theta}{\rho+\theta}\frac{u_{CH}}{u_{HH}} - \frac{2\theta^2}{(\rho+\theta)(\rho+2\theta)}$ *if the respondent chooses Universe One over Universe Two for a better future experience in the* $\frac{u_{CC}}{u_{HH}}$ *question.*[34]

$\frac{u_{CC}}{u_{HH}}$ is estimated to be 3.72 (Table 8). This is consistent with my assumption of $u_{CC} < 0$.

With the estimates of $-\frac{u_H}{u_C}$, $\frac{Hu_{HH}}{u_H}$, $\frac{u_{CH}}{u_{HH}}$, and $\frac{u_{CC}}{u_{HH}}$, I can calculate the statistics for testing additive and multiplicative habits. The point estimates show that none of the four statistics is equal to one (Table 9). Furthermore, one is not even in the 99% credible intervals of these statistics. This means that my survey evidence supports neither additive nor multiplicative habit.

Given the existence of habit formation, there are two possibilities as to why this happened. On the one hand, the habit utility function is indeed of neither the additive nor the multiplicative form. I have yet to see a third formulation of habit utility function that has been applied to data in the literature. If a micro-consistent form exists, it might be the key to solving the inconsistency I discussed in the previous section and to explaining phenomena that current habit formation models struggle to account for. On the other hand, habit formation itself might be insufficient. The assumption of a representative agent might need to be relaxed or other features might need to be introduced to the model or both to be able to match the reality fully. The first possibility might be the near-term direction for future research because my survey evidence is in line with the general idea of habit formation and does not exclude the possibility of habit preferences consistent with the evidence.

[34] Δe = \$2200 and Δf = \$500 in the example of Figure 14. See section F.8 of the appendix for proof.

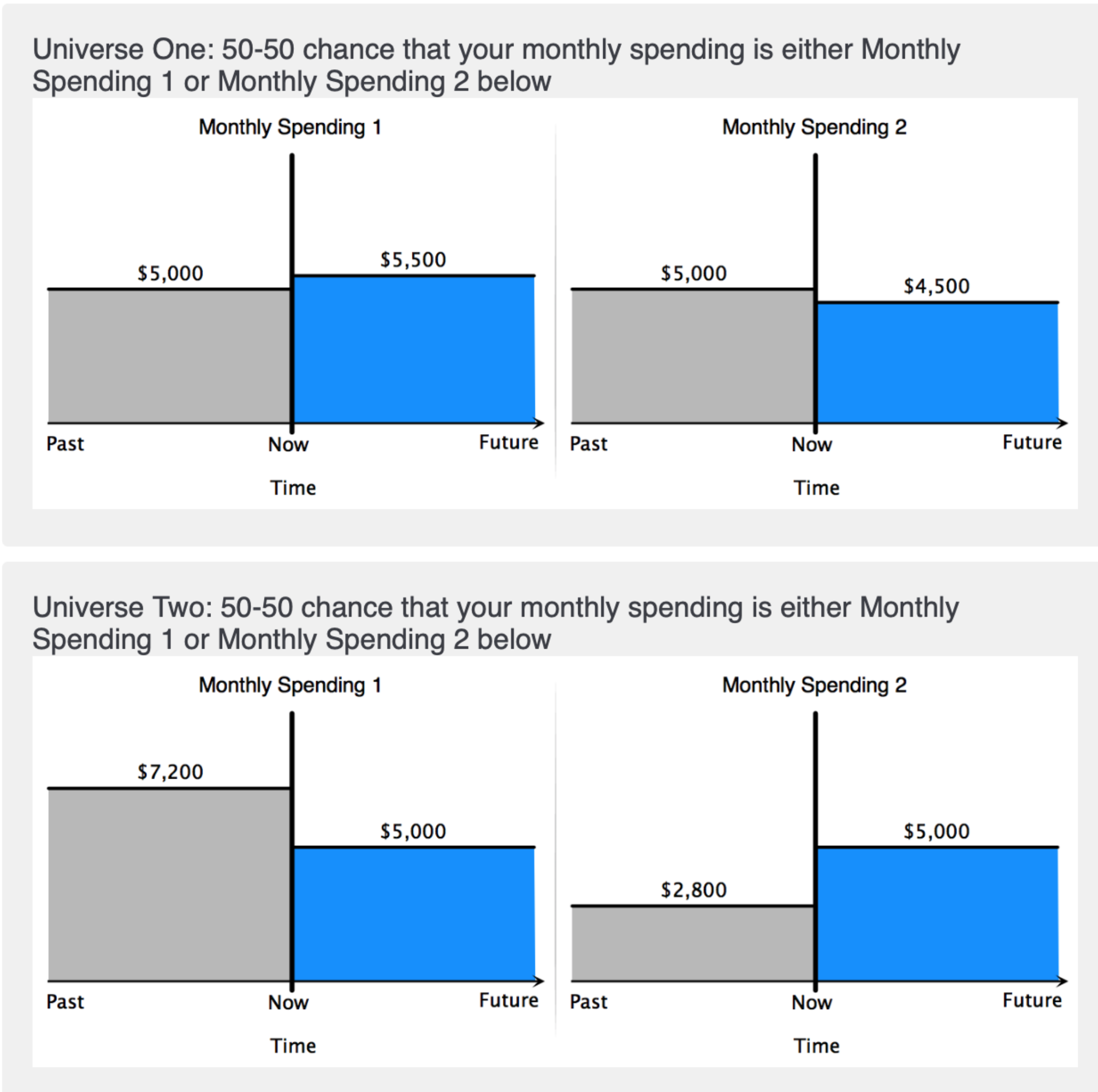


Figure 14: Monthly spending graphs—u_{CC}/u_{HH}.

4.4 External Habit

So far, I have been abstracting from the effect of other people's past spending on my current well-being, the external habit. This type of habit formation has prevailed in the literature, contrasting with both the existing empirical evidence and our day-to-day intuition that the formation of habits originates mostly from one's own behavior (internal habit). These two types of habits differ in their levels of tractability and can also have dramatically different implications for optimal tax policy and welfare analysis (Lettau and Uhlig, 2000; Ljungqvist and Uhlig, 2015). In this section, I explore the question of whether others' spending indeed affects one's habit.

If an increase in others' past spending, ceteris paribus, makes one worse off in the future, external habit exists. That is, external habit exists if

$$Q_{EH} \equiv u\left(C, H\left(\{C\}, \{C_{others} + \Delta C_{others}\}\right)\right) - u\left(C, H\left(\{C\}, \{C_{others}\}\right)\right) < 0$$

for $\Delta C_{others} > 0$, where $\{\cdot\}$ denotes historical path, and I explicitly allow the dependence of habit on other people's past spending, C_{others}.

To elicit the sign of Q_{EH} and the external habit mixture coefficient (see below), I use a set of questions where both others' and one's past spending vary. This leads to the monthly spending graphs in Figure 15 and the elicitation in Proposition 8.

Proposition 8. *Given the existence of internal habit, $Q_{EH} < 0$ if the respondent chooses Universe One over Universe Two for a better future experience in the external habit question.*[35]

The reverse of Proposition 8 does not hold: Choosing Universe Two over Universe One does not necessarily imply $Q_{EH} \geq 0$. As long as external habit is weak enough relative to internal habit, a respondent would always prefer Universe Two for a better future experience. Under this scenario, my survey does not contain information that allows me to separate the different possibilities for the sign of Q_{EH}. Therefore, I check every possibility. If always choosing Universe Two means $Q_{EH} < 0$ (most lenient), all the responses imply the existence of external habit. If always choosing Universe Two means $Q_{EH} > 0$ (strict), we are least likely to find evidence for external habit, whereas if always choosing Universe Two means $Q_{EH} = 0$ (lenient), we are relatively more likely to find external habit.

The response frequency in Table 10 indicates that the choice of the average respondent agrees with the existence of external habit. My estimate under the strict

[35]See section F.9 of the appendix for proof.

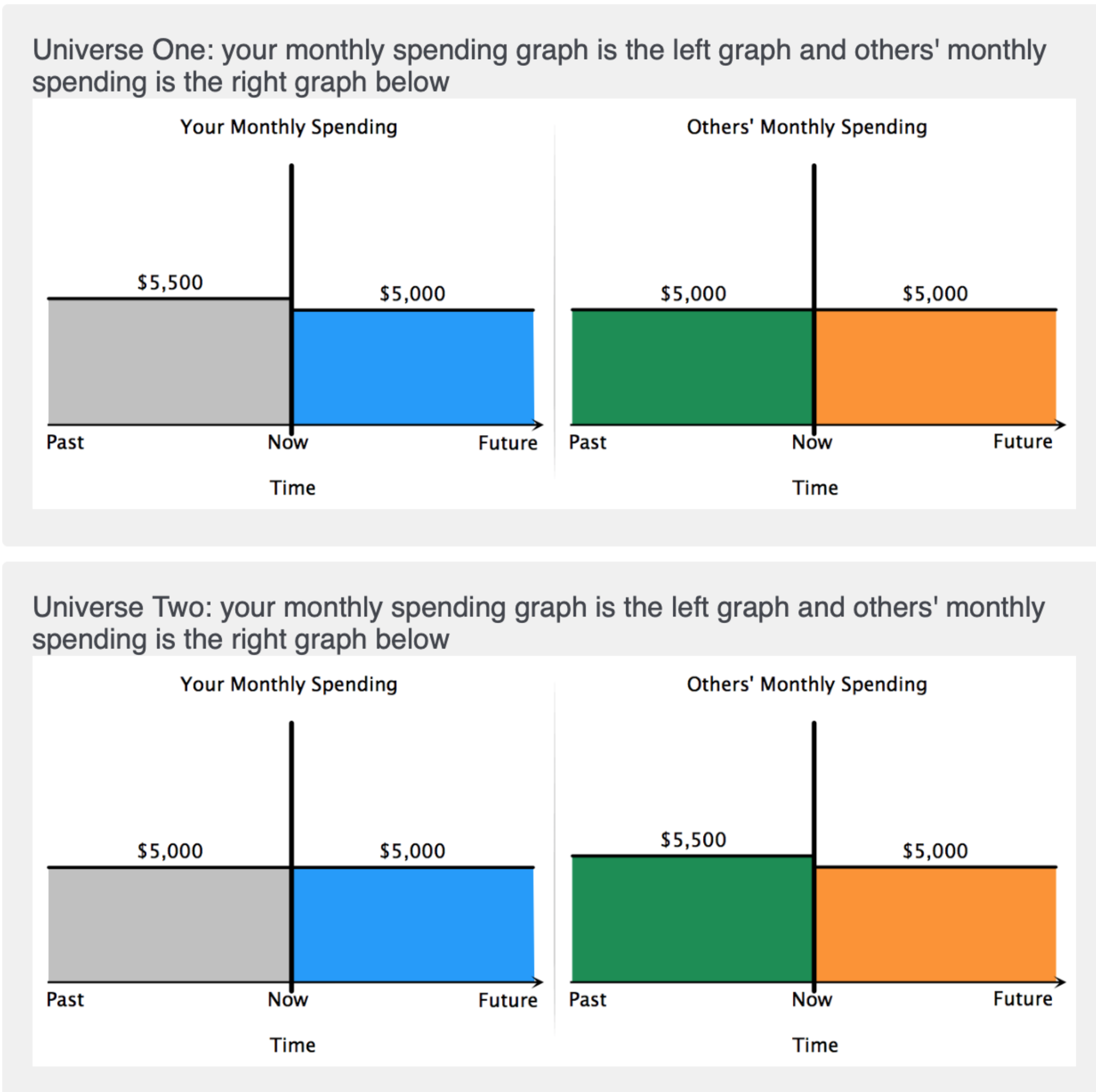


Figure 15: Monthly spending graphs—external habit.

Table 10: Response Frequency (Percentage) for External Habit Question

$\text{sgn}\left(Q_{EH}\right)$	-1	-1, 0, or 1
First Wave	54	46
Second Wave	54	46

Table 11: Estimate of Existence of External Habit

	MAP	Mean	Median	90% HPMI	95% HPMI	99% HPMI
$\text{sgn}\left(Q_{EH}\right)$ (strict)	-1.00	-0.84	-1.00	[-1.00, -1.00]	[-1.00, 1.00]	[-1.00, 1.00]
$\text{sgn}\left(Q_{EH}\right)$ (lenient)	-1.00	-1.00	-1.00	[-1.00, -1.00]	[-1.00, -1.00]	[-1.00, -1.00]

Note: HPMIs of additional levels are shown to help determine confidence level.

case shows that $Q_{EH} < 0$ in at least 90% HPMI, whereas the opposite sign could emerge in 95% or higher-level HPMI (Table 11). The lenient scenario, however, finds evidence of a negative sign for Q_{EH} at the 99% level. Overall, external habit exists at least at the 90% level.

Composition of Internal and External Habits

Given the existence of internal and external habits, what is their relative importance? To answer this question, I augment the habit formation model. I emphasize that the model's augmentation does not invalidate previous results because the augmented model nests the unaugmented model. Equivalently, one can think of my habit model as the most expanded version that includes everything that has been and will be discussed. I simply did not show some of the variables in previous sections because they either are constant or get canceled in elicitation, so not explicitly listing them results in no loss of generality.

I have already implicitly augmented the model in the first part of this section, where I allowed habit to depend on other people's spending. In this second part of the section, I specialize this dependence as per Grishchenko (2010):

$$\dot{H} = \theta\left(\left(1-\omega\right)C + \omega C_{others} - H\right) \tag{1}$$

where ω is the external habit mixture coefficient governing the contribution of others' spending on my habit. To elicit this parameter, I can use the same question that elicited the existence of external habit in the first part of this section.

Proposition 9. *Under exact elicitation,* $\omega > \frac{\Delta C}{\Delta C + \Delta C_{others}}$ *if the respondent chooses*

Table 12: Estimate of External Habit Mixture Coefficient

	MAP	Mean	Median	95% HPDI
ω	0.17	0.16	0.16	[0.07, 0.26]

Universe One over Universe Two for a better future experience in the external habit question.[36]

The point estimate indicates that others' spending contributes to about 17% of one's habit (Table 12). The 95% HPDI of the external habit mixture coefficient does not include 0, which is consistent with the existence of external habit, as shown above.

4.5 Relative Strength of Habit Formation and Keeping Up with the Joneses

To elicit the relative strength of habit formation and keeping up with the Joneses, I reveal others' spending in the utility function, $u(C, H, C_{others})$. I then elicit the ratio of $\frac{u_{C_{others}}}{u_H}$ by varying others' spending in both the past and the future. The survey question is illustrated by the monthly spending graphs in Figure 16.

Proposition 10. *Under first-order elicitation, $\frac{u_{C_{others}}}{u_H} < \frac{\omega}{\rho+\theta}\left(\rho\frac{\Delta C^{U2}_{others}}{\Delta C^{U1}_{others}} - \theta\right)$ if the respondent chooses Universe One over Universe Two for a better future experience in the $\frac{u_{C_{others}}}{u_H}$ question.*[37]

My point estimate for $\frac{u_{C_{others}}}{u_H}$ is about 1.05 (Table 13) and not significantly different from 1 at the 5% level. This supports that habit formation has about the same effect on utility as keeping up with the Joneses.

I draw two additional implications based on the significant negative sign of $u_{C_{others}}$ as implied from the estimate. First, keeping up with the Joneses exists separately from external habit. In the elicitation, external habit and keeping up with the Joneses are calculated separately. The fact that the estimate of $u_{C_{others}}$ is significantly negative means that keeping up with the Joneses exists after controlling for external habit. Second, keeping up with the Joneses is stronger than altruism. I did not restrict the sign of $u_{C_{others}}$ a priori, which can go both ways: altruism ($u_{C_{others}} > 0$) and keeping

[36] $\Delta C = \$500$ and $\Delta C_{others} = \$500$ in the example of Figure 15. See section F.10 of the appendix for proof.

[37] $\Delta C^{U1}_{others} = \$300$ and $\Delta C^{U2}_{others} = \$3,000$ in the example of Figure 16. See section F.11 of the appendix for proof.

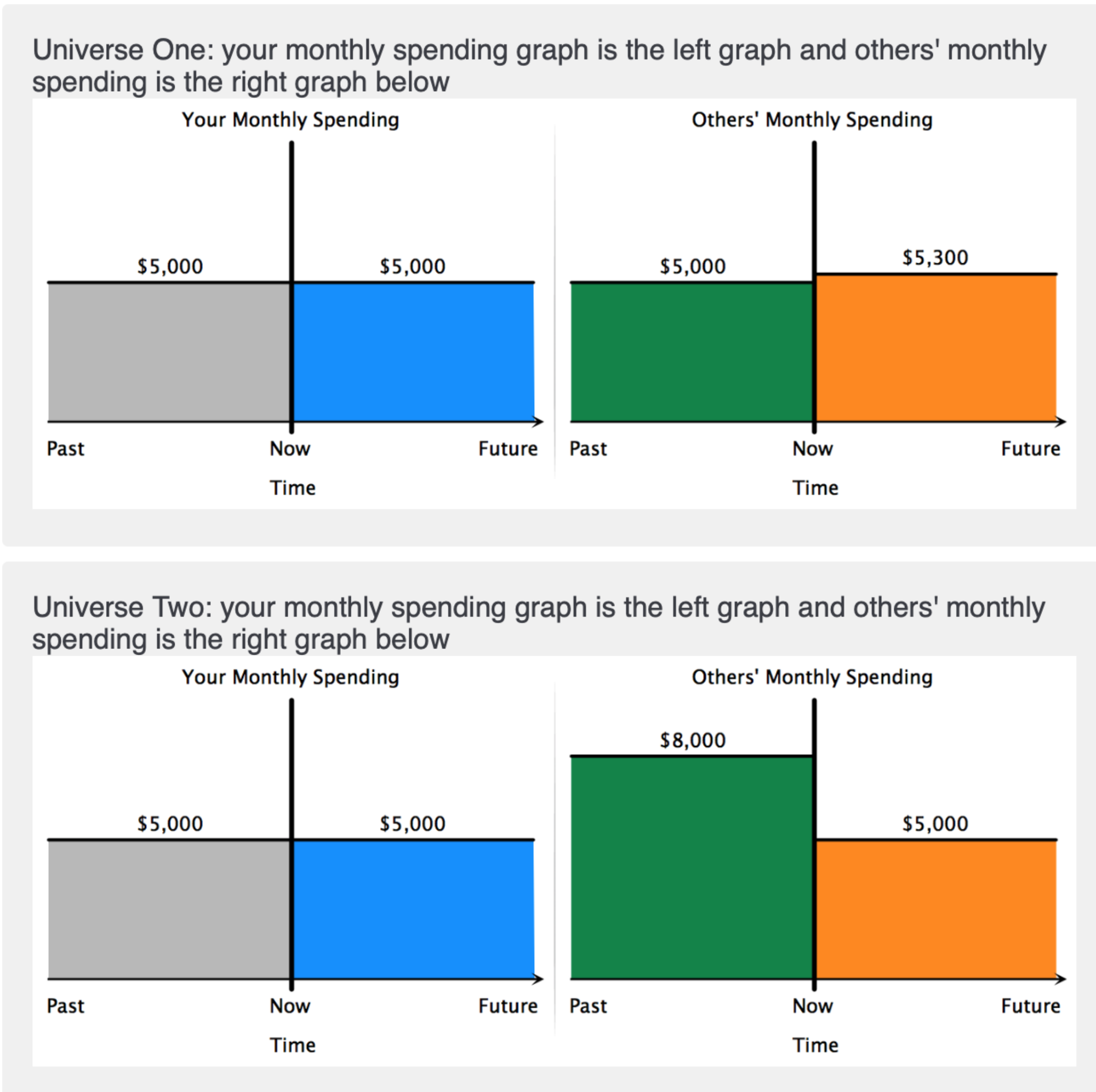


Figure 16: Monthly spending graphs—$u_{C_{others}}/u_H$.

Table 13: Estimates for $u_{C_{others}}/u_H$

	MAP	Mean	Median	95% HPDI
$\frac{u_{C_{others}}}{u_H}$	1.05	1.06	1.06	[0.70, 1.41]

up with the Joneses ($u_{C_{others}} < 0$). Essentially $u_{C_{others}}$ represents the net effect of these two phenomena. The significant negative sign of $u_{C_{others}}$, therefore, indicates that keeping up with the Joneses dominates altruism.

5 Explaining the Easterlin Paradox

The income–happiness paradox proposed by Easterlin states that income and happiness tend to be positively correlated in the short run and the cross section but uncorrelated in the long run (Easterlin, 1973, 1974, 1995, 2017; Kaiser and Vendrik, 2018). Alternative views have been proposed: Among others, that the U.S. data tend to be an outlier (Stevenson and Wolfers, 2008; Sacks, Stevenson and Wolfers, 2012) and that life satisfaction can be time intensive (Kimball and Willis, 2006). Despite the debate, the literature seems to be in broad agreement that the empirical gradient of happiness with respect to income is small and that the cross-section and short-run gradients tend to be larger than the long-run gradient. This section focuses on explaining the temporal heterogeneity of such gradients. To see the intuition most clearly, I will take the modal view of the paradox. Accommodation of the alternative views is readily achievable with slight changes in parameter values while keeping the intuition unchanged.

Explanations put forth for the paradox are generally of some reference-dependent form, among which habit formation and keeping up with the Joneses are the most popular (Easterlin, 1973, 1974, 1995; Clark, Frijters and Shields, 2008). Recent evidence on keeping up with the Joneses suggests that it is not powerful enough to fully explain the phenomenon (Luttmer, 2005; Lewbel et al., 2018; De Giorgi, Frederiksen and Pistaferri, 2019). Using my estimates on keeping up with the Joneses and habit formation of both the internal and external types, I show in this section that while each alone cannot generate the patterns of income and happiness of the Easterlin paradox, together they can.

Four clarifications merit discussion before proceeding. The first is that I focus on the causal channel that income changes happiness. Typical life experiences and studies exploiting exogenous variations support this view (Frijters, Haisken-DeNew and Shields, 2004; Gardner and Oswald, 2007). Evidence aside, this causality motivated the discovery of the paradox[38] and is the most counterintuitive, interesting,[39] and policy relevant. Non-income happiness-altering factors do not help explain the paradox because they generally improve with income, making the long-run income–happiness relationship even more mysterious (Di Tella and MacCulloch, 2008). The second clarification is that following the literature (Clark, Frijters and Shields, 2008; Benjamin et al., 2012; Perez-Truglia, 2019), I assume that the potential distinction between happiness and utility is of minimal effect on my discussion below. Third, the paradox holds when I replace income with consumption[40] because consumption is closely tied to income (Figure 17), while happiness still has a long-term trend of about zero (Figure 18). The relative lack of attention to the relationship between consumption and happiness is at least partly due to a lack of reliable individual-level panel data on consumption. Compared with income, consumption relates more directly to human welfare, as is widely accepted in the economics literature. Fourth, the literature provides at least three measures of happiness: affect measuring feelings of recent days, life satisfaction evaluation of life as a whole, and eudaemonia personal growth and meaning. I focus on the first two because their measurements are the most reliable (OECD, 2013), studied, and relevant to the paradox. I use the instantaneous utility as a proxy for affect[41] and lifetime utility for life satisfaction.

To assess how habit formation and keeping up with the Joneses potentially resolve the paradox, I will focus on a simplified environment that accentuates the mechanisms reconciling the paradox's tensions. I specify that people's utility is affected by both internal and external habits as well as keeping up with the Joneses. Habit evolves according to equation (1) with the habit depreciation rate and the external habit mixture coefficient calibrated to my estimates, 1.11 and 0.17, respectively. Keeping up with the Joneses and external habit emerge after the agent realizes that other people's consumption has changed, which is assumed to be k years after a universal consumption change occurs.[42] With no loss of generality, I assume that the full effect

[38] In addition to an interview where Easterlin discussed his motivation, one can get an idea of the question that interested Easterlin from the titles of his seminal papers: "Does Money Buy Happiness?"(Easterlin, 1973) and "Does Economic Growth Improve the Human Lot? Some Empirical Evidence"(Easterlin, 1974).

[39] This is evidenced by that the vast majority of speculative explanations of the paradox have focused on this channel.

[40] Throughout, consumption refers to real consumption.

[41] One can alternatively use the integral of instantaneous utility over the past one day or week to proxy affect, which are the typical time frames for survey questions on affect. I tried both and other time frames as well and found that the difference is trivial. This is a necessary implication of the fact that the curvature of the instantaneous utility is rather smooth (see below).

[42] The exact time of the delay does not matter for the resolution of the paradox. It only affects the speed at which people's utility converges to its steady state.

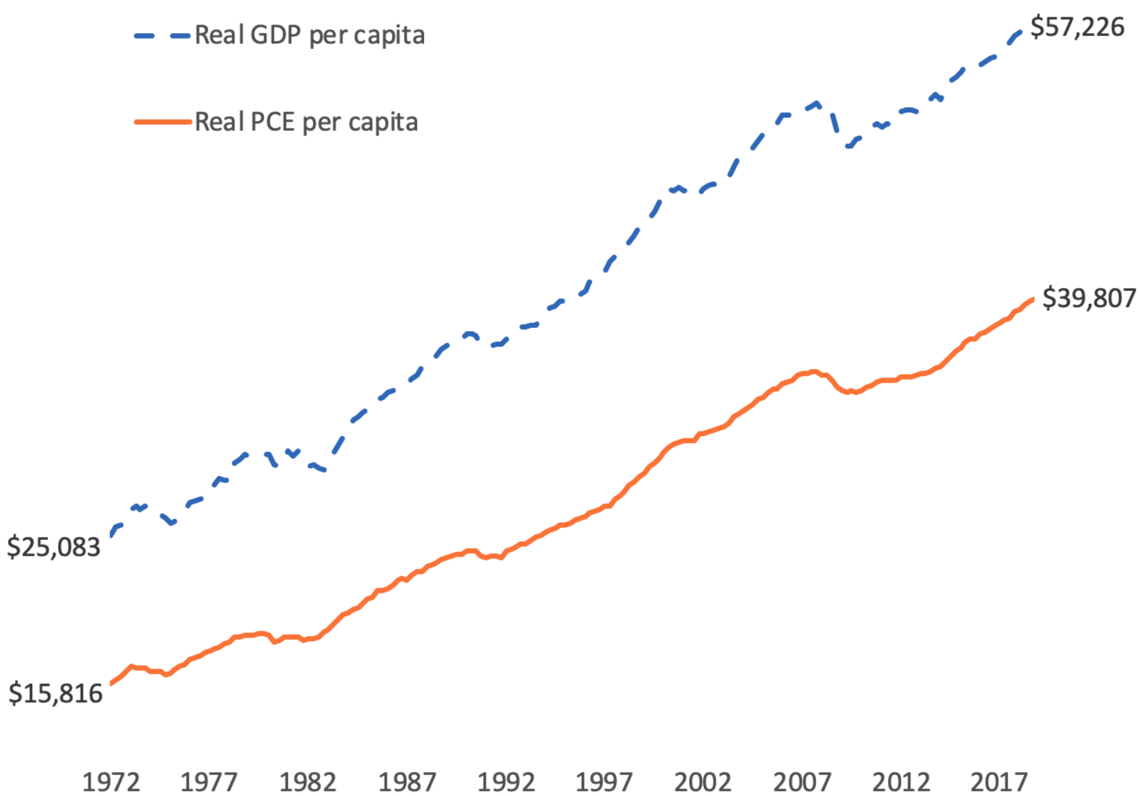


Figure 17: Real GDP and PCE per capita in the United States, 1972–2018. Chained 2012 dollars. Data from the U.S. Bureau of Economic Analysis.

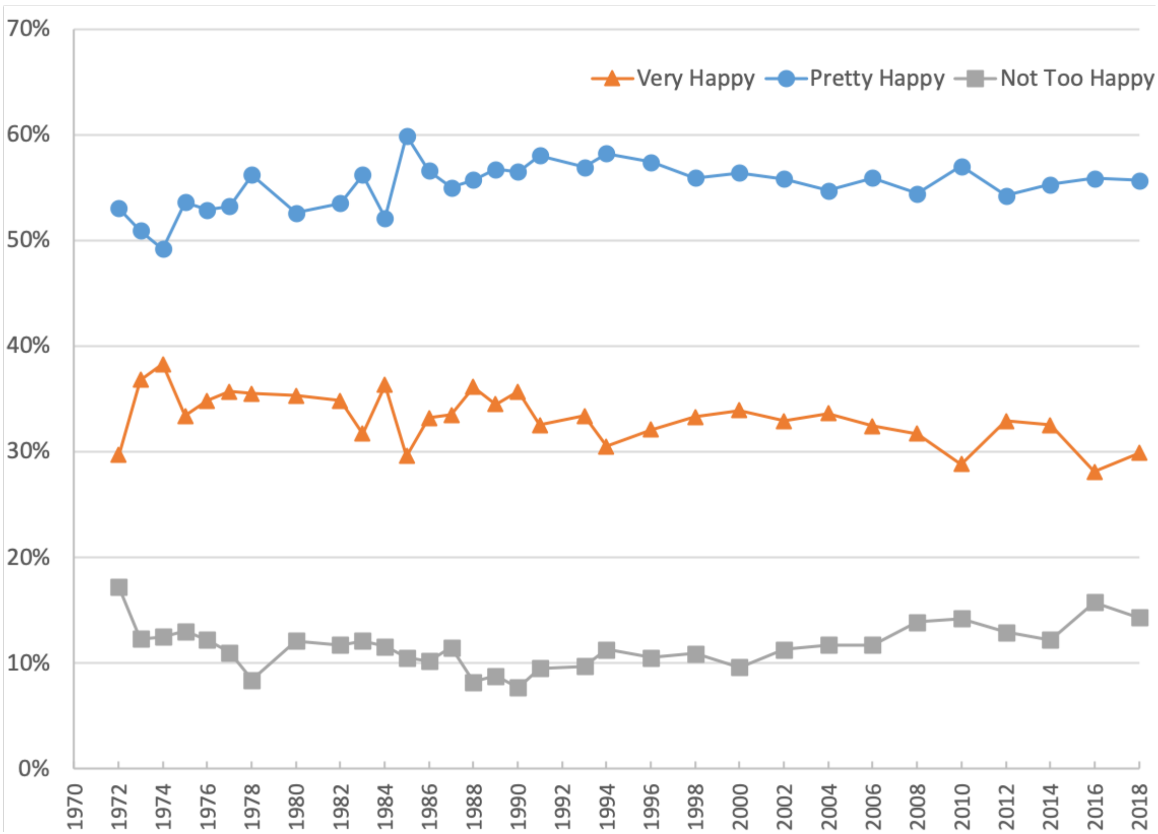


Figure 18: General happiness in the United States, 1972–2018. Survey response from the survey question: "Taken all together, how would you say things are these days—would you say that you are very happy, pretty happy, or not too happy?" Data from the General Social Surveys.

Table 14: Estimates for Easterlin Paradox

	MAP	Mean	Median	95% HPDI
$\frac{u_{C_{others}}}{u_C}$	-0.65	-0.65	-0.65	[-0.90, -0.41]
$\frac{u_H}{u_C} + \frac{u_{C_{others}}}{u_C}$	-1.26	-1.27	-1.27	[-1.60, -0.97]

of keeping up with the Joneses applies instantly once others' spending changes become known to the agent.

The effects of habit formation and keeping up with the Joneses on utility, to a first-order approximation, are captured by u_H/u_C and $u_{C_{others}}/u_C$. My estimates of these two ratios are both greater than -1 at the 95% level (Tables 14 and 8), which means that habit formation and keeping up with the Joneses alone cannot fully explain the paradox.

The long-run zero income–happiness slope dictates that

$$\frac{u_H}{u_C} + \frac{u_{C_{others}}}{u_C} = -1, \tag{2}$$

which my estimates support at the 95% level (Table 14). The point estimate of the left-hand side of the above equation is less than -1, which, aside from statistical precision considerations, provides the potential for the explanation of the paradox to be consistent with the general improvement of happiness-altering non-income factors (Di Tella and MacCulloch, 2008) and with the slightly negative long-run income–happiness slope in the United States (Stevenson and Wolfers, 2008; Firebaugh and Tach, 2012). Because I take the modal view of the Easterlin paradox for illustration purposes, I focus on the scenario where they sum to -1. For concreteness, let us choose $u_H/u_C = -0.55$ and $u_{C_{others}}/u_C = -0.45$, both of which are within their respective 95% HPDIs. As long as their sum is -1, the exact values of the two ratios only slightly affect the steady-state level of happiness and the convergence speed to the steady states, both of which do not alter the income–happiness pattern that is at the heart of the Easterlin paradox.

The intuition of equation (2) is that, to a first-order approximation, habit formation and keeping up with the Joneses entirely cancel the happiness effect of permanent consumption changes in the long run. To illustrate, imagine a single-episode-growth scenario where the economy was at some steady state such that its residents were having some constant level of happiness before the instant t_0. Imagine that at t_0 the economy grows so that everyone's consumption permanently increases by a small amount of Δc for all instants starting from t_0 onwards (Figure 19a). As a result, to a first-order approximation,[43] the agent's affect goes up by $u_C \Delta c$ at t_0. As time passes, people get used to this higher level of consumption, resulting in the buildup of internal habit that pulls affect down (Figure 19b). At the instant before keeping up with the Joneses kicks in, $t_0 + k$, the remaining effect of the higher consumption on affect is lowered to $\left[u_C + (1-\omega)\left(1 - e^{-\theta k}\right) u_H\right] \Delta c$. At $t_0 + k$, the agent realizes that everyone else also enjoys the same higher level of consumption as she does and feels worse as a result of social comparison—keeping up with the Joneses—which further pushes the gain of affect down to $\left[u_C + (1-\omega)\left(1 - e^{-\theta k}\right) u_H + u_{C_{others}}\right] \Delta c$. After that, external habit comes into play and, together with internal habit, erodes the remaining gain of affect until it completely disappears.

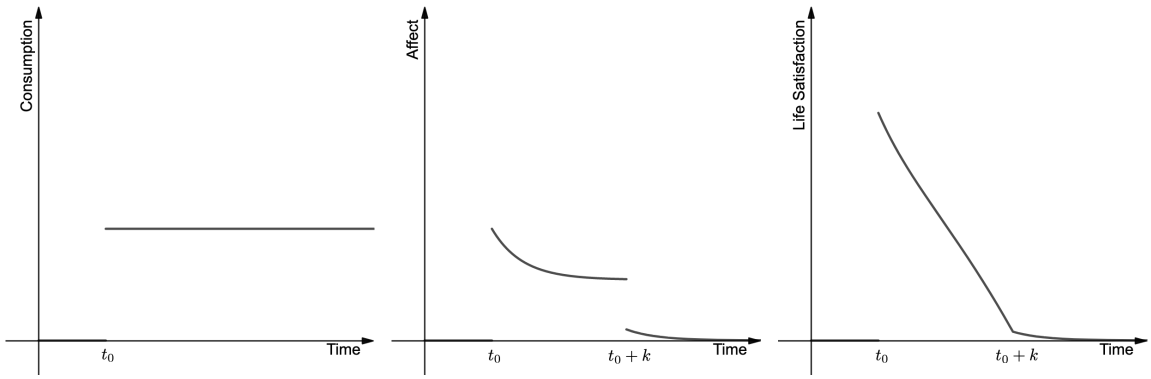
Integrating affect discounted by time preference,[44] I get life satisfaction (Figure 19c), the second measure of happiness. From the behavior of affect as analyzed above, it should come as no surprise that life satisfaction first increases, then gradually decreases to its previous level. For later reference, let me call this pattern the wear-off effect: Over time, habit formation and keeping up with the Joneses cancel out the happiness brought by a permanently higher level of consumption or income. Under the influence of this effect, in the distant future, the agent feels as if the permanent change in consumption did not happen, even though it did. Note that before the wear-off effect reaches its full potency, higher consumption indeed brings higher levels of happiness. The wear-off effect reveals the key intuition of how habit formation and keeping up with the Joneses resolve the income–happiness paradox: the agent feels happier in the short run after consumption increases, but in the long run, this excitement fades away.

In reality, economies tend to grow over time, and, as a result, people typically earn more and consume more over time. To capture the key aspect of this phenomenon, suppose everyone increases her or his consumption permanently by Δc every one year after t_0 (Figure 20a). Figures 20b and 20c plot the agent's happiness as time progresses. It should be of no surprise that habit formation wears off the gain of happiness within each year, as in the one-episode-growth scenario above.[45] What is

[43]Throughout my analysis in this section, I limit my attention to first-order approximations, because there is no point going beyond it when the value of $u_{C_{others}}/u_C$ is chosen based on its first-order elicitation. This is a good approximation when the change Δc is small, which I assume here.

[44]I calibrate the time discount rate to 0.15, based on my estimate of this parameter. See section C of the appendix for details.

[45]The discontinuities of the utility changes at the start of each year after t_0 are purely based on the simplifying assumption that consumption permanently increases at the start of each year after t_0, which is not essential. One can imagine that after an income increase, one increases consumption



(a) Consumption (b) Affect (c) Life satisfaction

Figure 19: One-episode growth.

new is the dynamics of happiness: Instead of eventually going back to its previous level, happiness gradually builds up and then plateaus. Again, for later reference, let me label these two phases of the happiness dynamics transition effect and plateau effect. The transition effect exists, contrasting with the decreasing trend as in the one-episode-growth scenario, because each year the growing consumption brings a new episode of the wear-off effect whose initial happiness-enhancing phase[46] stacks onto those from previous years. As habit formation and keeping up with the Joneses take effect, the happiness-reducing momentum gradually builds up and eventually cancels out the happiness-enhancing momentum that is driving the transition effect, leading the agent to the happiness plateau. The instant when such exact cancellation first happens is precisely the moment when the wear-off effect brought by the consumption increase of t_0 is in full swing for the first time.

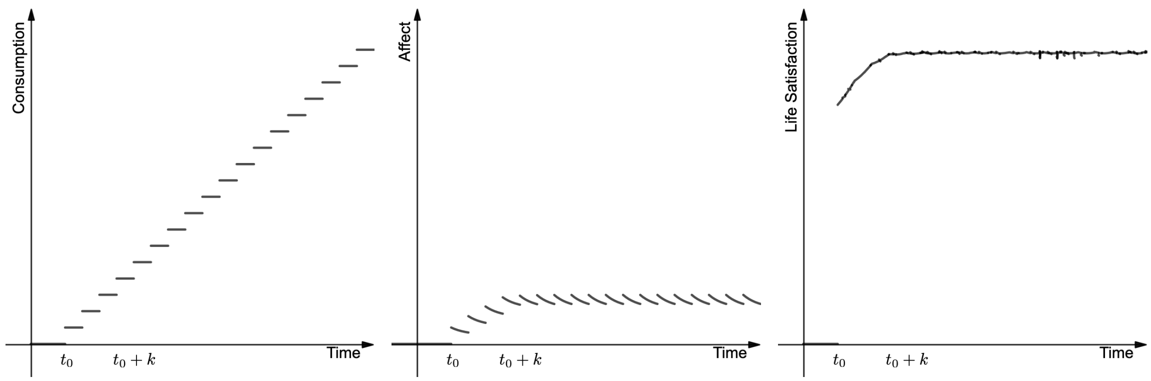
Because the wear-off effect on happiness is proportional to Δc,[47] the transition and plateau effects on happiness are also proportional to Δc (Figure 21). Let me call this the level effect: Higher consumption growth leads to higher levels of happiness during both the transition and the plateau phases.

The level effect explains the positive cross-sectional correlation between income and happiness. Richer people or countries enjoy higher levels of consumption and

slowly (e.g., because of habit formation) so that the jumps in the utility changes disappear. All my analysis remains unchanged in this scenario. Here I do not employ this to highlight the wear-off effect of habit formation.

[46]The time interval when happiness increases after t_0 in Figures 19b and 19c.

[47]This is because my analysis here focuses on first-order approximations. To the extent that people's marginal utility of consumption is always positive, the analysis still holds: Even though the utility difference of the high and low consumption changes will be smaller, the difference remains positive.



(a) Consumption (b) Affect (c) Life satisfaction

Figure 20: Multi-episode growth.

therefore are on higher happiness curves, which implies that they are happier than poorer people or countries on the lower happiness curves.[48] Constant economic fluctuations of the reality cause consumption to fluctuate, frequently putting the agent into transition phases. The transition effect, therefore, explains the short-run positive correlation between income and happiness. Note that regardless of income (or consumption) increase or decrease, transition effect always implies a positive relationship between income and happiness. Through the level effect, faster-growing economies such as developing countries tend to enjoy larger increases in happiness during the transition phase, as observed in reality (Frijters, Haisken-DeNew and Shields, 2004). The plateau effect explains the long-run nil correlation between income and happiness. Even though income constantly fluctuates, it fluctuates around its trend. This trend growth determines the plateau level of happiness, which underpins the long-run trend of happiness. To put it another way, in the long run, the trend of the happiness curve flattens even though consumption and income continually grow, hence the nil correlation.

To deepen the intuition, let me draw an analogy—run against an escalator. Imagine that you are about to run up a down escalator at a uniform speed of Δc stairs per unit of time. The escalator is initially stationary and, once you step onto it, will

[48]My above analysis has assumed a representative agent, purely to isolate the key mechanisms for resolving the paradox from potential complications implied in heterogeneities of the reference groups for social comparison. The analysis carries through with reasonable specifications of the reference groups. For example, one can assume that the income of the reference group changes by the same amount as does the (heterogeneous) agent's income. Note also that the level effect remains if everyone in the economy gets proportionally richer, implying that rising income inequality is not a requisite for resolving the paradox with the help of habit formation as is suspected in the literature (see, e.g., Clark, 2016).

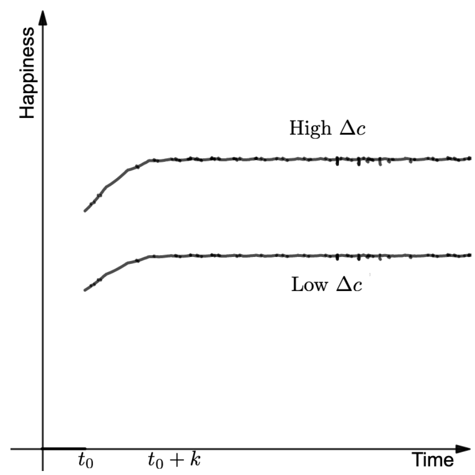


Figure 21: Level effect.

gradually accelerate to the speed of

$$\frac{u_H + u_{C_{others}}}{u_C} \Delta c = -\Delta c$$

stairs per unit of time. Suppose the escalator is long enough so that it catches up to (the negative of) your speed before you can reach the top. Your elevation represents happiness. The number of stairs you run represents the level of consumption, which implies that the speed you run, Δc stairs per unit of time, embodies the consumption growth, Δc per unit of time. The escalator symbolizes the joint effect of habit formation and keeping up with the Joneses.

With this analogy, allow me to propose and resolve another paradox, the Escalator paradox, which parallels the Easterlin paradox (Table 15). The Escalator (Easterlin) paradox states that running more stairs (increasing income) raises elevation (happiness) in the cross section and short run but not so in the long run. Why is this the case? In the long run, the escalator (habit formation and keeping up with the Joneses) eventually catches up to your running speed (consumption growth), after which the additional stairs you run (additional consumption you enjoy) do not affect your elevation (happiness). In the short run, you gain elevation (happiness) because your running speed (consumption growth) is faster than that of the escalator (the happiness-canceling effect of habit formation and keeping up with the Joneses). In the cross section, people who run faster (people or countries that are richer) are more elevated (happier) because the absolute difference between their running speed (consumption growth) and the speed of the escalator (the happiness-canceling effect of habit formation and keeping up with the Joneses) is larger during the transition

Table 15: Two Paradoxes

Time Frame	Easterlin Paradox	Escalator Paradox	Explanation
Long run	Why doesn't increasing *income* raise **happiness**?	Why doesn't running more *stairs* raise **elevation**?	Plateau effect
Short run	Why does increasing *income* raise **happiness**?	Why does running more *stairs* raises **elevation**?	Transition effect (+ fluctuation)
Cross section	Why are *richer* people/countries **happier**?	Why are *faster* people **more elevated**?	Level effect

phase, which eventually accumulates to a higher elevation (happiness level).

How does the above discussion speak to the questions that motivated the paradox: Does money buy happiness (Easterlin, 1973), and does economic growth improve human lot (Easterlin, 1974)? To phrase the questions in a slightly more accurate way, should we promote economic growth if happiness will eventually stop growing with economic growth? The answer is yes. Happiness decreases if the economy shrinks or grows at slower speeds. In other words, economic growth initially drives up happiness and eventually *maintains* it. If the economy grows slower or even shrinks, the resulting slower consumption growth will cause happiness to drop and to plateau at a level lower than the level at which it would plateau had the economy not slowed down.

Three implications for the alternative views of the Easterlin paradox follow from the discussion. First, the strong effects of habit formation and keeping up with the Joneses support the position that we should take the low happiness–income gradient seriously. Second, the existence of both types of habit formation and keeping up with the Joneses is consistent with the empirical observation that the aggregate-time-series happiness–income gradient is lower than the cross-sectional happiness–income gradient: Internal habit formation pushes down the gradient in both the cross section and the aggregate time series, while external habit and keeping up with the Joneses push down the gradient for the aggregate time series but not for the cross section. Third, the cross-country heterogeneities of happiness–income gradients for aggregate time series can potentially be explained by heterogeneous values of the preference parameters governing the effects of habit formation and keeping up with the Joneses.[49]

[49]Cross-country heterogeneities in the improvement of non-income happiness-altering factors could be another potential explanation.

6 Discussion and Robustness

In this section, I discuss the robustness of my results to departures from certain assumptions underlying the above results.

6.1 Demographic Heterogeneity

I collected information on the age, gender, household size, and household income of my respondents. It is interesting to see how my estimates vary with the demographics. Modifying the statistical model to allow the demographic variables to shift the means of the parameter distributions, I find that the demographics do not affect my estimates (Table 16). In particular, 0 is included in all of the 95% HPDIs of the estimated effects of demographic variables except that of household income on $\frac{Hu_{HH}}{u_H}$. After accounting for multiple hypothesis testing, this effect vanishes.[50]

This result supports the view that the parameters I elicited are deep preference parameters that do not vary with sociodemographic characteristics. Because the ratios of utility derivatives are dependent on the spending profiles, the fact that the estimates do not vary with the demographics of the respondents and, therefore, their heterogeneous spending profiles in reality is reassuring, for it implies that my respondents understood the hypothetical situations of the survey and were able to answer the survey questions without letting their own demographic situations confound their responses.

6.2 Finite Horizon

I assumed infinite horizon in my general habit formation model, as do almost all current habit formation models in the literature. To investigate the effect of this assumption on my results, I re-derived all my elicitations of the preference parameters under finite horizon, resulting in minimal changes: no change for the elicitations of some parameters and tiny changes for the rest.[51] I reran my estimation under the finite-horizon elicitations and found that the estimates are almost the same as those in the infinite horizon (Table 17).

[50] The adjusted probability of the estimate greater than 0 under the Holm algorithm for this effect is 0.39.

[51] The thresholds for the habit depreciation rate and external habit mixture coefficient are exactly the same for both time horizons. The changes to the thresholds of other parameters are simply replacing 1 by $1 - e^{-\rho T}$ or $1 - e^{-(\rho+\theta)T}$ or $1 - e^{-(\rho+2\theta)T}$, all of which are close or very close to 1 under reasonable values of T, the finite time horizon of interest.

Table 16: Effect of Demographics on Parameter Estimates

	Omitted Category	Age	Gender	Household Size	Household Income
Habit Depreciation Rate	1.31	0.00	-0.24	0.13	-0.04
	[0.99, 1.65]	[-0.02, 0.01]	[-0.67, 0.20]	[-0.03, 0.29]	[-0.10, 0.01]
External Habit Mixture Coefficient	0.13	0.00	0.06	-0.04	-0.02
	[0.00, 0.25]	[-0.01, 0.01]	[-0.13, 0.22]	[-0.11, 0.03]	[-0.04, 0.00]
$-u_H/u_C$	0.67	0.00	-0.09	0.05	0.01
	[0.50, 0.85]	[-0.01, 0.01]	[-0.33, 0.14]	[-0.04, 0.13]	[-0.02, 0.04]
Hu_{HH}/u_H	6.62	-0.02	1.00	-0.58	0.23
	[5.94, 8.12]	[-0.09, 0.04]	[-0.02, 1.48]	[-0.95, 0.02]	[0.03, 0.44]
u_{CH}/u_{HH}	-0.91	0.00	0.10	0.05	0.02
	[-1.12, -0.68]	[-0.01, 0.01]	[-0.19, 0.40]	[-0.06, 0.16]	[-0.02, 0.06]
u_{CC}/u_{HH}	4.04	0.03	-0.78	-0.32	-0.04
	[2.98, 4.89]	[-0.03, 0.09]	[-1.50, 0.46]	[-0.86, 0.16]	[-0.22, 0.12]
$u_{C^{others}}/u_H$	1.07	-0.02	0.06	0.18	0.05
	[0.52, 1.64]	[-0.05, 0.01]	[-0.68, 0.81]	[-0.07, 0.46]	[-0.04, 0.14]
Time Discount Rate	0.09	0.00	0.12	0.02	0.00
	[-0.09, 0.25]	[-0.01, 0.00]	[-0.09, 0.35]	[-0.06, 0.10]	[-0.02, 0.03]

Notes: The omitted category is that of 40-year-old males who live in three-member households with \$50,001–\$60,000 annual household income. 95% HPDIs below MAP estimates.

Table 17: Finite Horizon Estimates

	MAP	Mean	Median	HPDI
Habit Depreciation Rate	1.12	1.11	1.11	[0.90, 1.32]
External Habit Mixture Coefficient	0.16	0.16	0.16	[0.07, 0.26]
$-u_H/u_C$	0.59	0.60	0.59	[0.49, 0.71]
Hu_{HH}/u_H	7.66	7.77	7.75	[6.82, 8.73]
u_{CH}/u_{HH}	-0.86	-0.85	-0.86	[-1.00, -0.71]
u_{CC}/u_{HH}	3.83	3.85	3.85	[3.12, 4.57]
$u_{C_{others}}/u_H$	1.15	1.14	1.14	[0.78, 1.51]
$u_{C_{others}}/u_C$	-0.67	-0.68	-0.68	[-0.93, -0.42]
$u_{CH}u_H/u_{HH}u_C$	0.49	0.51	0.51	[0.34, 0.69]
$u_{CH}u_C/u_{CC}u_H$	0.37	0.38	0.37	[0.24, 0.56]
$Hu_Hu_{CH}/u_C\left(u_H + Hu_{HH}\right)$	0.45	0.45	0.45	[0.31, 0.61]
$Cu_Cu_{CH}/u_H\left(u_C + Cu_{CC}\right)$	-0.23	-0.24	-0.24	[-0.33, -0.16]
Time Discount Rate	0.15	0.14	0.14	[0.04, 0.25]

Note: The time horizon is 30 years in the future, because I instructed the respondents in the survey: "... think of ... 'Future' as the next 30 years." 99% HPDIs are reported for the four statistics for testing additive and multiplicative habits. 95% HPDIs are reported for other parameters.

6.3 Additional Attention Checks

I have already utilized the explicit attention checks to screen out respondents who did not understand the hypothetical situation or the monthly spending graphs through quizzes and deleted from the sample the responses of those who sped through the survey or were located outside of the United States. In this section, I make use of implicit attention checks to see whether a potential lack of attention biased my results. Because the implicit attention checks are not perfect proxies for attention, I apply them successively, from the relatively more reliable to the relatively less reliable.

At the end of the survey, I asked once again the quiz on the basic hypothetical situation of the survey. There were 120 respondents in wave one and 54 in wave two who made at least one mistake in answering the five-question quiz. Deleting these responses from my sample did not significantly change my estimates (Table 18), except that external habit no longer exists in the strict case, although it still exists in the lenient case.

I asked demographic questions in both waves. Within the relatively moderate amount of time that separated the two waves, the demographics should not have changed. In other words, the time consistency of answers to the demographic ques-

Table 18: Estimates Using Subsample of No Failed Quiz

	MAP	Mean	Median	HPDI/HPMI
Habit Depreciation Rate	1.11	1.12	1.12	[0.86, 1.37]
External Habit Mixture Coefficient	0.08	0.10	0.10	[0.00, 0.21]
$-u_H/u_C$	0.69	0.70	0.70	[0.56, 0.84]
Hu_{HH}/u_H	8.21	8.30	8.27	[7.02, 9.61]
u_{CH}/u_{HH}	-0.80	-0.81	-0.81	[-1.00, -0.60]
u_{CC}/u_{HH}	4.45	4.42	4.42	[2.82, 5.97]
$u_{C_{others}}/u_H$	1.53	1.52	1.52	[0.80, 2.27]
$u_{C_{others}}/u_C$	-1.02	-1.07	-1.05	[-1.64, -0.52]
$u_{CH}u_H/u_{HH}u_C$	0.56	0.57	0.56	[0.33, 0.82]
$u_{CH}u_C/u_{CC}u_H$	0.24	0.27	0.26	[0.13, 0.51]
$Hu_Hu_{CH}/u_C\left(u_H + Hu_{HH}\right)$	0.50	0.51	0.50	[0.30, 0.73]
$Cu_Cu_{CH}/u_H\left(u_C + Cu_{CC}\right)$	-0.18	-0.20	-0.19	[-0.37, -0.10]
$\text{sgn}\left(Q_H\right)$	-1.00	-1.00	-1.00	[-1.00, -1.00]
$\text{sgn}\left(Q_{EH}\right)$ (strict)	1.00	0.11	1.00	0.55
$\text{sgn}\left(Q_{EH}\right)$ (lenient)	-1.00	-1.00	-1.00	[-1.00, -1.00]
Time Discount Rate	0.11	0.11	0.11	[-0.06, 0.26]

Notes: 99% HPDIs or HPMIs are reported for the seven statistics for testing additive and multiplicative habits and existence of habit formation. 95% HPDIs are reported for other parameters. In cases where an estimate is not significant at the 90% level, the probability of the estimate greater than 0 is shown.

Table 19: Estimates Using Subsample of No Demographic Mistake

	MAP	Mean	Median	HPDI/HPMI
Habit Depreciation Rate	1.15	1.13	1.13	[0.85, 1.41]
External Habit Mixture Coefficient	0.10	0.12	0.12	[0.00, 0.24]
$-u_H/u_C$	0.67	0.67	0.67	[0.52, 0.82]
Hu_{HH}/u_H	8.35	8.47	8.45	[7.27, 9.92]
u_{CH}/u_{HH}	-0.87	-0.86	-0.86	[-1.06, -0.65]
u_{CC}/u_{HH}	4.20	4.15	4.15	[2.47, 5.86]
$u_{C_{others}}/u_H$	1.45	1.44	1.43	[0.64, 2.22]
$u_{C_{others}}/u_C$	-0.91	-0.96	-0.94	[-1.54, -0.41]
$u_{CH}u_H/u_{HH}u_C$	0.57	0.57	0.57	[0.34, 0.83]
$u_{CH}u_C/u_{CC}u_H$	0.29	0.33	0.31	[0.14, 0.67]
$Hu_Hu_{CH}/u_C\left(u_H + Hu_{HH}\right)$	0.51	0.51	0.51	[0.30, 0.75]
$Cu_Cu_{CH}/u_H\left(u_C + Cu_{CC}\right)$	-0.20	-0.23	-0.22	[-0.46, -0.10]
$\text{sgn}\left(Q_H\right)$	-1.00	-1.00	-1.00	[-1.00, -1.00]
$\text{sgn}\left(Q_{EH}\right)$ (strict)	-1.00	-0.31	-1.00	0.34
$\text{sgn}\left(Q_{EH}\right)$ (lenient)	-1.00	-1.00	-1.00	99: [-1.00, -1.00]
Time Discount Rate	0.12	0.11	0.12	[-0.05, 0.27]

Notes: 99% HPDIs or HPMIs are reported for the seven statistics for testing additive and multiplicative habits and existence of habit formation. 95% HPDIs are reported for other parameters. In cases where an estimate is not significant at the 90% level, the probability of the estimate greater than 0 is shown.

tions can serve as an implicit attention check. Applying this check eliminated another 18 responses from the remaining sample. My estimates were essentially unchanged (Table 19) except, again, that the evidence for existence of external habit was weaker.

A third implicit attention check is that people should be indifferent toward the universes when there is no difference between them. In the time discount rate question[52], past spendings are the same in both universes, where the respondents should choose the same past experience. This led me to further delete 104 and 19 responses from waves one and two, respectively. The estimates remain in line with my baseline estimates (Table 20). Two points are worth noting. The external habit exists at the 90% level in the strict case, as in my main results. The credible intervals inflate as a result of the much smaller sample size. This reduces the precision of some of my statistics for the tests of the existence of habit formation and of additive and multiplicative habits. However, the rest of the statistics remain accurate enough to keep

[52]See section C of the appendix.

Table 20: Estimates Using Subsample of Same Past Experience

	MAP	Mean	Median	HPDI/HPMI
Habit Depreciation Rate	1.27	1.27	1.27	[0.84, 1.74]
External Habit Mixture Coefficient	0.27	0.27	0.26	[0.00, 0.50]
$-u_H/u_C$	0.49	0.50	0.50	[0.30, 0.69]
Hu_{HH}/u_H	9.12	8.78	8.87	[7.35, 10.00]
u_{CH}/u_{HH}	-0.69	-0.68	-0.68	[-0.95, -0.41]
u_{CC}/u_{HH}	4.33	4.41	4.39	[2.31, 6.57]
$u_{C_{others}}/u_H$	0.93	0.98	0.98	[-0.20, 2.12]
$u_{C_{others}}/u_C$	-0.44	-0.49	-0.47	[-1.11, 0.12]
$u_{CH}u_H/u_{HH}u_C$	0.32	0.34	0.33	[0.11, 0.60]
$u_{CH}u_C/u_{CC}u_H$	0.22	0.35	0.31	[0.06, 0.92]
$Hu_Hu_{CH}/u_C\left(u_H+Hu_{HH}\right)$	0.28	0.30	0.30	[0.10, 0.54]
$Cu_Cu_{CH}/u_H\left(u_C+Cu_{CC}\right)$	-0.14	-0.17	-0.16	[-0.44, -0.04]
$\operatorname{sgn}\left(Q_H\right)$	-1.00	-0.97	-1.00	95: [-1.00, -1.00] 99: [-1.00, 1.00]
$\operatorname{sgn}\left(Q_{EH}\right)$ (strict)	-1.00	-0.85	-1.00	90: [-1.00, -1.00] 95: [-1.00, 1.00]
$\operatorname{sgn}\left(Q_{EH}\right)$ (lenient)	-1.00	-1.00	-1.00	99: [-1.00, -1.00]
Time Discount Rate	0.10	0.10	0.10	[-0.27, 0.46]

Notes: 99% HPDIs or HPMIs are reported for the seven statistics for testing additive and multiplicative habits and existence of habit formation. 95% HPDIs are reported for other parameters. Additional HPDI or HPMI of relevant level is also shown to help determine the confidence level.

my results robust.

Finally, I use a measure of response consistency across the waves as an attention check. Considering that this attention check involved more speculation, I eliminated only those who gave at least one polar response—any response corresponded to the first (last) extreme range of parameter values in wave one and the last (first) in wave two. This check led to the deletion of another 31 responses from both waves, resulting in further expansion of the credible intervals of my estimates, but, again, the estimates are not significantly different, and my results remain robust (Table 21).

6.4 Response Error with Nonzero and Wave-Varying Mean

In the statistical model, I assume that the response error is zero-mean across both waves of the survey. Relaxing this assumption, I arrive at a statistical model with response errors of nonzero means that potentially vary across waves. Without loss

Table 21: Estimates Using Sample of No Polar Response

	MAP	Mean	Median	HPDI/HPMI
Habit Depreciation Rate	1.43	1.43	1.43	[0.95, 1.96]
External Habit Mixture Coefficient	0.15	0.22	0.20	[0.00, 0.48]
$-u_H/u_C$	0.54	0.53	0.53	[0.25, 0.78]
Hu_{HH}/u_H	8.82	8.54	8.61	[6.94, 10.00]
u_{CH}/u_{HH}	-0.73	-0.72	-0.73	[-0.99, -0.44]
u_{CC}/u_{HH}	4.69	4.56	4.57	[1.91, 7.23]
$u_{C_{others}}/u_H$	1.02	1.07	1.05	[-0.25, 2.44]
$u_{C_{others}}/u_C$	-0.45	-0.56	-0.53	[-1.36, 0.20]
$u_{CH}u_H/u_{HH}u_C$	0.36	0.38	0.37	[0.08, 0.71]
$u_{CH}u_C/u_{CC}u_H$	0.10	0.34	0.31	[0.07, 1.41]
$Hu_Hu_{CH}/u_C\,(u_H + Hu_{HH})$	0.33	0.34	0.33	[0.08, 0.65]
$Cu_Cu_{CH}/u_H\,(u_C + Cu_{CC})$	-0.16	-0.19	-0.17	[-0.67, -0.03]
$\mathrm{sgn}\,(Q_H)$	-1.00	-0.99	-1.00	99: [-1.00, -1.00]
$\mathrm{sgn}\,(Q_{EH})$ (strict)	-1.00	-0.32	-1.00	0.34
$\mathrm{sgn}\,(Q_{EH})$ (lenient)	-1.00	-1.00	-1.00	99: [-1.00, -1.00]
Time Discount Rate	0.02	-0.05	-0.02	[-0.62, 0.47]

Notes: 99% HPDIs or HPMIs are reported for the seven statistics for testing additive and multiplicative habits and existence of habit formation. 95% HPDIs are reported for other parameters. Additional HPDI of relevant level is also shown to help determine the credible level. In cases where an estimate is not significant at the 90% level, the probability of the estimate greater than 0 is shown.

Table 22: Means of Response Errors in Wave One

μ_ε of	MAP	Mean	Median	95% HPDI
Habit Depreciation Rate	-0.01	-0.02	-0.02	[-0.25, 0.21]
External Habit Mixture Coefficient	-0.05	-0.04	-0.04	[-0.13, 0.06]
$-u_H/u_C$	-0.01	-0.01	-0.01	[-0.10, 0.07]
Hu_{HH}/u_H	-0.85	-1.15	-1.03	[-3.06, 0.48]
u_{CH}/u_{HH}	0.08	0.08	0.08	[-0.07, 0.23]
u_{CC}/u_{HH}	-0.41	-0.41	-0.40	[-1.14, 0.34]
$u_{C^{others}}/u_H$	-0.13	-0.13	-0.13	[-0.52, 0.25]
Time Discount Rate	-0.05	-0.05	-0.05	[-0.16, 0.07]

Note: μ_ε denotes mean of response error.

of generality,[53] the joint distribution of parameter $\tilde{x}$ for individual i in both waves becomes

$$\begin{bmatrix} \tilde{x}_{i,1} \\ \tilde{x}_{i,2} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x + \mu_\varepsilon \\ \mu_x - \mu_\varepsilon \end{bmatrix}, \begin{bmatrix} \sigma_x^2 + \sigma_{\varepsilon_x}^2 + \sigma_{\varepsilon_{x,1}}^2 & \sigma_x^2 + \sigma_{\varepsilon_x}^2 \\ \sigma_x^2 + \sigma_{\varepsilon_x}^2 & \sigma_x^2 + \sigma_{\varepsilon_x}^2 + \sigma_{\varepsilon_{x,2}}^2 \end{bmatrix} \right).$$

The estimates of the means of the response errors in wave one or equivalently negative of the means of the response errors in wave two are indistinguishable from zero at the 5% level (Table 22). This implies that my estimates are robust to the nonzero-wave-varying-mean response errors, which is exactly what I find (Table 23).

7 Conclusion

In this paper, I provide an extensive set of survey evidence for habit formation. I find that people's spending behavior exhibits habit formation, partly explaining why adding habit formation tends to improve the explanatory power of models. Considering that there are phenomena that current habit formation models are struggling to match, I suggest that future models should be built upon habit formation rather than ignoring it, especially when it alone does *not* seem to be able to match data.

Both internal and external habits exist, with the latter accounting for a small fraction (17%) of habit. This implies that in terms of micro validity, internal habit is a better choice than external habit. Better still, a composition of internal and

[53]Only two means, one for each wave, can be identified. The specification here identifying the difference of the means is equivalent to a specification that specifies the two means using two parameters, one for each mean. If the two means are different, μ_ε should be statistically different from 0.

external habits with the estimated weights could potentially deliver a superior match with people's behavior.

Habit depreciates by about two thirds per year. Future habit formation models should match or at least report how the model performs under this speed of habit depreciation. Calibrating to this habit adjustment speed could potentially reduce the fit between model and data. When happens, it signals, I believe, that habit formation alone[54] is not sufficient to fully explain the data at hand, necessitating the introduction of additional features to the model.

None of the current habit formation models are consistent with people's behavior because their specific utility functions fail to pass the four tests on the validities of the preference specifications. This justifies a search for habit utility functions that match the survey evidence, which potentially can explain phenomena current habit models cannot.

In terms of welfare effects, habit formation is as important as keeping up with the Joneses. Both external habit and keeping up with the Joneses are part of people's behavior. Keeping up with the Joneses dominates altruism.

Combining habit formation with keeping up with the Joneses can generate the patterns of income (consumption) and happiness highlighted by the Easterlin paradox. This implies that happiness can increase with income but only for a while before the wear-off effect induced by habituation and social comparison ends the happiness-enhancing phase. Level and transition effects explain the cross-sectional and short-run positive correlations of income and happiness, whereas plateau effect explains the long-run nil correlation. Although happiness eventually plateaus even though income keeps growing, continued income growth is necessary to maintain the plateaued level of happiness.

Future research could explore how my results compare with estimates from different populations. Heterogeneous cross-country estimates could potentially explain the cross-country heterogeneities in happiness–income dynamics. It would also be interesting to see how other ways of eliciting the same preference parameters I elicited here affect my results.

[54] Alone in the sense that one is exploring adding habit formation to help explain the data at hand.

Table 23: Estimates under Nonzero and Wave-Varying Means of Response Errors

	MAP	Mean	Median	HPDI/HPMI
Habit Depreciation Rate	1.11	1.12	1.12	[0.86, 1.36]
External Habit Mixture Coefficient	0.20	0.19	0.19	[0.07, 0.31]
$-u_H/u_C$	0.63	0.63	0.62	[0.51, 0.75]
Hu_{HH}/u_H	8.55	8.81	8.69	[7.03, 10.85]
u_{CH}/u_{HH}	-0.90	-0.89	-0.89	[-1.05, -0.72]
u_{CC}/u_{HH}	4.03	4.05	4.05	[3.15, 5.00]
$u_{C_{others}}/u_H$	1.14	1.14	1.14	[0.74, 1.59]
$u_{C_{others}}/u_C$	-0.70	-0.71	-0.71	[-1.02, -0.42]
$u_{CH}u_H/u_{HH}u_C$	0.55	0.56	0.55	[0.37, 0.76]
$u_{CH}u_C/u_{CC}u_H$	0.34	0.36	0.35	[0.22, 0.56]
$Hu_Hu_{CH}/u_C\left(u_H + Hu_{HH}\right)$	0.50	0.50	0.50	[0.33, 0.68]
$Cu_Cu_{CH}/u_H\left(u_C + Cu_{CC}\right)$	-0.23	-0.23	-0.23	[-0.35, -0.15]
$\operatorname{sgn}\left(Q_H\right)$	-1.00	-0.97	-1.00	95: [-1.00, -1.00] 99: [-1.00, 1.00]
$\operatorname{sgn}\left(Q_{EH}\right)$ (strict)	-1.00	-0.85	-1.00	90: [-1.00, -1.00] 95: [-1.00, 1.00]
$\operatorname{sgn}\left(Q_{EH}\right)$ (lenient)	-1.00	-1.00	-1.00	99: [-1.00, -1.00]
Time Discount Rate	0.18	0.18	0.18	[0.04, 0.34]

Notes: 99% HPDIs or HPMIs are reported for the seven statistics for testing additive and multiplicative habits and existence of habit formation. 95% HPDIs are reported for other parameters. Additional HPDI or HPMI of relevant level is also shown to help determine the credible level.

References

Abel, Andrew B. 1990. "Asset Prices under Habit Formation and Catching up with the Joneses." *American Economic Review*, 80(2): 38–42.

Alvarez-Cuadrado, Francisco, Jose Maria Casado, and Jose Maria Labeaga. 2015. "Envy and Habits: Panel Data Estimates of Interdependent Preferences." *Oxford Bulletin of Economics and Statistics*, 78(4): 443–469.

Ameriks, John, Joseph Briggs, Andrew Caplin, Matthew D Shapiro, and Christopher Tonetti. 2019. "Long-Term-Care Utility and Late-in-Life Saving." *Journal of Political Economy.*

Barsky, Robert B, F Thomas Juster, Miles S Kimball, and Matthew D Shapiro. 1997. "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study." *The Quarterly Journal of Economics*, 112(2): 537–579.

Benjamin, Daniel J, Kristen B Cooper, Ori Heffetz, and Miles Kimball. 2019. "A Well-Being Snapshot in a Changing World." *American Economic Review Papers and Proceedings*, 109: 344–49.

Benjamin, Daniel J, Ori Heffetz, Miles S Kimball, and Alex Rees-Jones. 2012. "What Do You Think Would Make You Happier? What Do You Think You Would Choose?" *American Economic Review*, 102(5): 2083–2110.

Benjamin, Daniel J, Ori Heffetz, Miles S Kimball, and Nichole Szembrot. 2014. "Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference." *American Economic Review*, 104(9): 2698–2735.

Berinsky, Adam J, Gregory A Huber, and Gabriel S Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon. Com's Mechanical Turk." *Political Analysis*, 20(3): 351–368.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. "Stereotypes." *The Quarterly Journal of Economics*, 131(4): 1753–1794.

Campbell, John Y, and John H Cochrane. 1999. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior." *Journal of Political Economy*, 107(2): 205–251.

Carroll, Christopher D, Jody Overland, and David N Weil. 2000. "Saving and Growth with Habit Formation." *American Economic Review*, 341–355.

Cavallo, Alberto. 2017. "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers." *The American Economic Review*, 107(1): 283–303.

Chetty, Raj, and Adam Szeidl. 2016. "Consumption Commitments and Habit Formation." *Econometrica*, 84(2): 855–890.

Clark, Andrew E. 2016. "Adaptation and the Easterlin Paradox." In *Advances in happiness research*. 75–94. Springer.

Clark, Andrew E, Paul Frijters, and Michael A Shields. 2008. "Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles." *Journal of Economic literature*, 46(1): 95–144.

Cochrane, John H. 2017. "Macro-Finance." *Review of Finance*, 21(3): 945–985.

Constantinides, George M. 1990. "Habit Formation: A Resolution of the Equity Premium Puzzle." *Journal of Political Economy*, 98(3): 519–543.

Debreu, Gerard. 1972. "Smooth Preferences." *Econometrica*, 40(4): 603–615.

De Giorgi, Giacomo, Anders Frederiksen, and Luigi Pistaferri. 2019. "Consumption Network Effects." *The Review of Economic Studies*. rdz026.

Di Tella, Rafael, and Robert MacCulloch. 2008. "Gross National Happiness as an Answer to the Easterlin Paradox?" *Journal of Development Economics*, 86(1): 22–42.

Dou, Winston Wei, Andrew W Lo, Ameya Muley, and Harald Uhlig. 2017. "Macroeconomic Models for Monetary Policy: A Critical Review from a Finance Perspective."

Easterlin, Richard A. 1973. "Does Money Buy Happiness?" *The public interest*, 30: 3.

Easterlin, Richard A. 1974. "Does Economic Growth Improve the Human Lot? Some Empirical Evidence." In *Nations and households in economic growth.* 89–125. Elsevier.

Easterlin, Richard A. 1995. "Will Raising the Incomes of All Increase the Happiness of All?" *Journal of Economic Behavior & Organization*, 27(1): 35–47.

Easterlin, Richard A. 2017. "Paradox Lost?" *Review of Behavioral Economics*, 4(4): 311–339.

Firebaugh, Glenn, and Laura Tach. 2012. *Income, Age, and Happiness in America. Social trends in American life: Findings from the General Social Survey since 1972*, Princeton University Press.

Flavin, Marjorie, and Shinobu Nakagawa. 2008. "A Model of Housing in the Presence of Adjustment Costs: A Structural Interpretation of Habit Persistence." *The American Economic Review*, 98(1): 474–495.

Folger, Harry T. 1926. "The Effects of Mechanical Shock on Locomotion in Amoeba Proteus." *Journal of Morphology*, 42(2): 359–370.

Frijters, Paul, John P Haisken-DeNew, and Michael A Shields. 2004. "Money Does Matter! Evidence from Increasing Real Income and Life Satisfaction in East Germany Following Reunification." *American Economic Review*, 94(3): 730–740.

Fuhrer, Jeffrey C. 2000. "Habit Formation in Consumption and Its Implications for Monetary-Policy Models." *American Economic Review*, 367–390.

Gardner, Jonathan, and Andrew J Oswald. 2007. "Money and Mental Wellbeing: A Longitudinal Study of Medium-Sized Lottery Wins." *Journal of health economics*, 26(1): 49–60.

Grishchenko, Olesya V. 2010. "Internal Vs. External Habit Formation: The Relative Importance for Asset Pricing." *Journal of Economics and Business*, 62(3): 176–194.

Havranek, Tomas, Marek Rusnák, and Anna Sokolova. 2017. "Habit Formation in Consumption: A Meta-Analysis." *European Economic Review.*

Houthakker, Hendrik S, and Lester D Taylor. 1970. "Consumer Demand in the United States."

Jha, Saumitra, and Moses Shayo. 2019. "Valuing Peace: The Effects of Financial

Market Exposure on Votes and Political Attitudes." *Econometrica*, 87(5): 1561–1588.

Johnson, David, and John Ryan. 2018. "Amazon Mechanical Turk Workers Can Provide Consistent and Economically Meaningful Data."

Johnston, Robert J, Kevin J Boyle, Wiktor Adamowicz, Jeff Bennett, Roy Brouwer, Trudy Ann Cameron, W Michael Hanemann, Nick Hanley, Mandy Ryan, Riccardo Scarpa, et al. 2017. "Contemporary Guidance for Stated Preference Studies." *Journal of the Association of Environmental and Resource Economists*, 4(2): 319–405.

Kaiser, Casper, and Maarten Vendrik. 2018. "Different Versions of the Easterlin Paradox: New Evidence for European Countries."

Kapteyn, Arie, and Federica Teppa. 2003. "Hypothetical Intertemporal Consumption Choices." *The Economic Journal*, 113(486): C140–C152.

Kimball, Miles, and Robert Willis. 2006. "Happiness and Utility." Working Paper.

Kimball, Miles, Claudia R Sahm, and Matthew D Shapiro. 2008. "Imputing Risk Tolerance from Survey Responses." *Journal of the American Statistical Association*, 103(483): 1028–1038.

Kimball, Miles, Claudia R. Sahm, and Matthew D. Shapiro. 2009. "Risk Preferences in the PSID: Individual Imputations and Family Covariation." *American Economic Review*, 99(2): 363–368.

Kimball, Miles, Fumio Ohtake, Daniel H Reck, Yoshiro Tsutsui, and Fudong Zhang. 2015. "Diminishing Marginal Utility Revisited."

Kimball, Miles S, and Matthew D Shapiro. 2008. "Labor Supply: Are the Income and Substitution Effects Both Large or Both Small?" National Bureau of Economic Research.

Kozicki, Sharon, and P.A. Tinsley. 2002. "Dynamic Specifications in Optimizing Trend-Deviation Macro Models." *Journal of Economic Dynamics and Control*, 26(9): 1585 – 1611. Special issue in honour of David Kendrick.

Kuziemko, Ilyana, Michael I Norton, Emmanuel Saez, and Stefanie Stantcheva. 2015. "How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments." *The American Economic Review*, 105(4): 1478–1508.

Lettau, Martin, and Harald Uhlig. 2000. "Can Habit Formation be Reconciled with Business Cycle Facts?" *Review of Economic Dynamics*, 3(1): 79–99.

Lewbel, Arthur, Samuel Norris, Krishna Pendakur, and Xi Qu. 2018. "Consumption Peer Effects and Utility Needs in India." Boston College Department of

Economics.

Ljungqvist, Lars, and Harald Uhlig. 2015. “Comment on the Campbell-Cochrane Habit Model.” *Journal of Political Economy*, 123(5): 1201–1213.

Luttmer, Erzo F. P. 2005. “Neighbors as Negatives: Relative Earnings and Well-Being.” *The Quarterly Journal of Economics*, 120(3): 963–1002.

Martínez-Marquina, Alejandro, Muriel Niederle, and Emanuel Vespa. 2019. “Failures in Contingent Reasoning: The Role of Uncertainty.” *American Economic Review*, 109(9): 3437–3474.

Matyskova, Ludmila, Brian Rogers, Jakub Steiner, and Keh-Kuan Sun. 2019. “Habits as Adaptations: An Experimental Study.”

OECD. 2013. *OECD Guidelines on Measuring Subjective Well-Being.*

Oster, Emily, Ira Shoulson, and E Dorsey. 2013. “Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease.” *The American Economic Review*, 103(2): 804–830.

Perez-Truglia, Ricardo. 2019. “The Effects of Income Transparency on Well-Being: Evidence from a Natural Experiment.” National Bureau of Economic Research.

Ravina, Enrichetta. 2007. “Habit Persistence and Keeping up with the Joneses: Evidence from Micro Data.”

Rayo, Luis, and Gary S Becker. 2007. “Habits, Peers, and Happiness: An Evolutionary Perspective.” *American Economic Review*, 97(2): 487–491.

Sacks, Daniel W, Betsey Stevenson, and Justin Wolfers. 2012. “The New Stylized Facts about Income and Subjective Well-Being.” *Emotion*, 12(6): 1181.

Sahm, Claudia R. 2007. *Stability of Risk Preference.* Division of Research & Statistics and Monetary Affairs, Federal Reserve Board.

Smets, Frank, and Rafael Wouters. 2007. “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach.” *The American Economic Review*, 97(3): 586–606.

Stevenson, Betsey, and Justin Wolfers. 2008. “Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox.” 1.

Thurstone, Louis L. 1931. “The Indifference Function.” *The Journal of Social Psychology*, 2(2): 139–167.

Van Praag, Bernard M.S. 1971. “The Welfare Function of Income in Belgium: An Empirical Investigation.” *European Economic Review*, 2(3): 337 – 369.

Van Praag, Bernard M.S., and Paul Frijters. 1999. “The Measurement of Welfare and Well-Being; the Leyden Approach.” *Well-Being: The Foundations of Hedonic Psychology*, , ed. D. Kahneman, E. Diener and N. Schwarz, 413–432. Russel Sage Foundation.

Appendices

A Aggregation

What does the elicited preference parameters of individual respondents tell us about the preference parameters of the representative agent? This question is of particular interest because almost all current models with habit formation assume a representative agent. In this section, I show that given its existence, the representative agent has preference parameters that are averages of the individuals' preference parameters.

To aggregate individuals, the welfare of the individuals needs to be comparable with that of each other (comparability), and the representative agent's welfare should represent the average of individuals' welfare (representativeness). To formalize the idea of comparability, I assume that at the homogenous steady state of $\bar{C}_i = \bar{H}_i = \$5,000\ \forall i$, spending an extra dollar,[55] while holding habit constant, brings the same marginal utility to every one: $u_{i,C}\left(\bar{C}_i, \bar{H}_i\right) = u_{j,C}\left(\bar{C}_j, \bar{H}_j\right)\ \forall i, j$.

With the comparability of individuals' utilities, the representativeness of the representative agent means that $Nu_R\left(C_R, H_R\right) = \sum_i u_i\left(C_i, H_i\right)$ when $C_R = C_i = A$ and $H_R = H_i = B\ \forall i$ and $\forall A, B$ in the domains of the utility functions, where N is the number of individuals in the economy. That is, when the heterogeneities in behaviors (consumption and habit) are homogenized, the representative agent is the average individual agent in terms of welfare. To see what this condition means, first note that the difference between a representative-agent model and a heterogeneous-agent model is that in the former, everyone in the economy is the same, while in the latter, each individual can be different. Imagine that everyone in the heterogeneous economy becomes the same (i.e., homogeneous in consumption, habit, and utility function, etc.), the representative agent model should behave exactly the same as the homogenized heterogeneous agents model, and hence the equality of $Nu_R\left(C_R, H_R\right) = \sum_i u_i\left(C_i, H_i\right)$. Now, allowing the individual to be heterogeneous along the dimension of utility function after the normalization of the comparability condition, this representativeness condition simply requires that the representative agent represents the individuals along the welfare dimension, after controlling for consumption and habit.

[55] An epsilon dollar, to be exact.

A.1 Aggregation of Habit Depreciation Rate

Even though habit depreciation rate (θ) and habit (H) are mapped one to one at each instant of time[56] for any given consumption profile, there are infinitely many pairs of them that satisfy the representative agent's habit evolution ($\dot{H}_R = \theta_R \left(C_R - H_R\right)$), individuals' habit evolutions ($\dot{H}_i = \theta_i \left(C_i - H_i\right)$), the comparability condition, and the representativeness condition. The intuition is that while H depends on the habit depreciation rate, its steady-state level not. In other words, different θ leads to different H before a steady state is reached, and this difference vanishes after the steady state is reached.

To find the mapping between aggregate habit depreciation rate (θ_R) and individual habit depreciation rate (θ_i), imagine that everyone starts at the homogenous steady state and increases consumption by the same iota amount. That is, starting from $C_i = C_j = C_R = H_i = H_j = H_R \ \forall i, j$, increase consumption by the iota amount $\Delta C_i = \Delta C_j = \Delta C_R \ \forall i, j$. The changes to the utilities are

$$\begin{aligned}\Delta u_R \left(C_R, H_R\right) &= u_{R,C} \left(C_R, H_R\right) \Delta C_R + u_{R,H} \left(C_R, H_R\right) \Delta H_R \\ &= u_{R,C} \left(C_R, H_R\right) \Delta C_R + u_{R,H} \left(C_R, H_R\right) \theta_R \Delta C_R\end{aligned}$$

and

$$\begin{aligned}\Delta u_i \left(C_i, H_i\right) &= u_{i,C} \left(C_i, H_i\right) \Delta C_i + u_{i,H} \left(C_i, H_i\right) \Delta H_i \\ &= u_{i,C} \left(C_i, H_i\right) \Delta C_i + u_{i,H} \left(C_i, H_i\right) \theta_i \Delta C_i.\end{aligned}$$

By $N \Delta u_R \left(C_R, H_R\right) = \sum_i \Delta u_i \left(C_i, H_i\right)$, we have

$$u_{R,C} \left(C_R, H_R\right) \Delta C_R + u_{R,H} \left(C_R, H_R\right) \theta_R \Delta C_R = \frac{1}{N} \sum_i \left[u_{i,C} \left(C_i, H_i\right) \Delta C_i + u_{i,H} \left(C_i, H_i\right) \theta_i \Delta C_i\right].$$

Because $\Delta C_i = \Delta C_j = \Delta C_R$ and $u_{R,C} \left(C_R, H_R\right) = u_{i,C} \left(C_i, H_i\right)$ (comparability) $\forall i, j$,

$$\begin{aligned}\frac{u_{R,H} \left(C_R, H_R\right)}{u_{R,C} \left(C_R, H_R\right)} \theta_R &= \frac{1}{N} \sum_i \frac{u_{i,H} \left(C_i, H_i\right)}{u_{i,C} \left(C_i, H_i\right)} \theta_i \\ &= \frac{1}{N} \sum_i \frac{u_{i,H} \left(C_i, H_i\right)}{u_{i,C} \left(C_i, H_i\right)} \cdot \frac{1}{N} \sum_i \theta_i\end{aligned}$$

where the second equality holds because of the independence between slope of indif-

[56] Before reaching a steady state.

ference curve and habit depreciation rate. With $u_{R,H}\left(C_R,H_R\right)=\frac{1}{N}\sum_i u_{i,H}\left(C_i,H_i\right)$ (see section A.2) and $u_{R,C}\left(C_R,H_R\right)=u_{i,C}\left(C_i,H_i\right)\ \forall i$, I arrive at

$$\theta_R=\frac{1}{N}\sum_i\theta_i.$$

That is, the representative agent's habit depreciation rate is the average of the individuals' habit depreciation rates.

A.2 Aggregation of Ratios of Utility Derivatives

First, I derive the relationships between utility derivatives of the representative agent and the heterogeneous agents at the baseline steady state ($\bar{C}_R=\bar{H}_R=\bar{C}_i=\bar{H}_i=\bar{C}=\bar{H}\ \forall i$).

Because $Nu_R\left(C_R,H_R\right)=\sum_i u_i\left(C_i,H_i\right)$ for $C_R=C_i=A$ and $H_R=H_i=B\ \forall i$ and $\forall A,B$ in the domains of the utility functions, utility derivatives of the representative agent is the average of the utility derivatives of the individuals:

$$u_{R,X}\left(C_R,H_R\right)=\frac{1}{N}\sum_i u_{i,X}\left(C_i,H_i\right)\xrightarrow{p}\mathbb{E}\left(u_{i,X}\left(C_i,H_i\right)\right)$$

where X denotes the variable and order of the utility derivatives (e.g., C, H, CC, CH, HH).

Next, I derive the relationships between ratios of utility derivatives of the representative agent and the heterogeneous agents at the baseline steady state.

1. Under the normalization of $u_{R,C}\left(\bar{C},\bar{H}\right)=u_{i,C}\left(\bar{C},\bar{H}\right)$, the distribution of $-\frac{u_{i,H}\left(\bar{C},\bar{H}\right)}{u_{i,C}\left(\bar{C},\bar{H}\right)}$ is simply the distribution of $-u_{i,H}\left(\bar{C},\bar{H}\right)$ scaled by $\bar{u}$, the level of the marginal utility of consumption at the baseline ($\bar{u}=u_{R,C}\left(\bar{C},\bar{H}\right)=u_{i,C}\left(\bar{C},\bar{H}\right)$). Thus, $\mathbb{E}\left(u_{i,H}\right)$ can be calculated from

   $$\mathbb{E}\left(u_{i,H}\right)=-\mathbb{E}\left(-\frac{u_{i,H}}{u_{i,C}}\cdot u_{i,C}\right)=-\mathbb{E}\left(-\frac{u_{i,H}}{u_{i,C}}\right)\cdot u_{i,C}=-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\cdot\bar{u},$$

   where $\mu_{-\frac{u_{i,H}}{u_{i,C}}}$ is the mean of the preference parameter $-\frac{u_{i,H}}{u_{i,C}}$ across individuals.

2. The distribution of $u_{i,HH}$ is the distribution of $\frac{Hu_{i,HH}}{u_{i,H}}$ multiplied by $u_{i,H}$ and

scaled by the inverse of H. Because the parameters are independent,

$$\begin{aligned}\mathbb{E}\left(u_{i,HH}\right) &= \mathbb{E}\left(\frac{Hu_{i,HH}}{u_{i,H}} \cdot u_{i,H} \cdot \frac{1}{H}\right) \\ &= \mathbb{E}\left(\frac{Hu_{i,HH}}{u_{i,H}}\right) \mathbb{E}\left(u_{i,H}\right) \frac{1}{H} \\ &= \mu_{\frac{Hu_{i,HH}}{u_{i,H}}} \cdot \left(-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\right) \cdot \bar{u} \cdot \frac{1}{H}.\end{aligned}$$

3. Similarly,

$$\begin{aligned}\mathbb{E}\left(u_{i,CH}\right) &= \mathbb{E}\left(\frac{u_{i,CH}}{u_{i,HH}} \cdot u_{i,HH}\right) \\ &= \mathbb{E}\left(\frac{u_{i,CH}}{u_{i,HH}}\right) \mathbb{E}\left(u_{i,HH}\right) \\ &= \mu_{\frac{u_{i,CH}}{u_{i,HH}}} \cdot \mu_{\frac{Hu_{i,HH}}{u_{i,H}}} \cdot \left(-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\right) \cdot \bar{u} \cdot \frac{1}{H},\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left(u_{i,CC}\right) &= \mathbb{E}\left(\frac{u_{i,CC}}{u_{i,HH}} \cdot u_{i,HH}\right) \\ &= \mathbb{E}\left(\frac{u_{i,CC}}{u_{i,HH}}\right) \mathbb{E}\left(u_{i,HH}\right) \\ &= \mu_{\frac{u_{i,CC}}{u_{i,HH}}} \cdot \mu_{\frac{Hu_{i,HH}}{u_{i,H}}} \cdot \left(-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\right) \cdot \bar{u} \cdot \frac{1}{H},\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\left(u_{i,C_{others}}\right) &= \mathbb{E}\left(\frac{u_{i,C_{others}}}{u_{i,H}} \cdot u_{i,H}\right) \\ &= \mathbb{E}\left(\frac{u_{i,C_{others}}}{u_{i,H}}\right) \cdot \mathbb{E}\left(u_{i,H}\right) \\ &= \mu_{\frac{u_{i,C_{others}}}{u_{i,H}}} \cdot \left(-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\right) \cdot \bar{u}.\end{aligned}$$

4. With these I then calculate

$$-\frac{u_{R,H}}{u_{R,C}} = -\frac{\mathbb{E}\left(u_{i,H}\right)}{\mathbb{E}\left(u_{i,C}\right)} = -\frac{\mathbb{E}\left(u_{i,H}\right)}{\bar{u}} = \mu_{-\frac{u_{i,H}}{u_{i,C}}},$$

$$\frac{Hu_{R,HH}}{u_{R,H}} = \frac{H\mathbb{E}\left(u_{i,HH}\right)}{\mathbb{E}\left(u_{i,H}\right)} = \frac{H\mu_{\frac{Hu_{i,HH}}{u_{i,H}}}\left(-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\right)\bar{u}\frac{1}{H}}{-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\bar{u}} = \mu_{\frac{Hu_{i,HH}}{u_{i,H}}},$$

$$\frac{u_{R,CH}}{u_{R,HH}} = \frac{\mathbb{E}\left(u_{i,CH}\right)}{\mathbb{E}\left(u_{i,HH}\right)} = \frac{\mu_{\frac{u_{i,CH}}{u_{i,HH}}}\mu_{\frac{Hu_{i,HH}}{u_{i,H}}}\left(-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\right)\bar{u}\frac{1}{H}}{\mu_{\frac{Hu_{i,HH}}{u_{i,H}}}\left(-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\right)\bar{u}\frac{1}{H}} = \mu_{\frac{u_{i,CH}}{u_{i,HH}}},$$

$$\frac{u_{R,CC}}{u_{R,HH}} = \frac{\mathbb{E}\left(u_{i,CC}\right)}{\mathbb{E}\left(u_{i,HH}\right)} = \frac{\mu_{\frac{u_{i,CC}}{u_{i,HH}}}\mu_{\frac{Hu_{i,HH}}{u_{i,H}}}\left(-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\right)\bar{u}\frac{1}{H}}{\mu_{\frac{Hu_{i,HH}}{u_{i,H}}}\left(-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\right)\bar{u}\frac{1}{H}} = \mu_{\frac{u_{i,CC}}{u_{i,HH}}},$$

and

$$\frac{u_{R,C_{others}}}{u_{R,H}} = \frac{\mathbb{E}\left(u_{i,C_{others}}\right)}{\mathbb{E}\left(u_{i,H}\right)} = \frac{\mu_{\frac{u_{i,C_{others}}}{u_{i,C}}}\cdot\left(-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\right)\cdot\bar{u}}{\left(-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\right)\cdot\bar{u}} = \mu_{\frac{u_{i,C_{others}}}{u_{i,H}}}.$$

In summary, the representative agent's ratios of utility derivatives are averages of individuals' ratios of utility derivatives.

A.3 Aggregation of External Habit Mixture Coefficient

Imagine that everyone increases her or his consumption by the same iota amount so that the representative agent also increases her consumption by this same amount. Looking at the representativeness condition reveals that the changes in utilities satisfy

$$N\left[u_{R,C}\left(C_R,H_R\right)\Delta C_R + u_{R,H}\left(C_R,H_R\right)\Delta H_R\right] = \sum_i \left[u_{i,C}\left(C_i,H_i\right)\Delta C_R + u_{i,H}\left(C_i,H_i\right)\Delta H_i\right].$$

Utilizing the comparability condition to get

$$N\frac{u_{R,H}\left(C_R,H_R\right)}{u_{R,C}\left(C_R,H_R\right)}\Delta H_R = \sum_i \frac{u_{i,H}\left(C_i,H_i\right)}{u_{i,C}\left(C_i,H_i\right)}\Delta H_i.$$

Because $\Delta H_R/\Delta H_i = \theta_R(1-\omega_R)/[\theta_i(1-\omega_i)]$, I have

$$\begin{aligned}\frac{u_{R,H}(C_R,H_R)}{u_{R,C}(C_R,H_R)}\theta_R(1-\omega_R) &= \frac{1}{N}\sum_i \frac{u_{i,H}(C_i,H_i)}{u_{i,C}(C_i,H_i)}\theta_i(1-\omega_i) \\ &= \left(\frac{1}{N}\sum_i \frac{u_{i,H}(C_i,H_i)}{u_{i,C}(C_i,H_i)}\right)\cdot\left(\frac{1}{N}\sum_i \theta_i\right)\cdot\left(\frac{1}{N}\sum_i (1-\omega_i)\right),\end{aligned}$$

where the second equality because of the independence between the preference parameters. Finally, by $\frac{u_{R,H}(C_R,H_R)}{u_{R,C}(C_R,H_R)} = \frac{1}{N}\sum_i \frac{u_{i,H}(C_i,H_i)}{u_{i,C}(C_i,H_i)}$ and $\theta_R = \frac{1}{N}\sum_i \theta_i$,

$$\omega_R = \frac{1}{N}\sum_i \omega_i.$$

B Equivalence of Linear and Nonlinear Habit Evolutions

In this section, I show that the model with the linear habit evolution (Model L below) and models with nonlinear habit evolutions (Model N below) are observationally equivalent (Definition 1) by a (potentially nonlinear) monotonic transformation of the scale on which the habit is measured.[57]

- Model L:

$$\begin{gathered}\mathbb{E}_0 \int_0^\infty e^{-\rho t} u(C,H)\,dt \\ s.t.\quad \dot{H} = \theta(C-H)\end{gathered}$$

- Model N:

$$\begin{gathered}\mathbb{E}_0 \int_0^\infty e^{-\rho t} v(C,\mathcal{H})\,dt \\ s.t.\quad \dot{\mathcal{H}} = f(C,\mathcal{H})\end{gathered}$$

  where f can be a nonlinear function of C and $\mathcal{H}$.

Note that $H_t = h(C_0,H_0,t)$ if $C_t = C_{t+1}$ for all $t \geq 0$ (subscripts index time). Similarly, $\mathcal{H}_t = k(C_0,\mathcal{H}_0,t)$ if $C_t = C_{t+1}$ for all $t \geq 0$. That is, when the consumption

[57] For example, the geometric habit evolution, as in Kozicki and Tinsley (2002), synchronizes with the linear habit evolution of Model L.

path does not change, H_t and $\mathcal{H}_t$ are functions of only time while C_0, H_0, and $\mathcal{H}_0$ are their parameters.

Definition 1. Two models are observationally equivalent if they lead to the same set of optimal choices.

Definition 2. Monotonicities of two functions are entangled with respect to a variable if 1) the two functions share this variable as an argument and 2) when one function is monotonic in the argument, the other function is also monotonic in the argument.

Because H and $\mathcal{H}$ are two measurements of one fundamental - habit, they change at the same time (though in potentially different ways) when habit changes and stop changing when habit stops changing. By Definition 2, their monotonicities[58] are entangled with respect to time.

Proposition 11. *Model L and Model N are observationally equivalent if monotonicities of H and $\mathcal{H}$ are entangled with respect to time.*

Proof. Because H and $\mathcal{H}$ are entangled monotonically with respect to time, without loss of generality, suppose that H and $\mathcal{H}$ are monotonic from period 0 to period T and flat afterward (i.e., remain at constant levels), say at levels $\bar{H}$ and $\bar{\mathcal{H}}$. Suppose also consumption changes at period 0 and stays at that level: $C_t = C_{t+1} \neq C_{-\varepsilon}$ for all $t \geq 0$ and $\varepsilon > 0$. As a result, $H_t = a\left(t|C_0, H_0\right)$ and $\mathcal{H}_t = b\left(t|C_0, \mathcal{H}_0\right)$ where $a\left(\cdot\right)$ and $b\left(\cdot\right)$ are monotonic functions of t for $0 \leq t \leq T$ and flat for $t > T$, where T is the instant when habit reaches its new steady state.

Because

$$\mathcal{H}_t = b\left(t|C_0, H_0\right) = b\left(a^{-1}\left(a\left(t|C_0, H_0\right)\right)|C_0, H_0\right) = b\left(a^{-1}\left(H_t\right)|C_0, H_0\right)$$

for $0 \leq t \leq T$ and

$$\mathcal{H}_t = \frac{\bar{\mathcal{H}}}{\bar{H}} H_t$$

for $t > T$, there always exists an injective function G that maps H_t into $\mathcal{H}_t$:

$$\mathcal{H}_t = G\left(H_t\right) = \begin{cases} b\left(a^{-1}\left(H_t\right)|C_0, H_0\right) & 0 \leq t \leq T \\ \frac{\bar{\mathcal{H}}}{\bar{H}} H_t & t > T \end{cases}$$

. For other patterns of monotonicities of the two functions (e.g. flat to monotonic to flat to monotonic, etc.), the function G can be derived analogously.

[58] That is, H' and $\mathcal{H}'$ can equal to 0 but they will not change sign: i.e., $H' \cdot \mathcal{H}'$ does not change sign around the region that $H' \cdot \mathcal{H}' = 0$.

Because

$$v\left(C\left(t\right),\mathcal{H}\left(t\right)\right)=v\left(C\left(t\right),G\left(H\left(t\right)\right)\right)\equiv u\left(C\left(t\right),H\left(t\right)\right),$$

Model N gives the same utility as Model L for any consumption path that is constant for $t\geq 0$.

When consumption path is not constant over time, the above remains true. To see this, start from the period when consumption is changed for the last time and apply the above logic to get the same utility from the two models starting from that period onwards. Then go back to the period when consumption is changed for the second to last time and apply the above logic. Same utility results again for the two models. Continue this process until the first period of interest.

Because the utilities from the two models are the same, the consumption choices generated from these two models coincide. Suppose not. That is, the two models generate different optimal consumption paths, $\{C_L^*\}\neq\{C_N^*\}$ for at least one instant, where

$$\{C_L^*\}=\arg\max_{\{C\}}\mathbb{E}_0\int_0^\infty e^{-\rho t}u\left(C,H\right)dt\equiv\arg\max_{\{C_L\}}U\left(\{C_L\},H_0\right)$$

and

$$\{C_N^*\}=\arg\max_{\{C\}}\mathbb{E}_0\int_0^\infty e^{-\rho t}v\left(C,\mathcal{H}\right)dt\equiv\arg\max_{\{C_N\}}V\left(\{C_N\},\mathcal{H}_0\right).$$

If $U\left(\{C_L^*\},H_0\right)\neq V\left(\{C_N^*\},\mathcal{H}_0\right)$, at least one of the two consumption paths is not the optimal solution, contradicting that both of them are optimal solutions. If $U\left(\{C_L^*\},H_0\right)=V\left(\{C_N^*\},\mathcal{H}\right)$, then the consumption path $\{C_L^*\}$ will also be a solution to the Model N while $\{C_N^*\}$ also a solution to Model L. Therefore, $\{C_H^*\}$ and $\{C_{\mathcal{H}}^*\}$ are both solutions to the two models. In other words, the two models share the same set of solutions. Thus, by Definition 1, the two models are observationally equivalent. $\square$

Because the monotonicities of H and $\mathcal{H}$ are entangled with respect to time, by Proposition 11, Model L and Model N are observationally equivalent.

C Elicitation of Time Discount Rate

Because the time discount rate is of only indirect interest,[59] it is relegated to here.

To elicit the time discount rate, I increase spending in the next year and the year after next year. The resulting survey question has monthly spending graphs as in

[59] That is, helping the estimation of other preference parameters.

Figure 22.

Proposition 12. *Under exact elicitation, $\rho > -\ln \frac{\Delta e}{\Delta f}$ if the respondent chooses Universe One over Universe Two for a better future experience in the time discount rate question.*[60]

Estimation pins down a value of about 0.15 for the time discount rate (Table 24).

Table 24: Estimate of Time Discount Rate

	MAP	Mean	Median	95% HPDI
ρ	0.15	0.14	0.14	[0.04, 0.25]

D Response Frequency

See Table 25.

Table 25: Response Frequency (Percentage) for Preference Parameters Identifiable to Scale

Question	Wave	Choice					
		1	2	3	4	5	6
Habit Depreciation Rate	1	28	10	16	11	7	28
	2	30	10	14	11	6	30
External Habit Mixture Coefficient	1	46	9	13	8	4	18
	2	46	4	7	16	5	23
$-u_H/u_C$	1	32	3	7	7	12	39
	2	31	4	6	2	20	36
Hu_{HH}/u_H	1	14	4	8	10	3	60
	2	25	2	6	5	1	61
u_{CH}/u_{HH}	1	7	4	11	10	27	41
	2	9	3	10	9	34	36
u_{CC}/u_{HH}	1	24	30	10	9	5	22
	2	24	19	8	10	9	30
$u_{C_{others}}/u_H$	1	26	18	10	8	2	36
	2	24	19	8	6	3	40
ρ	1	34	4	7	7	4	43
	2	39	3	5	5	4	45

[60] $\Delta e = \$2,000$ and $\Delta f = \$2,200$ in the example of Figure 22. See section F.12 of the appendix for proof.

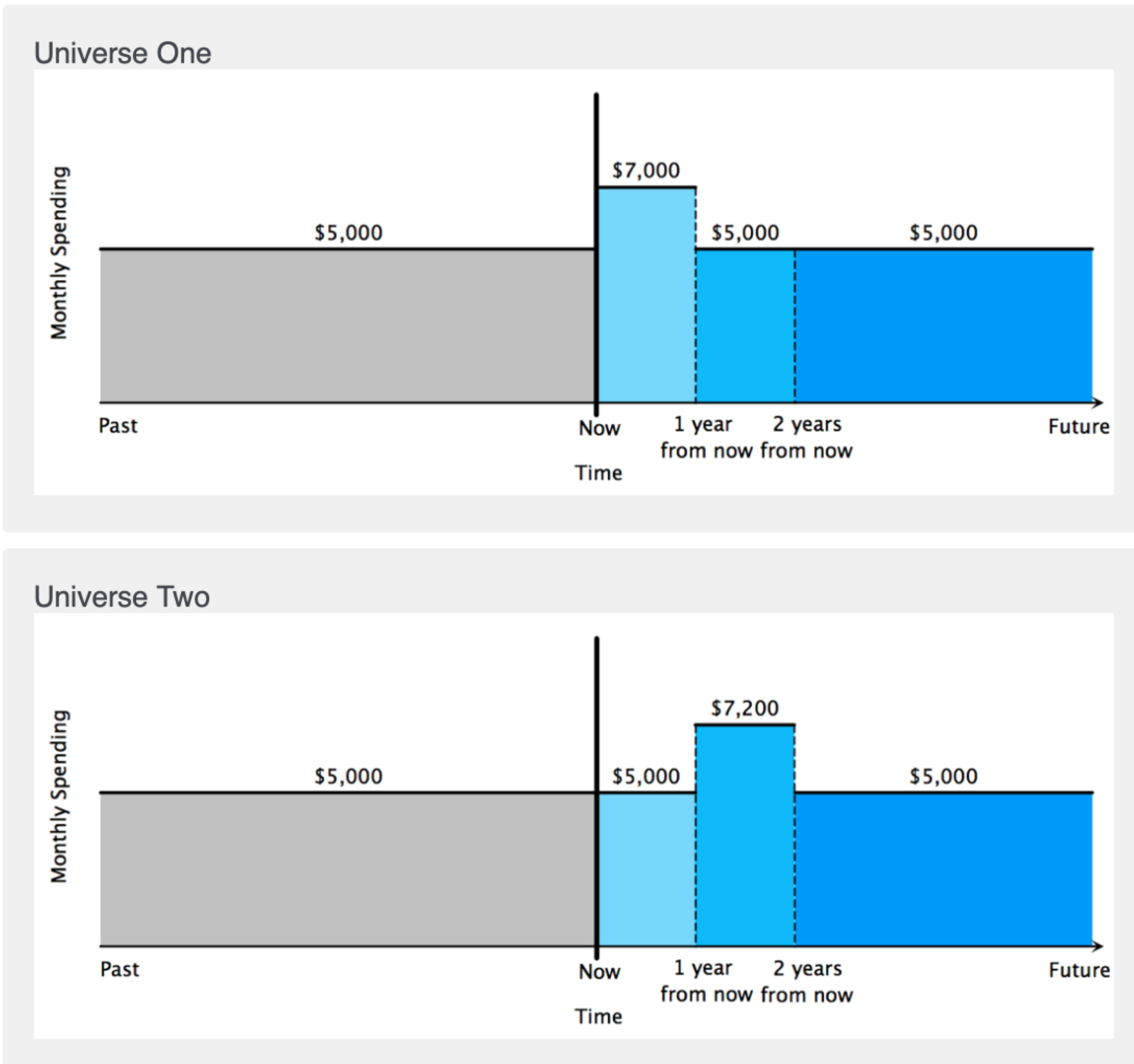


Figure 22: Monthly spending graphs—time discount rate.

E Survey

E.1 Instructions

Instruction 1/2

The survey started with instruction on the hypothetical situation discussed in section 2.2. A set of practice questions followed to test the respondents' understanding of the hypothetical situation:

> With no inflation and prices of everything staying the same, if you can buy 3 bananas with one dollar in the last year, how many bananas can you buy with one dollar in the next year?
>
> - 5
> - 3
> - 1
> - No idea
>
> If you rent the durable goods you consume, select any of the following that you own (that is, not rent):
>
> - Residence
> - Car
> - Furniture
> - I do not own any of the above
> - No idea

Instruction 2/2

After passing the practice questions, the respondents were presented with instruction on reading monthly spending graphs.

> In this survey, you'll compare your experience in several universes that are identical except that your monthly spending differs.
>
> - Monthly spending refers to the total amount of money you spend, rather than earn, in each month.

- You will be asked to find out in which universe you will have (had) a better experience given how much you spend (spent).
- 'Better' means more satisfying.
- You can afford the monthly spending specified in the questions.

The difference of your monthly spending between the universes is detailed in monthly spending graphs, like the one below.

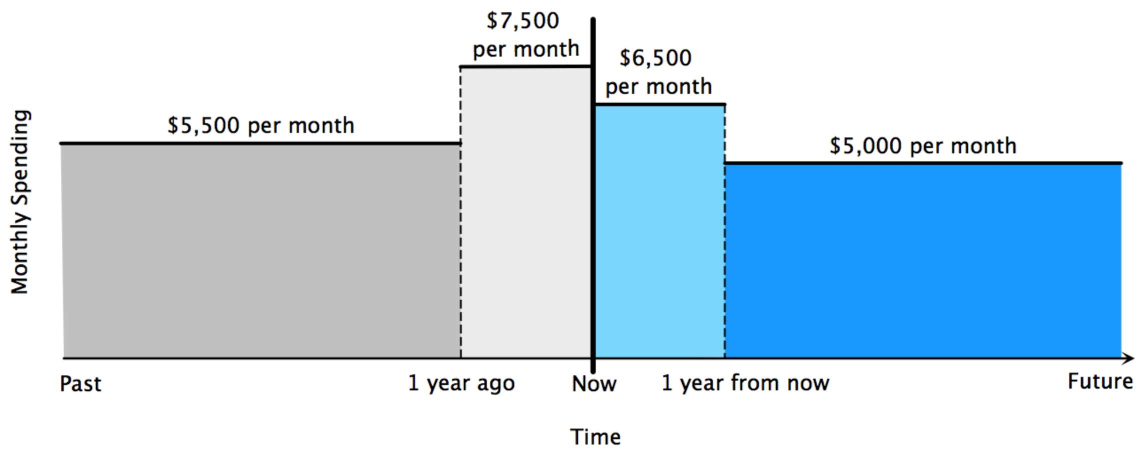


Figure 23: Instruction—monthly spending graph 1.

Now let's learn to read a monthly spending graph.

The first element of a monthly spending graph is the timeline, with past on the left, now in the middle, future on the right. A thick vertical line representing now separates the past from the future.

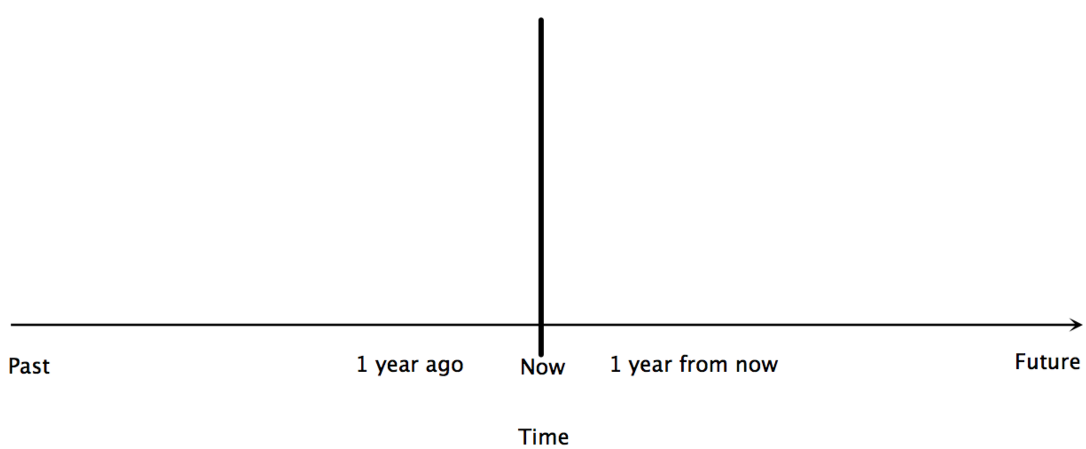


Figure 24: Instruction—timeline.

To fix the idea, the 'Past' means as far back in the past as you can remember and the 'Future' as far in the future as you can imagine. If easier, think of the 'Past' as the past 30 years and the 'Future' as the next 30 years.

The second element of a monthly spending graph is the bars above the timeline.

- The height of bars represents the level of monthly spending (again, not income) in time frames covered by the bars.
- The exact level of monthly spending is labeled on top of the corresponding bar. The words 'per month' are saved for space consideration from now on, but you should always remember that the numbers are per month spending.
- The bars are colored differently to help you distinguish different time frames.

For example, if the following monthly spending graph describes your monthly spending,

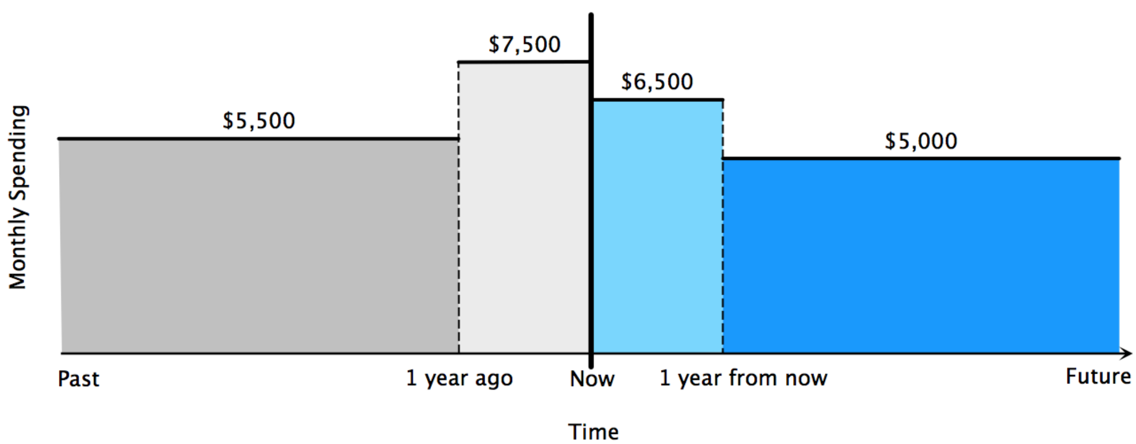


Figure 25: Instruction—monthly spending graph 2.

you spent/spend

- $5,500 per month in the 'past' until '1 year ago';
- $7,500 per month from '1 year ago' until 'now' (or in the 'past year');

- \$6,500 per month from 'now' to '1 year from now' (or in the 'next year');
- \$5,000 per month from '1 year from now' onward.

To highlight the difference of monthly spending, the time frames as in the above example are sometimes collapsed into three or two timeframes. For instance, if in Universe One your monthly spending graph is

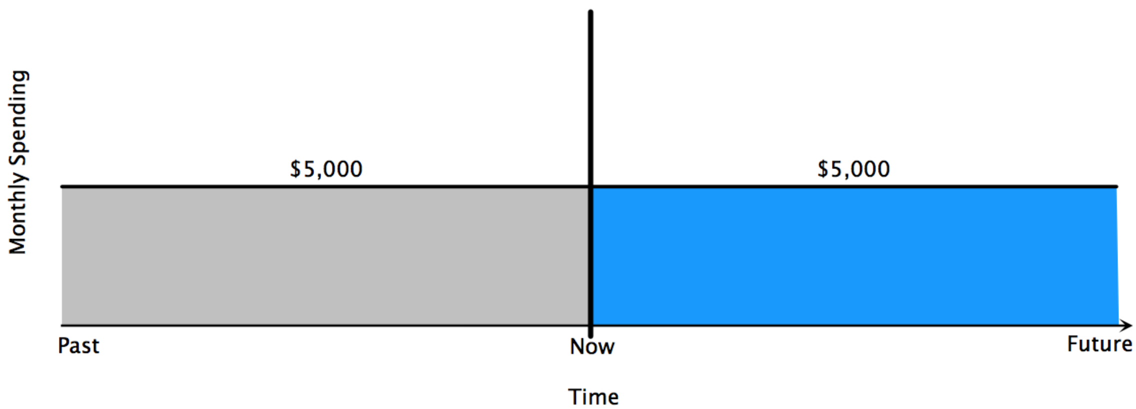


Figure 26: Instruction—monthly spending graph 3.

while in Universe Two your monthly spending is

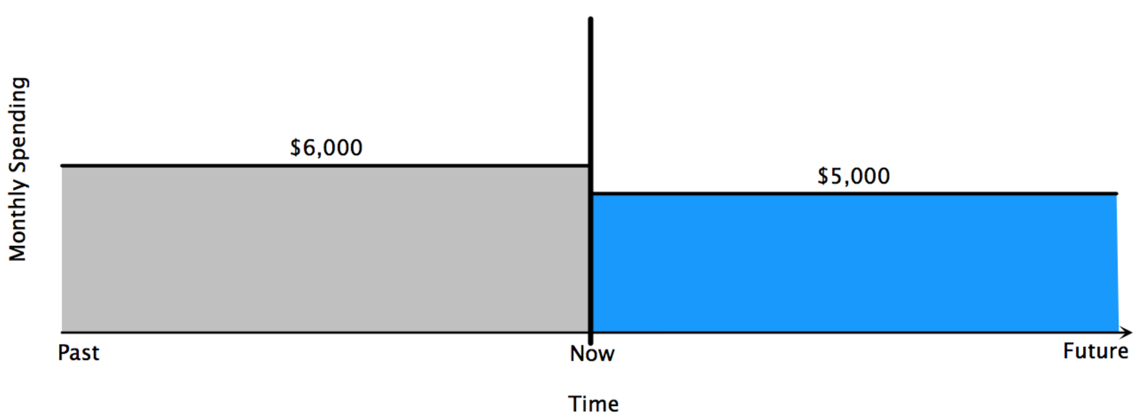


Figure 27: Instruction—monthly spending graph 4.

then the difference and the similarity of your monthly spending in the two universes are that

- in Universe Two you spent \$1,000 more per month in the 'past' than you did in Universe One where you spent \$5,000 per month in the 'past'; and
- in both universes, you will spend \$5,000 per month from 'now' onward.

Then, the respondents were presented with a set of practice questions testing their understanding of the above instruction.

Imagine that your monthly spending is detailed in the following monthly spending graph

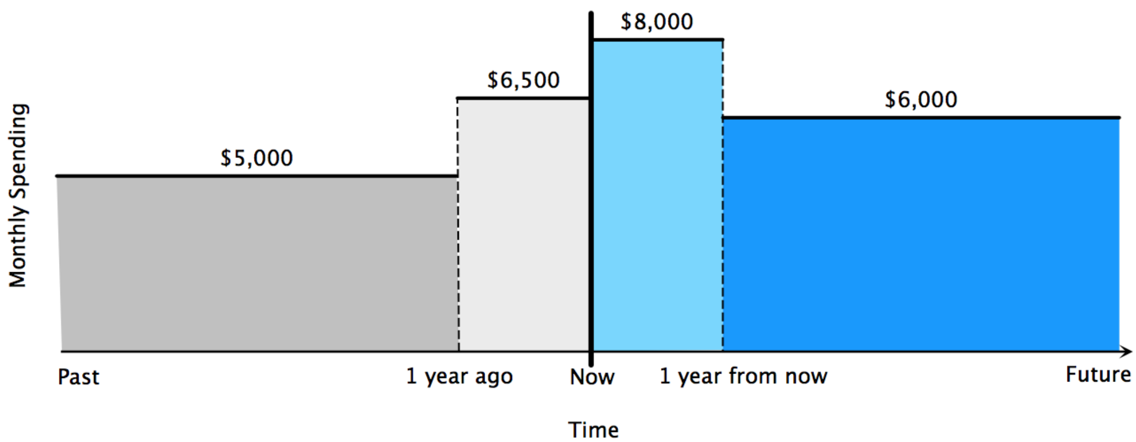


Figure 28: Instruction—monthly spending graph 5.

How much will you spend per month in the next year?

- \$5,000
- \$6,000
- \$6,500
- \$8,000

How much did you spend per month from 'as far back as you can remember in the past' until '1 year ago'?

- \$5,000
- \$6,000
- \$6,500

- $8,000

- Imagine that your monthly spending in Universe One and Universe Two are detailed in the following monthly spending graphs.

- [The graphs are in the choices below. You can directly click the graph to give your answer.]

  In which universe did you spend more per month in the 'past year'?

  - Universe One

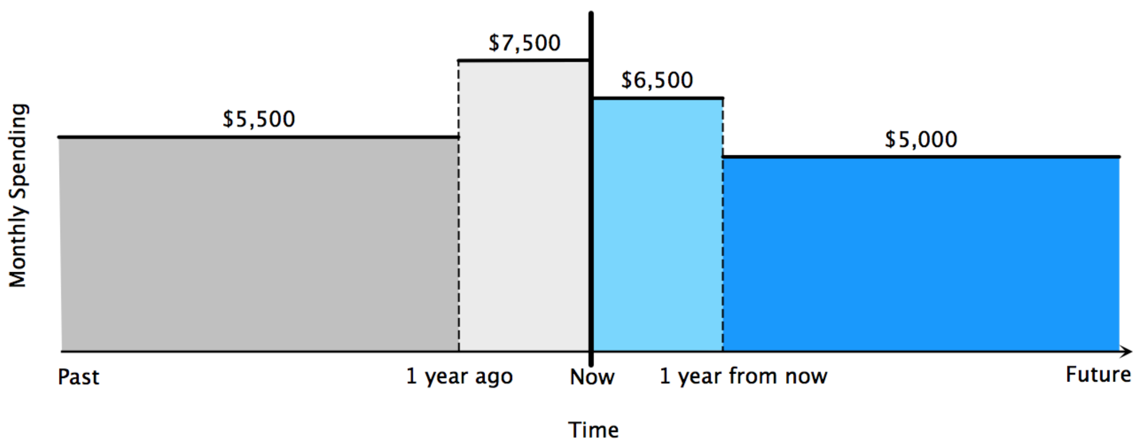


Figure 29: Instruction—monthly spending graph 2.

  - Universe Two

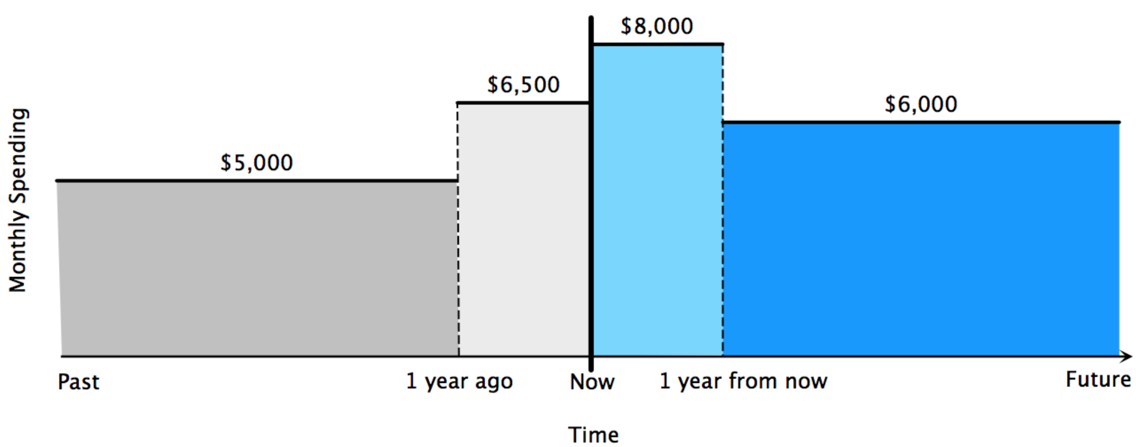


Figure 30: Instruction—monthly spending graph 5.

In which universe will you spend more per month from '1 year from now' onward?

- Universe One

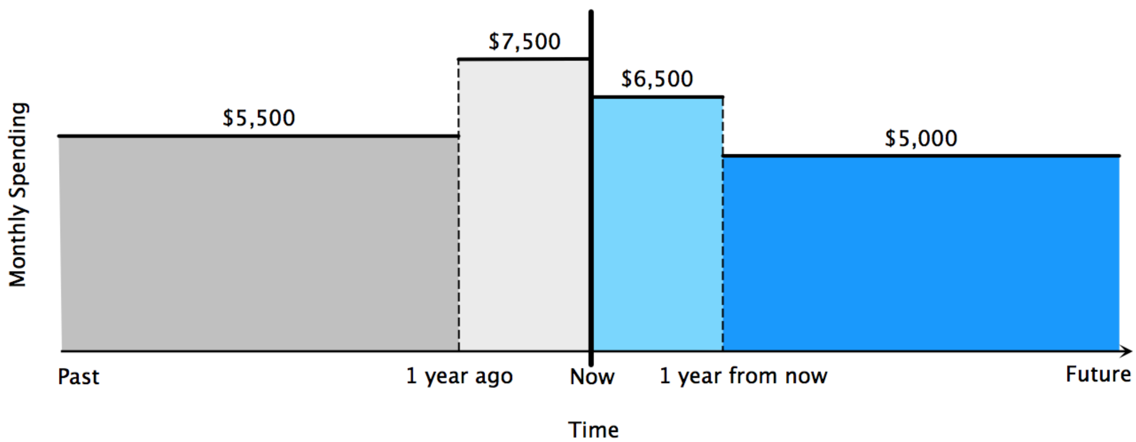


Figure 31: Instruction—monthly spending graph 2.

- Universe Two

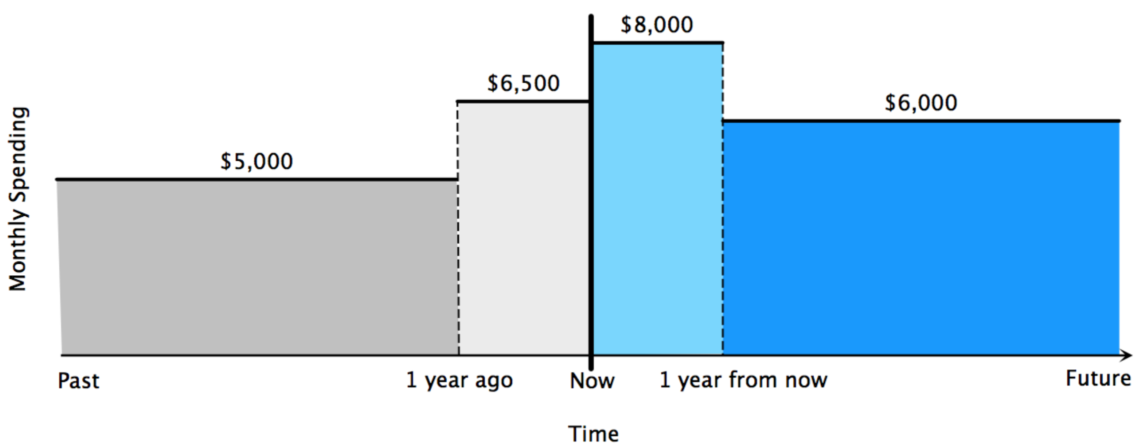


Figure 32: Instruction—monthly spending graph 5.

In the graphs of last question, how much more did you spend in Universe One than in Universe Two from 'as far back in the past as you can remember' until '1 year ago'?

- $0
- $500

- \$1,000
- \$5,500

Final Instruction

A final instruction vaccinated the respondents against the seeming repetitiveness of follow-up questions and encouraged effort with attention checks and a potential bonus reward. An opportunity to review the instructions was also presented.

> We designed the survey to learn as much as possible from your answers. To increase the power of the study result, each question is normally followed by additional questions that vary slightly from previous questions. Although the survey may look repetitive, please pay careful attention and answer each question the best you can.
>
> Implicit and explicit attention checks are integrated into this survey. Responses show signs of inattentiveness will be rejected. A small lottery (\$1) will be randomly paid as a bonus to workers who show excellence in the responses. The chance to get this lottery reward is 1 in 100. Your normal HIT payment won't be affected by this lottery.
>
> If you would like to view these instructions again before beginning the survey, please check the following box.
>
> ☐ View Instructions Again

E.2 Core Modules

The survey has nine core modules, each corresponding to one or two preference parameters of interest.

The flow of each core module is

1. ask the respondent to choose which universe brings her a better past experience (Figure 2);

2. ask the respondent to choose which universe brings her a better future experience (Figure 3);

3. present the respondent with a set of follow-up monthly spending graphs that varies slightly from the original set. The total number of sets of follow-up questions ranges from 1 to 2 (Figure 7).

The main element of the survey questions in the core modules is the monthly spending graphs, which are in the main text of the paper. To save space, they are not repeated here. In Table 26, I present the changes between follow-up questions and initial questions in terms of the quantities that are necessary to elicit the preference parameters in Propositions 1, 4 to10, and12.

E.3 End-of-Survey Check of Understanding of Hypothetical Situation

At the end of the survey, I checked the respondents' understanding of the hypothetical situation again using the following questions, which serve as implicit attention checks.

> Under the hypothetical situation of this survey, if you can buy 3 bananas with one dollar in the last year, how many bananas can you buy with one dollar in the next year?
>
> - 5
> - 3
> - 1
> - No idea
>
> Under the hypothetical situation of this survey, select any of the following that you own (that is, not rent):
>
> - Residence
> - Car
> - Furniture
> - I do not own any of the above
> - No idea
>
> Under the hypothetical situation of this survey, do things you want change over time?
>
> - Yes
> - Maybe
> - No

Table 26: Quantities in Monthly Spending Graphs

		If choosing U2 in initial and 1st follow-up questions (2nd follow-up question)	If choosing U2 in initial question (1st follow-up question)	Initial question	If choosing U1 in initial question (1st follow-up question)	If choosing U1 in initial and 1st follow-up questions (2nd follow-up question)
Habit Depreciation	ΔC_{U1}	400	1200	2000	2000	2000
Speed	ΔC_{U2}	4000	4000	4000	2800	2200
External Habit	ΔC	4500	1200	500	500	500
Mixture Coefficient	ΔC_{others}	500	500	500	1200	4500
$-u_H/u_C$	Δe	2000	2000	2000	2000	2000
	Δf	20	80	200	400	1000
Hu_{HH}/u_H	Δe	500	500	500	500	500
	Δf	540	600	650	700	800
u_{CH}/u_{HH}	Δe	2000	1600	1000	600	100
	Δf	200	200	200	200	200
u_{CC}/u_{HH}	Δe	500	1500	2200	3000	3500
	Δf	500	500	500	500	500
$u_{C_{others}}/u_H$	ΔC^{U1}_{others}	3000	600	300	150	100
	ΔC^{U2}_{others}	3000	3000	3000	3000	3000
Time Discount	Δe	2000	2000	2000	2000	2000
Rate	Δf	3300	2500	2200	2100	2040

Notes: U1 and U2 denote Universe One and Universe Two of the monthly spending graphs, respectively. Choosing U1 in the initial question and then U2 in the 1st follow-up question or choosing U2 in the initial question and then U1 in the 1st follow-up question ends the module at the end of the 1st follow-up question (see Figure 7). All amounts are in U.S. dollar.

Under the hypothetical situation of this survey, do things not mentioned in the questions change?

- Yes
- Maybe
- No

Under the hypothetical situation of this survey, how much do people not mentioned in questions always spend per month?

- $4,000
- $5,000
- $6,500
- $8,000
- No idea

F Proofs

F.1 Lemmas

I prove three lemmas that will be used in proving propositions that have not been proved thus far.

Lemma 1. *For $a, b, c \in \mathbb{R}$, if $a\,(a+b) > 0$, then $a\,(a+cb) > 0$ as long as $0 \leq c \leq 1$.*

Proof. $a\,(a+b) > 0$ is equivalent to $a+b < 0$ if $a < 0$ and $a+b > 0$ if $a > 0$.

Suppose $a < 0$ and $a+b < 0$. Note that $a+cb = a+b+(c-1)\,b$. If $b \geq 0$, then $(c-1)\,b \leq 0$ and thus $a+cb \leq a+b < 0$. If $b < 0$, then by $a < 0$ and $c \geq 0$, $a+cb < 0$. Therefore $a\,(a+cb) > 0$.

Suppose $a > 0$ and $a+b > 0$. If $b \leq 0$, then $(c-1)\,b \geq 0$ and therefore $a+cb \geq a+b > 0$. If $b > 0$, then by $a > 0$ and $c \geq 0$, $a+cb > 0$. Therefore $a\,(a+cb) > 0$. $\square$

Lemma 2. *For $\Delta e, \Delta f, M, \theta \in \mathbb{R}^+$, if $M - \Delta f - \Delta e \geq 0$, $\frac{\sum_{s=0}^{n} (\Delta e)^s (\Delta f)^{n-s}}{M^n}$ is non-increasing in $n \in \mathbb{N}^+$.*

Proof. $\frac{\sum_{s=0}^{n}(\Delta e)^s(\Delta f)^{n-s}}{M^n}$ is non-increasing in $n \in \mathbb{N}^+$ if $\frac{\sum_{s=0}^{n}(\Delta e)^s(\Delta f)^{n-s}}{M^n} - \frac{\sum_{s=0}^{n+1}(\Delta e)^s(\Delta f)^{n+1-s}}{M^{n+1}} \geq 0 \ \forall n \in \mathbb{N}^+$.

Now,

$$\begin{aligned}
&\frac{\sum_{s=0}^{n}\left(\Delta e\right)^s\left(\Delta f\right)^{n-s}}{M^n} - \frac{\sum_{s=0}^{n+1}\left(\Delta e\right)^s\left(\Delta f\right)^{n+1-s}}{M^{n+1}} \\
&= \frac{\sum_{s=0}^{n}\left(\Delta e\right)^s\left(\Delta f\right)^{n-s}}{M^n} - \frac{\sum_{s=0}^{n}\left(\Delta e\right)^s\left(\Delta f\right)^{n-s} + \frac{(\Delta e)^{n+1}}{\Delta f}}{M^n}\frac{\Delta f}{M} \\
&= \frac{\sum_{s=0}^{n}\left(\Delta e\right)^s\left(\Delta f\right)^{n-s}}{M^n}\left(1 - \frac{\Delta f}{M}\right) - \frac{\left(\Delta e\right)^{n+1}}{M^{n+1}} \\
&= \frac{1}{M^{n+1}}\left[\sum_{s=0}^{n}\left(\Delta e\right)^s\left(\Delta f\right)^{n-s}\left(M - \Delta f\right) - \left(\Delta e\right)^{n+1}\right] \\
&= \frac{\left(\Delta e\right)^n}{M^{n+1}}\left[\sum_{s=0}^{n}\left(\frac{\Delta f}{\Delta e}\right)^{n-s}\left(M - \Delta f\right) - \Delta e\right] \\
&= \frac{\left(\Delta e\right)^n\left(M - \Delta f\right)}{M^{n+1}}\left[\sum_{s=0}^{n}\left(\frac{\Delta f}{\Delta e}\right)^{n-s} - \frac{\Delta e}{M - \Delta f}\right] \\
&= \frac{\left(\Delta e\right)^n\left(M - \Delta f\right)}{M^{n+1}}\left[\sum_{s=0}^{n-1}\left(\frac{\Delta f}{\Delta e}\right)^{n-s} + 1 - \frac{\Delta e}{M - \Delta f}\right] \\
&= \frac{\left(\Delta e\right)^n\left(M - \Delta f\right)}{M^{n+1}}\left[\sum_{s=0}^{n-1}\left(\frac{\Delta f}{\Delta e}\right)^{n-s} + \frac{M - \Delta f - \Delta e}{M - \Delta f}\right] \\
&\geq \frac{\left(\Delta e\right)^n\left(M - \Delta f\right)}{M^{n+1}}\left[\sum_{s=0}^{n-1}\left(\frac{\Delta f}{\Delta e}\right)^{n-s}\right] \\
&\geq 0,
\end{aligned}$$

where the first inequality holds because $M - \Delta f - \Delta e \geq 0$ and $\Delta e > 0$. □

Lemma 3. *For $\Delta e, \Delta f, M, \theta \in \mathbb{R}^+$, if $M - \Delta f - \Delta e \geq 0$, $u_H < 0$, and utility function is analytic, then*

$$u_H e^{-\theta t} + \frac{1}{2}u_{HH}\left(\Delta f + \Delta e\right)\left(e^{-\theta t}\right)^2 + \cdots + \frac{1}{n!}\frac{\partial^n u}{\partial H^n}\left(\sum_{s=0}^{n-1}\left(\Delta e\right)^s\left(\Delta f\right)^{n-s}\right)\left(e^{-\theta t}\right)^n + \cdots < 0.$$

Proof. By $u_H < 0$ and analyticity of the utility function,

$$u\left(C, H+M\right) - u\left(C, H\right) = u_H M + \frac{1}{2}u_{HH}M^2 + \cdots + \frac{1}{n!}\frac{\partial^n u}{\partial H^n}M^n + \cdots < 0$$

for $M > 0$. Because $0 \leq \frac{(\Delta f + \Delta e)}{M}e^{-\theta t} \leq 1$ for $t \in \mathbb{N}^+$, by Lemma 1

$$u_H M + \frac{\Delta f + \Delta e}{M}e^{-\theta t}\left[\frac{1}{2}u_{HH}M^2 + \cdots + \frac{1}{n!}\frac{\partial^n u}{\partial H^n}M^n + \cdots\right] < 0.$$

By Lemma 2,

$$\frac{\frac{\sum_{s=0}^{2}(\Delta e)^s(\Delta f)^{2-s}}{M^2}\left(e^{-\theta t}\right)^2}{\frac{(\Delta f + \Delta e)}{M}e^{-\theta t}} = \frac{\frac{\sum_{s=0}^{2}(\Delta e)^s(\Delta f)^{2-s}}{M^2}}{\frac{(\Delta f + \Delta e)}{M}}e^{-\theta t} \leq 1.$$

Apply Lemma 1 again to get

$$\begin{aligned}
&u_H M + \frac{1}{2}u_{HH}M^2\frac{\Delta f + \Delta e}{M}e^{-\theta t} \\
&+ \frac{\frac{\sum_{s=0}^{2}(\Delta e)^s(\Delta f)^{2-s}}{M^2}}{\frac{(\Delta f + \Delta e)}{M}}e^{-\theta t}\frac{(\Delta f + \Delta e)}{M}e^{-\theta t}\left[\frac{1}{3!}u_{HHH}M^3 + \cdots + \frac{1}{n!}\frac{\partial^n u}{\partial H^n}M^n + \cdots\right] \\
=&u_H M + \frac{1}{2}u_{HH}M^2\frac{\Delta f + \Delta e}{M}e^{-\theta t} \\
&+ \frac{\sum_{s=0}^{2}\left(\Delta e\right)^s\left(\Delta f\right)^{2-s}}{M^2}\left(e^{-\theta t}\right)^2\left[\frac{1}{3!}u_{HHH}M^3 + \cdots + \frac{1}{n!}\frac{\partial^n u}{\partial H^n}M^n + \cdots\right] \\
<&0.
\end{aligned}$$

Repeating this process for all other $n \in \mathbb{N}^+$ to get

$$u_H e^{-\theta t} + \frac{1}{2}u_{HH}\left(e^{-\theta t}\right)^2\left(\Delta f + \Delta e\right) + \cdots + \frac{1}{n!}\frac{\partial^n u}{\partial H^n}\left(\sum_{s=0}^{n-1}\left(\Delta e\right)^s\left(\Delta f\right)^{n-s}\right)\left(e^{-\theta t}\right)^n + \cdots < 0.$$

□

F.2 Proof of Proposition 1

Proof. That θ is habit depreciation rate implies $\theta \in \mathbb{R}^+$. Taking $M = \$5000$, then $M - \Delta C_{U1} - \left(1 - e^{-\theta}\right)\Delta C_{U2} > 0$ in all the questions for habit depreciation rate.[61]

[61] See Table 26 for all the values of ΔC_{U1} and ΔC_{U2} used in survey.

A respondent preferring Universe One for a better future experience implies

$$\begin{aligned}
&\int_0^\infty e^{-\rho t}\left[u_H e^{-\theta t}\Delta C_{U1} + \frac{1}{2}u_{HH}\left(e^{-\theta t}\Delta C_{U1}\right)^2 + \cdots\right]dt \\
&-\int_0^\infty e^{-\rho t}\left[u_H e^{-\theta t}\left(1-e^{-\theta}\right)\Delta C_{U2} + \frac{1}{2}u_{HH}\left(e^{-\theta t}\left(1-e^{-\theta}\right)\Delta C_{U2}\right)^2 + \cdots\right]dt \\
=&\int_0^\infty e^{-\rho t}\left\{u_H e^{-\theta t}\left[\Delta C_{U1} - \left(1-e^{-\theta}\right)\Delta C_{U2}\right]\right. \\
&\left. + \frac{1}{2}u_{HH}\left(e^{-\theta t}\right)^2\left[\left(\Delta C_{U1}\right)^2 - \left(\left(1-e^{-\theta}\right)\Delta C_{U2}\right)^2\right] + \cdots\right\}dt \\
=&\left[\Delta C_{U1} - \left(1-e^{-\theta}\right)\Delta C_{U2}\right]\int_0^\infty e^{-\rho t}\left\{u_H e^{-\theta t}\right. \\
&\left. + \frac{1}{2}u_{HH}\left(e^{-\theta t}\right)^2\left[\Delta C_{U1} + \left(1-e^{-\theta}\right)\Delta C_{U2}\right] + \cdots\right\}dt \\
>&0,
\end{aligned}$$

which, by Lemma 3, implies

$$\Delta C_{U1} - \left(1-e^{-\theta}\right)\Delta C_{U2} < 0$$

or

$$\theta > -\ln\left(1 - \frac{\Delta C_{U1}}{\Delta C_{U2}}\right).$$

□

F.3 Proof of Proposition 2

Proof.

$$\frac{u_{CH}}{u_{HH}}\frac{u_H}{u_C} = \frac{-\gamma v''}{\gamma^2 v''}\frac{-\gamma v'}{v'} = 1,$$

and

$$\frac{u_{CH}}{u_{CC}}\frac{u_C}{u_H} = \frac{-\gamma v''}{v''}\frac{v'}{-\gamma v'} = 1.$$

□

F.4 Proof of Proposition 3

Proof.

$$\frac{Hu_H u_{CH}}{u_C\left(u_H+Hu_{HH}\right)}=\frac{H\left(-\gamma\frac{C}{H^{\gamma+1}}v'\right)\left(-\gamma\frac{1}{H^{\gamma+1}}v'-\gamma\frac{C}{H^{2\gamma+1}}v''\right)}{\frac{1}{H^{\gamma}}v'\left\{-\gamma\frac{C}{H^{\gamma+1}}v'+H\gamma\frac{C}{H^{2}}\left[\left(\gamma+1\right)\frac{1}{H^{\gamma}}v'+\gamma\frac{C}{H^{2\gamma}}v''\right]\right\}}$$
$$=1,$$

and

$$\frac{Cu_C u_{CH}}{u_H\left(u_C+Cu_{CC}\right)}=\frac{C\frac{1}{H^{\gamma}}v'\left(-\gamma\frac{1}{H^{\gamma+1}}v'-\gamma\frac{C}{H^{2\gamma+1}}v''\right)}{\left(-\gamma\frac{C}{H^{\gamma+1}}v'\right)\left(\frac{1}{H^{\gamma}}v'+C\frac{1}{H^{2\gamma}}v''\right)}$$
$$=1.$$

□

F.5 Proof of Proposition 4

Proof. Under second-order elicitation, a respondent preferring Universe One for a better future experience implies

$$\begin{aligned}
&\frac{1}{\rho}\left\{u_C\Delta f+\frac{1}{\rho+\theta}u_H\left(\rho\Delta e+\theta\Delta f\right)+\frac{1}{2}\left[u_{CC}\left(\Delta f\right)^2+2\frac{1}{\rho+\theta}u_{CH}\Delta f\left(\rho\Delta e+\theta\Delta f\right)\right.\right.\\
&\left.\left.+\frac{1}{\left(\rho+\theta\right)\left(\rho+2\theta\right)}u_{HH}\left(\rho\left(\rho+\theta\right)\left(\Delta e\right)^2+2\rho\theta\Delta e\Delta f+2\theta^2\left(\Delta f\right)^2\right)^2\right]\right\}\\
&-\frac{1}{\rho}\left\{u_C\left(-\Delta f\right)+\frac{1}{\rho+\theta}u_H\left(\rho\left(-\Delta e\right)+\theta\left(-\Delta f\right)\right)+\frac{1}{2}\left[u_{CC}\left(\Delta f\right)^2\right.\right.\\
&+2\frac{1}{\rho+\theta}u_{CH}\Delta f\left(\rho\Delta e+\theta\Delta f\right)+\frac{1}{\left(\rho+\theta\right)\left(\rho+2\theta\right)}u_{HH}\left(\rho\left(\rho+\theta\right)\left(\Delta e\right)^2\right.\\
&\left.\left.\left.+2\rho\theta\Delta e\Delta f+2\theta^2\left(\Delta f\right)^2\right)^2\right]\right\}\\
&=\frac{2}{\rho}\left[u_C\Delta f+\frac{1}{\rho+\theta}u_H\left(\rho\Delta e+\theta\Delta f\right)\right]\\
&>0,
\end{aligned}$$

which, by $u_C>0$, implies

$$-\frac{u_H}{u_C}<\frac{\left(\rho+\theta\right)\Delta f}{\rho\Delta e+\theta\Delta f}.$$

□

F.6 Proof of Proposition 5

Proof. Under second-order elicitation, a respondent preferring Universe One for a better future experience implies

$$\begin{aligned}
&\frac{1}{2}\left[\frac{1}{\rho+\theta}u_H\left(-\Delta f\right)+\frac{1}{2}\frac{1}{\rho+2\theta}u_{HH}\left(\Delta f\right)^2\right]+\frac{1}{2}\left[\frac{1}{\rho+\theta}u_H\Delta e+\frac{1}{2}\frac{1}{\rho+2\theta}u_{HH}\left(\Delta e\right)^2\right]\\
&=\frac{1}{2}\left\{\frac{1}{\rho+\theta}u_H\left(\Delta e-\Delta f\right)+\frac{1}{2}\frac{1}{\rho+2\theta}u_{HH}\left[\left(\Delta f\right)^2+\left(\Delta e\right)^2\right]\right\}\\
&>0,
\end{aligned}$$

which, by $u_H < 0$,[62] implies

$$\frac{Hu_{HH}}{u_H} < \frac{2\left(\rho+2\theta\right)}{\rho+\theta}\frac{\Delta f/\Delta e-1}{\left(\Delta f/\Delta e\right)^2+1}\frac{H}{\Delta e}.$$

□

[62]This sign is elicited in the existence of internal habit formation question.

F.7 Proof of Proposition 6

Proof. Under third-order elicitation, a respondent preferring Universe One for a better future experience implies

$$\begin{aligned}
&\frac{1}{2}\frac{1}{\rho}\Bigg\{u_C\Delta f+\frac{1}{\rho+\theta}u_H\left(\rho\Delta e+\theta\Delta f\right)+\frac{1}{2}\Bigg[u_{CC}\left(\Delta f\right)^2+2\frac{1}{\rho+\theta}u_{CH}\Delta f\left(\rho\Delta e+\theta\Delta f\right)\\
&\quad+\frac{1}{\left(\rho+\theta\right)\left(\rho+2\theta\right)}u_{HH}\left(\rho\left(\rho+\theta\right)\left(\Delta e\right)^2+2\rho\theta\Delta e\Delta f+2\theta^2\left(\Delta f\right)^2\right)\Bigg]\Bigg\}\\
&\quad+\frac{1}{2}\frac{1}{\rho}\Bigg\{-u_C\Delta f-\frac{1}{\rho+\theta}u_H\left(\rho\Delta e+\theta\Delta f\right)+\frac{1}{2}\Bigg[u_{CC}\left(\Delta f\right)^2+2\frac{1}{\rho+\theta}u_{CH}\Delta f\left(\rho\Delta e+\theta\Delta f\right)\\
&\quad+\frac{1}{\left(\rho+\theta\right)\left(\rho+2\theta\right)}u_{HH}\left(\rho\left(\rho+\theta\right)\left(\Delta e\right)^2+2\rho\theta\Delta e\Delta f+2\theta^2\left(\Delta f\right)^2\right)\Bigg]\Bigg\}\\
&\quad-\frac{1}{2}\frac{1}{\rho}\left\{\left[u_C+\frac{\theta}{\rho+\theta}u_H\right]\Delta f+\frac{1}{2}\left[u_{CC}+2u_{CH}\frac{\theta}{\rho+\theta}+u_{HH}\frac{2\theta^2}{\left(\rho+\theta\right)\left(\rho+2\theta\right)}\right]\left(\Delta f\right)^2\right\}\\
&\quad-\frac{1}{2}\frac{1}{\rho}\left\{\left[u_C+\frac{\theta}{\rho+\theta}u_H\right]\left(-\Delta f\right)+\frac{1}{2}\left[u_{CC}+2u_{CH}\frac{\theta}{\rho+\theta}+u_{HH}\frac{2\theta^2}{\left(\rho+\theta\right)\left(\rho+2\theta\right)}\right]\left(\Delta f\right)^2\right\}\\
=&\frac{1}{2}\frac{1}{\rho}\Bigg[u_{CC}\left(\Delta f\right)^2+2\frac{1}{\rho+\theta}u_{CH}\Delta f\left(\rho\Delta e+\theta\Delta f\right)\\
&\quad+\frac{1}{\left(\rho+\theta\right)\left(\rho+2\theta\right)}u_{HH}\left(\rho\left(\rho+\theta\right)\left(\Delta e\right)^2+2\rho\theta\Delta e\Delta f+2\theta^2\left(\Delta f\right)^2\right)\Bigg]\\
&\quad-\frac{1}{2}\frac{1}{\rho}\left[u_{CC}+2u_{CH}\frac{\theta}{\rho+\theta}+u_{HH}\frac{2\theta^2}{\left(\rho+\theta\right)\left(\rho+2\theta\right)}\right]\left(\Delta f\right)^2\\
=&\frac{1}{2}\frac{1}{\rho}\left[2u_{CH}\frac{\rho}{\rho+\theta}\Delta f\Delta e+\frac{1}{\left(\rho+\theta\right)\left(\rho+2\theta\right)}u_{HH}\left(\rho\left(\rho+\theta\right)\left(\Delta e\right)^2+2\rho\theta\Delta e\Delta f\right)\right]\\
>&0,
\end{aligned}$$

which, by $u_{HH}<0$,[63] implies

$$\frac{u_{CH}}{u_{HH}}<-\frac{\left(\rho+\theta\right)\Delta e+2\theta\Delta f}{2\left(\rho+2\theta\right)\Delta f}.$$

□

[63] This sign is elicited in the Hu_{HH}/u_H question.

F.8 Proof of Proposition 7

Proof. Under third-order elicitation, a respondent preferring Universe One for a better future experience implies

$$
\begin{aligned}
&\frac{1}{2}\frac{1}{\rho}\left\{\left[u_C+\frac{\theta}{\rho+\theta}u_H\right]\Delta f+\frac{1}{2}\left[u_{CC}+2u_{CH}\frac{\theta}{\rho+\theta}+u_{HH}\frac{2\theta^2}{(\rho+\theta)(\rho+2\theta)}\right](\Delta f)^2\right\}\\
&+\frac{1}{2}\frac{1}{\rho}\left\{\left[u_C+\frac{\theta}{\rho+\theta}u_H\right](-\Delta f)+\frac{1}{2}\left[u_{CC}+2u_{CH}\frac{\theta}{\rho+\theta}+u_{HH}\frac{2\theta^2}{(\rho+\theta)(\rho+2\theta)}\right](\Delta f)^2\right\}\\
&-\frac{1}{2}\left[\frac{1}{\rho+\theta}u_H\Delta e+\frac{1}{2}\frac{1}{\rho+2\theta}u_{HH}(\Delta e)^2\right]-\frac{1}{2}\left[\frac{1}{\rho+\theta}u_H(-\Delta e)+\frac{1}{2}\frac{1}{\rho+2\theta}u_{HH}(\Delta e)^2\right]\\
=&\frac{1}{2}\frac{1}{\rho}\left[u_{CC}+2u_{CH}\frac{\theta}{\rho+\theta}+u_{HH}\frac{2\theta^2}{(\rho+\theta)(\rho+2\theta)}\right](\Delta f)^2-\frac{1}{2}\frac{1}{\rho+2\theta}u_{HH}(\Delta e)^2\\
>&0,
\end{aligned}
$$

which, by $u_{HH}<0$,[64] implies

$$
\frac{u_{CC}}{u_{HH}}<\frac{\rho}{\rho+2\theta}\left(\frac{\Delta e}{\Delta f}\right)^2-\frac{2\theta}{\rho+\theta}\frac{u_{CH}}{u_{HH}}-\frac{2\theta^2}{(\rho+\theta)(\rho+2\theta)}.
$$

□

F.9 Proof of Proposition 8

Proof. Choosing Universe One over Universe Two implies

$$
u(C,H(\Delta C>0,\Delta C_{others}=0))>u(C,H(\Delta C=0,\Delta C_{others}>0)).
$$

Subtracting the baseline utility from both sides to get

$$
\begin{aligned}
&u(C,H(\Delta C=0,\Delta C_{others}>0))-u(C,H(\Delta C=0,\Delta C_{others}=0))\\
&<u(C,H(\Delta C>0,\Delta C_{others}=0))-u(C,H(\Delta C=0,\Delta C_{others}=0))\\
&<0
\end{aligned}
$$

where the last inequality because of the existence of internal habit formation. □

[64] This sign is elicited in the Hu_{HH}/u_H question.

F.10 Proof of Proposition 9

Proof. Take $M = \$5000$, then $M - (1-\omega)\,\Delta C - \omega \Delta C_{others} > 0$ in all the questions for external habit.

A respondent preferring Universe One for a better future experience implies

$$\begin{aligned}
&\int_0^\infty e^{-\rho t}\left[u_H e^{-\theta t}\left((1-\omega)\,\Delta C\right) + \frac{1}{2}u_{HH}\left(e^{-\theta t}\left((1-\omega)\,\Delta C\right)\right)^2 + \cdots\right]dt \\
&- \int_0^\infty e^{-\rho t}\left[u_H e^{-\theta t}\left(\omega \Delta C_{others}\right) + \frac{1}{2}u_{HH}\left(e^{-\theta t}\left(\omega \Delta C_{others}\right)\right)^2 + \cdots\right]dt \\
=&\left[(1-\omega)\,\Delta C - \omega \Delta C_{others}\right]\int_0^\infty e^{-\rho t}\left[u_H e^{-\theta t}\right. \\
&\left. + \frac{1}{2}u_{HH}\left(e^{-\theta t}\right)^2\left(\left((1-\omega)\,\Delta C + \omega \Delta C_{others}\right)\right) + \cdots\right]dt \\
>&0,
\end{aligned}$$

which, by Lemma 3, implies

$$(1-\omega)\,\Delta C - \omega \Delta C_{others} < 0$$

or

$$\omega > \frac{\Delta C}{\Delta C + \Delta C_{others}}.$$

□

F.11 Proof of Proposition 10

Proof. Under first-order elicitation, a respondent preferring Universe One for a better future experience implies

$$\frac{1}{\rho}\left[u_{C_{others}} + \frac{\theta\omega}{\rho+\theta}u_H\right]\Delta C^{U1}_{others} - \frac{\omega}{\rho+\theta}u_H \Delta C^{U2}_{others} > 0$$

which, by $u_H < 0$,[65] implies

$$\frac{u_{C_{others}}}{u_H} < \frac{\omega}{\rho+\theta}\left(\rho\frac{\Delta C^{U2}_{others}}{\Delta C^{U1}_{others}} - \theta\right).$$

□

[65]This sign is elicited in the existence of habit formation question.

F.12 Proof of Proposition 12

Proof. Let

$$\begin{aligned} Q(a) \equiv \int_0^1 & e^{-\rho t} \left\{ u_C a + \frac{1}{2} u_{CC} a^2 + u_{CH} \left(\Delta c \left(1 - e^{-\theta t}\right) a \right) + \cdots \right. \\ & \left. + \left[u_H \left(1 - e^{-\theta t}\right) a + \frac{1}{2} u_{HH} \left(\left(1 - e^{-\theta t}\right) a \right)^2 + \cdots \right] \right\} dt \\ & + \int_1^\infty e^{-\rho t} \left[u_H e^{-\theta t} \left(1 - e^{-\theta}\right) a + \frac{1}{2} u_{HH} \left(e^{-\theta t} \left(1 - e^{-\theta}\right) a \right)^2 + \cdots \right] dt. \end{aligned}$$

Under exact elicitation, a respondent preferring Universe One for a better future experience implies

$$\begin{aligned} Q(\Delta e) - e^{-\rho} Q(\Delta f) &= \Delta e \frac{Q(\Delta e)}{\Delta e} - e^{-\rho} \Delta f \frac{Q(\Delta f)}{\Delta f} \\ &> \left(\Delta e - e^{-\rho} \Delta f \right) \frac{Q(\Delta e)}{\Delta e} \\ &> 0, \end{aligned}$$

where the first inequality holds because of diminishing marginal utility $\frac{Q(\Delta e)}{\Delta e} > \frac{Q(\Delta f)}{\Delta f} > 0$ for $\Delta f > \Delta e > 0$. This implies

$$\Delta e - e^{-\rho} \Delta f > 0$$

or

$$\rho > -\ln \frac{\Delta e}{\Delta f}.$$

□