Survey Evidence on Habit Formation

Jiannan Zhou*

April 30, 2020

Abstract

Habit formation has 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. However, habit formation is surrounded by controversies over its existence, specifications, and implications. To fill this gap, I document new and extensive micro evidence for habit formation, through designing and fielding a survey jointly eliciting ten preference parameters informative about habit formation. Habit forms both internally and externally, depreciates by around two-thirds annually, and has an about equisized welfare impact as peer effect. I also propose and implement four tests of additive and multiplicative habits and find that these ubiquitous preferences are rejected. Evidence-based simulations show that combining habit formation with peer effect could explain the Easterlin paradox. Finally, this paper advances the preference elicitation methods by introducing higher-than-first-order approximations, enabling elicitation of preference parameters not previously elicitable.

Keywords: Habit formation, micro evidence, peer effect, Easterlin paradox, preference elicitation.

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

1 Introduction

Habit formation refers to the phenomenon of response decrement to repetitive stimulation.[1] (Total consumption) habit formation has been used to explain many important phenomena in, among other areas, 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, specifications, and implications of habit formation, leaving open the foundational question of whether, how, and when to use habit formation in economics. To answer the question, this paper employs survey experiments to document new and extensive micro evidence on habit formation, as follows.

First, habit forms both internally and externally. Depending on its source, habit formation can be categorized into internal habit formation (habit based on one's own past consumption) and external habit formation (habit based on other people's past consumption). Literature investigating evidence of habit formation has mostly focused on the existence of internal habit formation (see column 1 of Table 1). Existing studies show that explaining macrodata tends to require more significant evidence for internal habit formation than microdata suggests. Existence of external habit formation has received much less attention, and its existing micro evidence does not support the popular modeling practice of assuming only external habit formation (column 2 of Table 1).[3] Internal and external habit formation can have dramatically different implications for optimal tax policy and welfare analysis (Ljungqvist and Uhlig, 2000, 2015). Through stated-preference experiments that differentiate between the two types of habit formation, this paper documents micro evidence for the existence of both internal and external habit formation. Few authors have studied the com-

[1]This definition of habit formation differs from the day-to-day notion of the cue-routine-reward habit, but is what the current economic models of habit formation are trying to capture and is consistent with the biology literature. Because habit formation is already a widely accepted term in economics, this paper will continue using it to label this phenomenon. This notion of habit formation sets it apart from path dependence, with which it is sometimes confused. It is also worth noting that habit formation is different from desensitization, which would imply reduced responses to small changes. Habit formation increases responses to small changes.

[2]For example, equity premium puzzle and stock market behavior (Constantinides, 1990; Detemple and Zapatero, 1991; Campbell and Cochrane, 1999); excess smoothness and excess sensitivity of consumption (Fuhrer, 2000; Boldrin, Christiano, and Fisher, 2001); and high growth causing high saving (Carroll, Overland, and Weil, 2000). Because the explanations of these and many other important economic phenomena are achieved through the habit formed on total consumption, the paper focuses on this type of habit formation and abstracts from habits formed on individual varieties of consumption (e.g., Ravn, Schmitt-Grohé, and Uribe, 2006). Hereafter, following the literature, this paper does not add qualifiers like *total* or *consumption* to the word *habit* when referring to this habit.

[3]For popular models with only external habit formation, see, e.g., Abel (1990); Campbell and Cochrane (1999); Smets and Wouters (2007); Dou, Lo, Muley, and Uhlig (2017).

position of these two types of habit formation, and its only formal estimate in the literature is macrodata-based (column 3 of Table 1). Using microdata and allowing habit to form both internally and externally as per Grishchenko (2010), this paper estimates that external habit formation accounts for a small portion (about 18%) of the habit.

Second, habit depreciates by about two thirds per year. Most habit formation specifications depend on two parameters: habit depreciation rate[4] and habit intensity. Existing research has focused primarily on estimating the habit intensity parameter (Havranek, Rusnak, and Sokolova, 2017; column 4 of Table 1) while treating the habit depreciation rate largely as a free parameter (column 5 of Table 1). A potential reason for the current state of the literature is a lack of recognition of the importance of the habit depreciation rate parameter. To illustrate this parameter's importance, I show that simply changing its value can significantly affect the performance of habit formation models. This paper provides an estimate of the parameter using micro-level variations that are cleaner than those used by existing estimates.[5] The estimate is also more robust to specification errors, due to the use of a nonparametric felicity function in the elicitation and estimation, and aggregated for a representative agent.

Third, neither the additive nor multiplicative habit preference is consistent with people's spending behavior. Almost all current habit formation models in the literature assume either of these two habit utility functions (column 6 of Table 1). The conclusions drawn from these models are, therefore, joint estimates and tests with specifications of questionable validity. In a general utility function naturally nesting these two formulations, this paper proposes and implements four tests of the preference specifications. This is the first time such tests have been done in the literature: not only are the tests new, but also their implementation requires advancing existing methods of preference elicitation.[6] The tests are based on the mapping between the preferences and their indifference curves. The test results imply that both habit utility functions are rejected with very high confidence, casting doubt on the validity of the predictions from the overwhelming majority of existing habit formation models. Even though these two common specifications are rejected, estimates of the signs of all the elicited utility derivatives in the general preference are consistent with the idea of habit formation,[7]

[4]This study focuses on the depreciation rate rather than the catch-up rate because the latter varies under different normalizations of habit, whereas the former is invariant to such normalizations. Habit depreciation rate fully pins down the habit catch-up rate for any given normalization of habit.

[5]For example, the variations are free from the confounding factors like liquidity constraints and the measurement errors in existing micro-level panel data on total consumption.

[6]See the end of this section for more details.

[7]Specifically, $u_H < 0$ and $u_{CH} > 0$.

suggesting that habit formation preferences consistent with the micro evidence could be found.[8]

Fourth, the welfare impacts of habit formation and peer effect are about the same in size. As two important interdependent preferences, peer effect allows interpersonal dependence, while habit formation allows intertemporal dependence (in internal and external habit formation) as well as interpersonal dependence (in external habit formation). Previous researchers have found a strong welfare impact from peer effect (Luttmer, 2005; De Giorgi, Frederiksen, and Pistaferri, 2020) but have disagreed on the strength of the welfare impact relative to habit formation. Through estimating linearized consumption Euler equations, Alvarez-Cuadrado, Casado, and Labeaga (2015) find internal habit formation to be as strong as peer effect, whereas Ravina (2019) finds internal habit formation to be about 70% stronger than peer effect. Allowing both internal and external habit formation, this paper provides a formal estimate of the relative strength of the two phenomena[9] without taking a problematic stance on the functional forms of the utility function.

Fifth, combining habit formation with peer effect could generate the happiness–income pattern of the Easterlin paradox. Easterlin (1973, 1974) highlighted the tension between the positive cross-sectional correlation and zero time-series correlation of happiness and income and proposed peer effect as an explanation in light of its effect on averaging happiness. As happiness data accumulated over time, the literature discovered that the zero time-series correlation tends to hold only in the long run, whereas the short-run correlation is generally positive (Stevenson and Wolfers, 2008; Sacks, Stevenson, and Wolfers, 2012; Easterlin, 2017).[10] Habit formation has been proposed as a potential explanation for the original version of the paradox (Easterlin, 1995) and, specifically, for the relatively newly discovered temporal heterogeneity of the correlation (Clark, Frijters, and Shields, 2008). To the best of my knowledge, evidence on whether habit formation can actually explain the paradox is absent from the literature. Since their specifications are rejected, structural simulations under any existing habit formation models have unknown credibility. Using this paper's extensive evidence on habit formation that is free from such specification errors, I conduct semi-structural simulations and find that, when coupled with peer effect, habit

[8]I leave this direction to future research.

[9]Formal in the sense of being able to report an interval estimate, in addition to the point estimate, of the parameter that governs the relative strength of the welfare impacts of the two phenomena.

[10]There is an ongoing debate on whether the long-run gradient is exactly zero or slightly positive. This paper intends not to participate in the debate, as it supplies no new evidence on happiness measures, and its discussion very easily accommodates both views. The goal is to show how habit formation (and peer effect) affects the relationship between income (or consumption) and happiness.

formation can explain the observed happiness–income pattern across all the dimensions: cross-section, short-run, and long-run.

The intuition is best illustrated by an analogy that could be called "running against an escalator." Imagine that you are about to run up, with a uniform speed, against a down escalator that initially is still but, once you step onto it, will gradually accelerate to your running speed. The number of 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 peer effect. For a while after you step onto the escalator, you run faster than the escalator and therefore your elevation increases, implying a positive correlation between number of stairs run and elevation reached, 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 number of stairs run and elevation reached, just like the long-run nil (or low) happiness–income gradient. People who run faster plateau at higher elevations, implying a positive correlation between number of stairs run and elevation reached, just like the cross-sectional 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.

To document the new and extensive micro evidence for habit formation, this paper uses a general model that is agnostic about the existence, specifications, and implications of habit formation while still allowing the extraction of useful information from data. The generality is essential, not just for nesting existing habit formation models that are heterogeneous along many dimensions but also for uncovering and reducing specification errors that potentially exist in all current habit formation models. Two recent papers have investigated habit formation models in this direction. Chen and Ludvigson (2009) allow habit to evolve in nonparametric ways and to form either internally or externally but maintain the parametric assumptions of additive habit and power utility. Crawford (2010) relaxes parametric assumptions for both the felicity function and habit evolution but allows only internal habit formation. Neither of the papers' models nests and therefore investigates the common multiplicative habit specification, and both papers assume limited numbers of lags, up to four quarters, in the habit evolution. The model in this paper is more general, in that it uses a nonparametric felicity function that relaxes the joint concavity of consumption and habit to nest the common multiplicative habit and allows an infinite number of lags in the habit evolution. To extract useful information in such a general framework, I identify ten structural preference parameters[11] that govern the existence, specifications, and implications of habit formation and are instrumental in addressing the controversies.

To estimate all the preference parameters, I designed stated-preference experiments that jointly elicit them, implemented the experiments in a survey, and fielded two waves of the survey on Amazon Mechanical Turk (MTurk). The experiments create variations in past and future consumption[12] and exploit the implied welfare variations to uncover the preference parameters. The survey was anonymous, and participation was voluntary. The benchmark estimation uses responses from 359 and 139 U.S. participants in the respective waves, who spread across the U.S. and match the U.S. population on all the demographic characteristics the survey collected.

The reliance on stated-preference data raises questions regarding the validity of the results. However, all other methods for addressing the controversies suffer from severe drawbacks. Real choices are often plagued by identification and data issues (Kimball and Shapiro, 2008). In the context of habit formation, a lack of required variations in real choices has confined the literature to mostly studying three, barely touching two more, and completely ignoring the other five of the ten preference parameters this paper estimates, all of which are crucial to addressing the controversies over habit formation. Furthermore, real choices often come from competitive markets where the price-taking behaviors rule out the possibility of testing the common multiplicative habit. This is because a utility-maximizing agent, taking prices as given, will never choose and therefore will never be observed choosing an interior bundle on the concave region(s) of the indifference curves of non-quasiconcave preferences (Samuelson, 1950), like the multiplicative habit.[13] Field and laboratory experiments are impracticable because of the prohibitive financial and time costs required to create variations in total consumption on the scale that is meaningful for macroeconomics and finance[14] (e.g., several thousand U.S. dollars of monthly consumption per person) and for the time span that is relevant for habit formation (e.g., two years).[15]

[11] 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 peer effect, and two quantities concerning the existence of internal and external habit formation.

[12] The exact variation for eliciting each parameter is different from that for another parameter. See Section 4 for details.

[13] Multiplicative habit is non-quasiconcave when habit intensity is less than 1. Habit intensity less than 1 is required for the intuitive regularity of higher consumption leading to higher steady-state utility. Existing estimates of habit intensity are also consistent with this restriction (see, e.g., Fuhrer, 2000; Kapteyn and Teppa, 2003; Lubik and Schorfheide, 2004; Ravina, 2019).

[14] Most current habit formation models are built to explain phenomena in these fields.

[15] Imitating the scale and time span using hypothetical elements in field and laboratory experiments make

Stated-preference experiments, therefore, seem to be the only feasible method for an extensive investigation of habit formation, which is required to address the controversies surrounding habit formation. The validity of the method rests on the assumption of truthful preference revelation, and potential response biases and errors cause deviations from the assumption. Response biases and errors can be and have been carefully studied and dealt with, and the economic studies applying stated-preference experiments can be traced back to Thurstone (1931) and span many fields: among others, behavioral economics (e.g., Kahneman and Tversky, 1979), environmental economics (e.g., Johnston, Boyle, Adamowicz, Bennett, Brouwer, Cameron, Hanemann, Hanley, Ryan, Scarpa, et al., 2017), and health economics (e.g., Ameriks, Briggs, Caplin, Shapiro, and Tonetti, 2019). This study deals with potential response biases and errors through the design and implementation of the stated-preference experiments, survey, estimation, and robustness checks. The stated-preference evidence is important, not only because it can feasibly shed light on the preference parameters that revealed-preference methods cannot, but also because it complements the revealed-preference evidence for the preference parameters that both the stated- and revealed-preference methods can illuminate. The complementarity derives from the fact that the limitations of the stated-preference methods—response biases and errors—tend to be orthogonal to the aforementioned limitations of the revealed-preference methods.

In addition to providing the new and extensive micro evidence on habit formation, this paper also advances the stated-preference method for eliciting structural preference parameters. Following Barsky, Juster, Kimball, and Shapiro (1997), most existing research that elicits structural preference parameters focuses on fully parametric preferences.[16] To be immune to specification errors, some researchers have dispensed with certain parametric assumptions and used first-order approximations in eliciting parameters of the nonparametric part(s) of semiparametric or nonparametric preferences (e.g., Benjamin, Heffetz, Kimball, and Szembrot, 2014; Benjamin, Cooper, Heffetz, and Kimball, 2019). This paper extends the literature by using higher, including the infinitieth, order of approximation in such preference elicitation. This advancement not only allows for potentially drastic improvement of the accuracy of preference elicitation, but also enables the elicitation of preference pa-

the experiments dependent on the same assumptions on which stated-preference experiments rely. Additional assumptions are also needed for field and laboratory experiments to create within-individual-time variations in habit, because they can specify only one historical path of consumption for each individual at any given point in time.

[16]See, e.g., Kapteyn and Teppa (2003); Sahm (2007); Kimball, Sahm, and Shapiro (2008); Kimball and Shapiro (2008); Kimball, Sahm, and Shapiro (2009); Kimball, Ohtake, Reck, Tsutsui, and Zhang (2015).

rameters that have not been elicitable: e.g., to elicit the (ratios of) utility derivatives of the second order, as required to implement the tests of additive and multiplicative habits, approximations of at least the second order are necessary.

This paper proceeds as follows. Section 2 presents the general model and survey instrument. Section 3 summarizes the data and statistical model. Section 4 contains the elicitation, estimate, and implication of each preference parameter of interest. Section 5 explores the explanation of the Easterlin paradox. Section 6 checks robustness, and Section 7 concludes.[17]

2 Methodology

2.1 Model

This section presents the general model that is agnostic about the existence, specifications, and implications of habit formation.

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 is habit, and ρ is time discount rate. Henceforth the time index will be omitted for brevity, and doing so will cause no ambiguity. Following the literature, this paper maintains expected utility and exponential time discounting.[18] As is always the case in existing habit formation models, the utility function is analytic and satisfies positive monotonicity of consumption ($u_C > 0$) and diminishing marginal utility of consumption ($u_{CC} < 0$). These assumptions aid elicitation and estimation without interfering with the evidence this paper provides. In particular, they leave open whether and how habit affects utility.[19] The respondent's utility can depend on other variables (e.g., labor), but because they will be kept constant in the survey, not explicitly listing them as the arguments of the utility function results in no loss of generality. In the discussion of survey questions involving changes in things other than self-spending and habit (e.g., other people's spending), the additional variable(s) of the utility function will be explicitly shown.

[17]Please see the online Appendix for observational equivalence of linear and nonlinear habit evolutions, aggregation of the preference parameters, response distributions, and proofs of elicitation propositions.

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

[19]As a technical note, to maintain the generality of the utility function under the infinitieth-order "approximation," if habit exists, it is necessary to assume and therefore this paper assumes that the positive $\partial^n u/\partial H^n$'s, if any, are bounded from above. See Lemma 3 of the online Appendix for details. Under common parameter values, the ubiquitous additive and multiplicative habits with power utility satisfy the bounds specified in the lemma.

Table 1: Estimates of Habit Parameters in Selected Literature

Study	Internal habit[a] (1)	External habit[a] (2)	ω (3)	α (4)	θ[e] (5)	Additive or multiplicative[g] (6)
Panel A. Microdata						
Naik and Moore (1996)	Y	*(N)*	*(0)*	0.08	*(Y)*	*(A)*
Dynan (2000)	N	*(N)*	*(0)*	-0.04	*(Y)*	*(A)*
Guariglia and Rossi (2002)	N	*(N)*	*(0)*	-0.27	*(Y)*	*(A)*
Lupton (2002)	Y	*(N)*	*(0)*	0.23	9.2%/Y	*(A)*
Kapteyn and Teppa (2003)	Y	*(N)*	*(0)*	0.78	[f]	*(M)*
Rhee (2004)	Y, N[b]	*(N)*	*(0)*	0.61, 0.62	*(Y)*	*(A)*
Browning and Collado (2007)	Y, N[b]	*(N)*	*(0)*	0.01-0.14	*(Q)*	*(A)*
Alessie and Teppa (2010)	Y	*(N)*	*(0)*	0.21	*(Y)*	*(A)*
Iwamoto (2013)	N	*(N)*	*(0)*	-0.38	*(Y)*	*(A)*
Khanal, Mishra, and Nedumaran (2018)	Y	*(N)*	*(0)*	0.55	*(Y)*	*(A)*
Ravina (2019)	Y	N	0.03[c]	0.50	*(Q)*	*(A, M)*
Panel B. Macrodata						
Ferson and Constantinides (1991)	Y	*(N)*	*(0)*	0.64-0.97	*(M, Q, Y)*	*(A)*
Fuhrer (2000)	Y	*(N)*	*(0)*	0.80	99.9%/Q	*(M)*
Stock and Wright (2000)	Y, N[b]	*(N)*	*(0)*	[d]	*(M, Y)*	*(A)*
Smets and Wouters (2003)	*(N)*	Y	*(1)*	0.57	*(Q)*	*(A)*
Lubik and Schorfheide (2004)	Y	*(N)*	*(0)*	0.57	*(Q)*	*(M)*
Christiano, Eichenbaum, and Evans (2005)	Y	*(N)*	*(0)*	0.65	*(Q)*	*(A)*
Adolfson, Laséen, Lindé, and Villani (2007)	Y	*(N)*	*(0)*	0.69	*(Q)*	*(A)*
Smets and Wouters (2007)	*(N)*	Y	*(1)*	0.71	*(Q)*	*(A)*
Grishchenko (2010)	Y	N	0.00	0.90[c]	70.7%/Q	*(A)*
Korniotis (2010)	N	Y	0.79[c]	0.33[c]	*(Y)*	*(A)*
Altig, Christiano, Eichenbaum, and Lindé (2011)	Y	*(N)*	*(0)*	0.76	*(Q)*	*(A)*

Notes: The studies are selected for representativeness based on citation count, number of habit parameters estimated, and publication year. Each character not in parentheses is a parameter estimate. Characters in parentheses (and italics for further distinction) are assumed parameter values of the studies. Preference parameters in this table are from specializations of the following habit formation model:

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

where C and C_{others} are self and others' consumption, respectively, H is habit, α is habit intensity, θ is habit depreciation rate, and ω is external habit mixture coefficient.

[a] Y/N—exist/not exist.

[b] Estimates depend on goods or time horizon.

[c] Implied estimates.

[d] The study provides only confidence sets.

[e] M/Q/Y—habit depreciates fully at the end of a month/quarter/year.

[f] Geometric habit evolution speed of 0.07 (0.01).

[g] A/M—additive/multiplicative habit.

Habit evolves according to $\dot{H} = \theta (C - H)$, where θ is the habit depreciation rate. This specification is chosen for two reasons. First, it has been in the literature since at least Houthakker and Taylor (1970) and is the most commonly used habit evolution in the literature. Researchers have used different formulations of the evolution. However, the difference is either a simple rescaling of the unit of habit (e.g., Constantinides, 1990) or disappears in the steady state (e.g., Campbell and Cochrane, 1999). For general habit evolutions that are potentially nonlinear (even in the steady state), I show that they are observationally equivalent to this linear habit evolution under the general habit formation preference.[20] Second, this habit evolution has an intuitive unit, the same as that of consumption. For example, a person who has been spending \$5,000 per month for as long as they can remember has a habit of spending \$5,000 per month.

To document the extensive micro evidence for habit formation, I need information on whether habit affects utility and, if it does, the values of the preference parameters governing the effects of habit on the utility: θ, ρ, ratios of utility derivatives up to the second order ($-\frac{u_H}{u_C}$, $\frac{H u_{HH}}{u_H}$, $\frac{u_{CH}}{u_{HH}}$, and $\frac{u_{CC}}{u_{HH}}$), external habit mixture coefficient, and strength of habit formation relative to peer effect.[21] Eliciting, estimating, and using these preference parameters to shed light on the controversies surrounding habit formation is the primary subject of the rest of the paper.

2.2 Survey

To elicit the preference parameters, I design simple stated-preference experiments that identify them while controlling for potential confounding factors and response biases and errors. Because past spending determines habit and habit potentially affects people's well-being, the basic idea behind the stated-preference experiments is to compare the welfare implications of different spending paths.[22] The exact experiments vary from one parameter to another and will be discussed in detail in Section 4. This section sets the stage for that discussion by presenting the survey design.[23]

As discussed earlier, the elicitation of the preference parameters of interest requires

[20]See Section A of the online Appendix for the proof.

[21]The last two parameters are related to changes in other people's spending, a currently "hidden" argument of the utility. For details on the model in which the argument is unhidden and on the two parameters, see Sections 4.4 and 4.5.

[22]Because income affects utility through spending, the survey does not specify the income process except to tell the respondents that they could afford the spending profiles in the survey.

[23]See Section H of the supplemental material for additional details of the survey.

placing the comparison of spending paths in stated-preference or hypothetical situations. The survey starts with a preamble module that specifies the basic hypothetical environment in which comparisons of spending paths will be performed and instructs the respondents on the format of the core survey questions. Nine core modules follow, each containing specific variations in spending paths that elicit one or two of the preference parameters of interest.

The basic hypothetical situation is designed to be as simple as possible while still allowing elicitation of the parameters of interest and avoiding potential confounding factors that plague real-choice data. In particular, it frees the respondents from worrying about changes to the purchasing power of money, about durable goods, and about changes in preferences. The basic hypothetical situation 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.*

The survey has a set of questions testing respondents' understanding of this basic hypothetical situation. Only those who passed the test were able to proceed to the core modules of the survey.

The respondents did not know that the survey is about habit formation. They were only told 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 habit formation is, as we economists call it. The second reason was to avoid potential biases; I cannot prime respondents with habit formation while attempting to test its very existence.

To make the representation of a spending path intuitive and to simplify comparison across several of them, I draw it in a monthly spending graph (Figure 1). In such graphs, time is on the horizontal axis: past on the left, now in the middle, and future on the right. The bars above the time axis represent monthly spending and are drawn to scale and colored differently to help distinguish time horizons. The spending path of Figure 1 represents spending $7,000 per month in the past until now and $5,000 per month in the future starting now. The respondents went through instructions and were tested on reading the monthly spending graphs before being qualified to answer questions in the core modules.

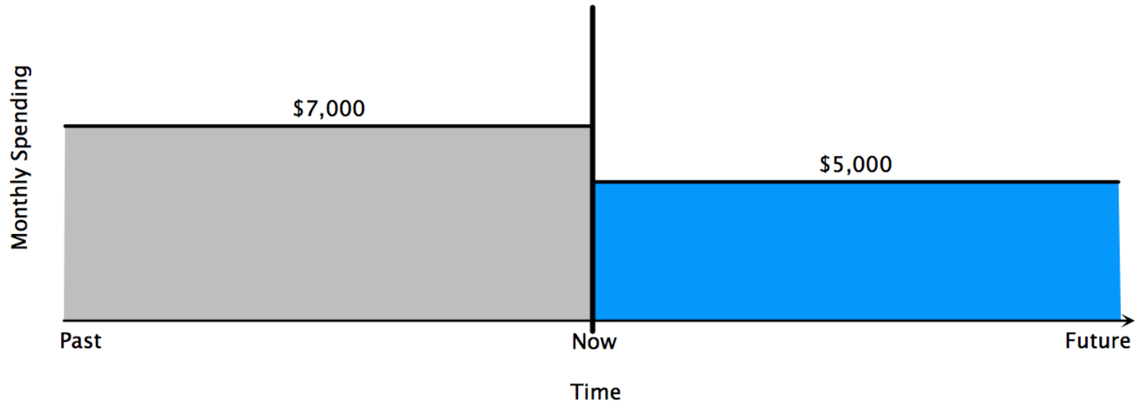


Figure 1: A typical monthly spending graph.

To alleviate the concern that each person has only one past spending path in reality, I invoke the metaphor of parallel universes, between which everything is the same except for the spending paths. I then ask the respondents which universe brings them a better future experience—how they feel in the future starting now. Figure 2 presents a screenshot of a typical survey question.

The survey is incentive-compatible for truthful preference revelation if the respondents truthfully reveal their preferences regardless of other respondents' choices. The anonymous online implementation of the survey rules out feasible mechanisms through which the respondents could know and influence each other. Due to the fact that the preference elicitation does not rely on, and therefore the survey does not elicit, respondents' exact valuation that is often the object of interest in willingness to pay or accept elicitation, concerns of under- or over-reporting of valuation do not apply, as long as the relative ranking of the (often two) options is truthfully reported. Because of the stated-preference nature of the core survey questions and because none of their options is inherently right or wrong, the only reasons for not revealing true ranking are misunderstanding of the survey questions, lack of attention, and protest responses.

To deal with these concerns, the stated-preference experiments and the survey are designed to minimize cognitive load as much as possible, which can be partly seen from the above discussion of the design of the representation of spending paths. To reinforce the idea that the only variation between universes is in spending, the survey reiterates it at the start of every core module. To help the respondents compare the graphical spending paths, the survey questions also tell them in words in what time horizon the spending differs. To help

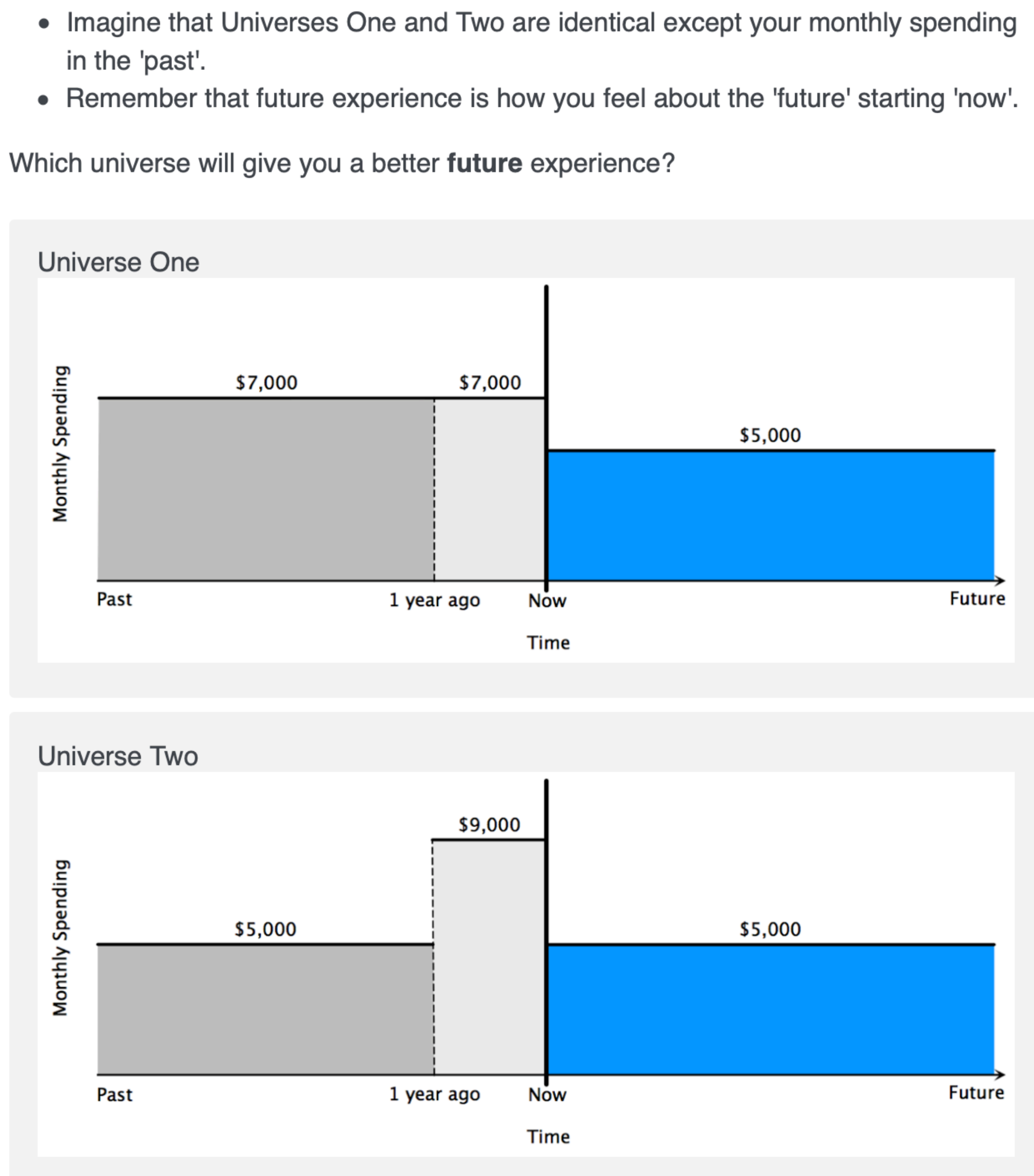


Figure 2: A typical survey question.

them distinguish past experience from future experience, they are asked to express preferences on both experiences. The survey also repeats the definitions of the experiences of interest and highlight the key words—past or future—to further remind the respondents of which experience a question asks about. To help the respondents avoid clicking on an option different from the one they want to choose, I integrate the spending graphs into clickable options. To help them confirm that they answer as they intend, I darken slightly the background of an option when their mice hover over it and darken completely the background of the option they select. As mentioned above, the survey tests respondents' understanding of the instructions whenever possible, and only those who pass the tests can enter the sample.

Attention checks spread throughout the survey, ranging from explicit ones, like the quiz on the basic hypothetical situation at the start of the survey, to implicit ones, such as time spent on each survey question. To encourage attention, I tell the respondents about the existence of such attention checks but do not tell them where they are or how to identify them. In addition, to encourage greater attention, I tell them in the survey's introduction that respondents whose responses are of high quality will be entered in a small ($1) lottery with winning odds of 1 in 100.[24] A series of relatively speculative checks on attention are also implemented through robustness checks (Section 6.3).

As an additional mechanism to guard against untruthful preference revelation, I conducted two waves of the survey and use a statistical model that jointly estimates the preference parameters to extract consistent responses across the two waves while taking care of remaining response biases and errors. To minimize the possibility of shirking or untruthful answers, the second wave was fielded 20 days after the first wave,[25] with the sequence of the core modules reordered and all options flipped.

3 Data and Statistical Model

3.1 Data

The two waves of the survey were fielded on MTurk, an online crowdsourcing platform for human intelligence tasks. A large number of psychology studies have conducted stated-preference or hypothetical experiments using this sample (Anderson, Allen, Plante, Quigley-McBride, Lovett, and Rokkum, 2018). The number of economic studies adopting this sample for stated-preference experiments has been growing.[26]

Participation in the study is voluntary and anonymous.[27] To avoid the potential influences of cultural differences, this study restricts the respondents to U.S. residents. From the

[24]Of the 550 responses I collected, six respondents were randomly chosen for this award.

[25]The two waves of the survey were fielded in July and August of 2018.

[26]See, e.g., Oster, Shoulson, and Dorsey, 2013; Kuziemko, Norton, Saez, and Stantcheva, 2015; Bordalo, Coffman, Gennaioli, and Shleifer, 2016; Benjamin, Cooper, Heffetz, and Kimball, 2019.

[27]According to the consent, each worker was paid $2.5 for the survey, corresponding to an hourly wage of about $4.5. The median hourly wage on MTurk was about $2 (Hara, Adams, Milland, Savage, Callison-Burch, and Bigham, 2018).

Table 2: Sample Statistics

	First wave (359 obs.)	Second wave (139 obs.)	United States
Age, median	38	37	38
Household income, median	\$50,001–\$60,000	\$50,001–\$60,000	\$57,652
Female percentage	53.2%	48.2%	50.8%
Household size, mean	2.69	2.71	2.63
Time on survey, mean	34'55"	33'36"	

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).

295 first-wave respondents who expressed affirmative 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 who were outside the United States, submitted duplicate responses, or were suspected of speeding, the sample has 359 and 139 responses from the respective waves.

Although the MTurk sample is potentially less representative of the U.S. population than national probability samples, it is more representative than in-person convenience samples (Berinsky, Huber, and Lenz, 2012), has been used widely in social sciences, and can provide consistent and economically meaningful data (Johnson and Ryan, 2018).

The sample I collected is consistent with the literature on the representativeness of the MTurk sample. Of all the demographic information reported by the respondents—age, gender, household income and size—the sample statistics are essentially the same as the national counterparts (Table 2). At the time of the survey, a typical respondent was about 38 years old, lived with another one or two people in a household with an annual income in the range of \$50,001 to \$60,000, was slightly more likely to be female if participating only in the first wave and slightly more likely to be male if participating in both waves, and spent a little over half an hour on the survey. Locations of the IP addresses associated with the survey responses indicate that the respondents spread across the United States (Figure 3) and show no sign of non-U.S. respondents pretending to be in U.S. locations using virtual private networks.

Eight of the ten preference parameters are identifiable to scale and estimated jointly. As a result of the joint estimation and potential remaining response biases and errors that need to be taken care of by the statistical model, response distributions for individual parameters alone are not particularly informative and are reported in Table A.1 of the online Appendix.

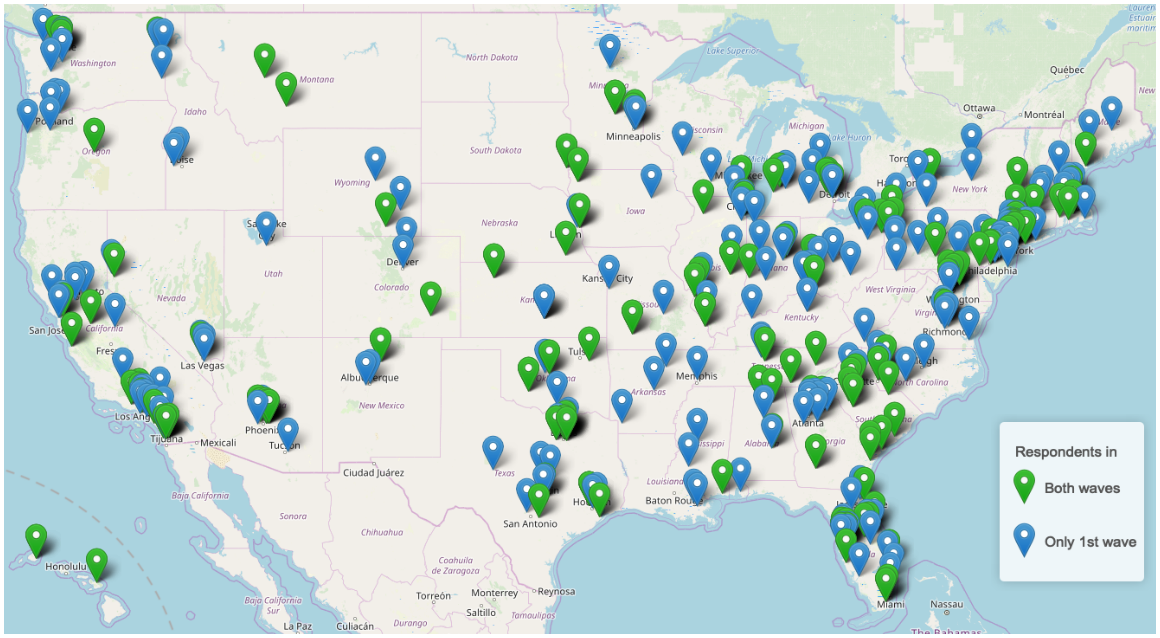


Figure 3: Locations of respondents.

Response distributions of parameters identifiable to sign only are presented at the places where their estimates are reported.

3.2 Statistical Model

The statistical model underlying the estimation addresses response biases and errors not addressed by the design and implementation of the stated-preference experiments and survey or by the elimination of invalid responses. Potential remaining response biases and errors are dealt with through robustness checks in Section 6.

I model an observed response for preference parameter x from individual i in wave w as $X_{i,w} \equiv \sum_k k \cdot 1\left(T_{k,\tilde{x}} \leq \tilde{x}_{i,w} \leq T_{k+1,\tilde{x}}\right)$, where the unobserved latent variable $\tilde{x}_{i,w} = x_i + \varepsilon_{i,x,w}$, and $T_{\{k\},\tilde{x}}$ denotes 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 bias and error $\varepsilon_{i,x,w}$ is drawn from $\mathcal{N}(0, \varsigma_{x,w}^2)$ independently of the true parameter value. A robustness check allows the means of the response biases and errors to be nonzero and to vary across waves and finds that the estimates of the means are indistinguishable from zero and that the estimates of the preference parameters are not significantly different from those under the specification here. For aggregation and computation, the parameters are assumed to be independent within a respondent. Because the respondents spread across the United States (Figure 3) and most likely did not know each other, the responses are assumed to be independent across respondents.

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

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

Given that almost all current habit formation models assume a representative agent, this paper focuses on the implication of the estimates for the representative-agent models with habit formation. In Section B of the online Appendix, I prove that 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 denotes the value of the representative agent's parameter x. Because $x_R = \mu_x$, the estimate of interest is that of μ_x.

To be consistent with the joint elicitation, this paper jointly estimates the preference parameters. To deal with the computational burden of the resulting high dimensional estimation, I use a Bayesian method to bypass the direct optimization associated with 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.

The implementation of Hamiltonian Monte Carlo uses uniform priors, not only to let data speak as much as possible but also to establish the equivalence between the posterior mode estimates and the maximum likelihood estimates.[28] I run ten Markov chains initialized from random diffuse starting points and collect 15,000 iterations of warmup and 25,000 draws of sample. I report all three Bayesian point estimators (posterior mode, mean, and median) and the highest posterior density or mass interval (HPDI or HPMI).

[28]Other common priors, like normal and conjugate priors, give the same estimates, suggesting that the information contained in the data override the influence of the priors.

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

Universe	1	2	3	4	5
First wave	56	4	10	2	29
Second wave	60	1	6	1	30

4 Elicitation, Estimation, and Implication

4.1 Existence of Internal Habit Formation

The fundamental characteristic of habit formation is response decrement to repetitive stimulation. 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. Therefore, internal habit formation is consistent with the utility difference $Q_H \equiv u(C, H + \Delta h) - u(C, H) < 0$ but not with $Q_H \geq 0$, for $\Delta h > 0$ due to higher past self-consumption.

To elicit the sign of Q_H, I vary the respondent's past spending while controlling for future spending (Figure 4), so that variation in future experience is induced only by different levels of habit. In this context, preferring a spending path with less past spending (low H) over one with more past spending (high H) implies $Q_H < 0$. It is worth emphasizing again that the survey does not prime the respondents with habit formation and that no assumption is made about the signs of derivatives of the felicity function with respect to habit.

The responses to this question show that the average respondent chose the lowest level of past spending for the best future experience—Universe One (Table 3), consistent with the existence of internal habit formation for the representative agent. The estimate of $\mathrm{sgn}(Q_H)$ confirms this (Table 4).

As a clarification, the existence of habit formation in people's spending behavior does not imply that habit formation is the deepest phenomenon influencing people's spending behavior. A phenomenon exists in people's behavior if the definition of the phenomenon matches people's behavior at the level of magnification closest to the phenomenon. This paper's evidence shows that people's spending behavior exhibits response decrement to repetitive stimulation, matching the definition of habit formation. Therefore, habit formation exists (in people's spending behavior). That habit formation exists, however, says nothing about whether it is the deepest possible explanation of people's spending behavior. The fact that biologists have found evidence for habituation of various behaviors across both humans and animals, including the amoeba, an organism without a neural system (Folger,

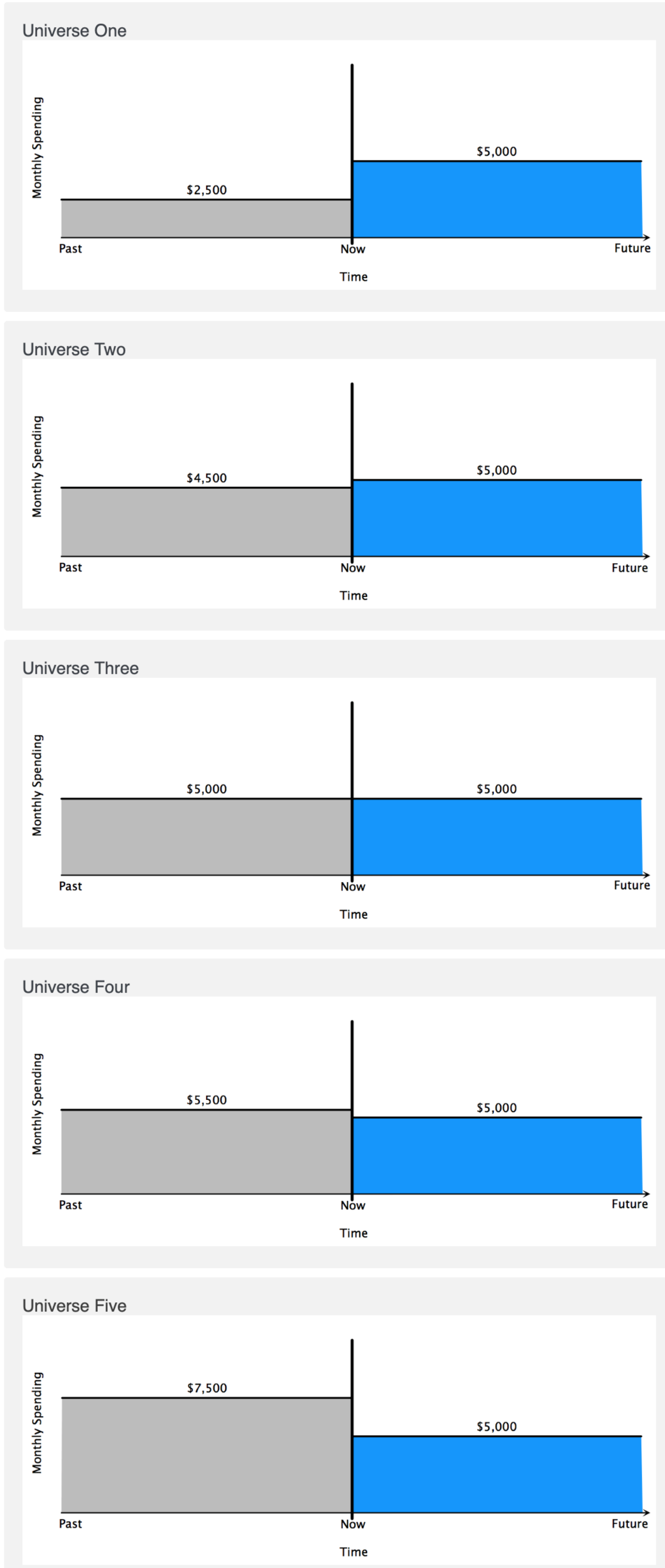


Figure 4: Monthly spending graphs of a survey question for the existence of internal habit formation.

Table 4: Estimates of Preference Parameters

	Mode	Mean	Median	95% HPDI/HPMI
sgn (Q_H)	-1.00	-1.00	-1.00	[-1.00, -1.00]
Habit depreciation rate	1.07	1.09	1.09	[0.88, 1.28]
Habit depreciation factor, annual	0.67	0.66	0.66	[0.59, 0.73]
$-u_H/u_C$	0.60	0.59	0.59	[0.49, 0.70]
Hu_{HH}/u_H	7.46	7.54	7.52	[6.70, 8.46]
u_{CH}/u_{HH}	-0.88	-0.88	-0.88	[-1.03, -0.73]
u_{CC}/u_{HH}	3.71	3.68	3.68	[3.01, 4.34]
External habit mixture coefficient	0.18	0.18	0.18	[0.09, 0.27]
$u_{C_{\text{others}}}/u_H$	1.03	1.04	1.04	[0.69, 1.39]
$u_{C_{\text{others}}}/u_C$	-0.61	-0.62	-0.61	[-0.86, -0.39]
$u_H/u_C + u_{C_{\text{others}}}/u_C$	-1.17	-1.21	-1.21	[-1.52, -0.91]

Note: The annual habit depreciation factor is calculated based on the habit depreciation rate: $\theta_{Factor} = 1 - e^{-\theta_{Rate}}$.

1926), seems to suggest the existence of deeper, and possibly universal, explanations for habit formation.[29] But at the level of magnification most closely associated with the phenomenon of habit formation—people's spending behavior—habit formation does describe people's behavior.

4.2 Habit Depreciation Speed

The speed at which habit depreciates is governed by θ as in $\dot{H} = \theta (C - H)$. The survey question eliciting θ varies the persistence and level of past spending (Figure 2) to induce a surjective mapping from θ to future experience (Proposition 1).

Proposition 1. $\theta > -\ln\left(1 - \frac{\Delta C_{U1}}{\Delta C_{U2}}\right)$ *if the respondent chooses Universe One over Universe Two for a better future experience in a habit depreciation rate question.*[30]

The intuition of the proposition is that choosing the spending path with more persistent past spending (Universe One) for a better future experience means that recent past spending matters more to habit than distant past spending does, which implies a fast habit depreciates speed. ΔC_{U1} (ΔC_{U2}) of the proposition denotes the difference between the monthly spending in Universe One (Universe Two) and the baseline monthly spending, \$5,000 per

[29]For example, one potential explanation for habit formation focuses on its evolutionary advantage: Rayo and Becker (2007) argue that habituating living standards increases the motivation to strive for more, which likely helps with survival.

[30]Because they are sufficient for the results of this paper, the elicitation propositions in the paper are stated as conditional statements, even though all of them can be strengthened to biconditional (if and only if) statements.

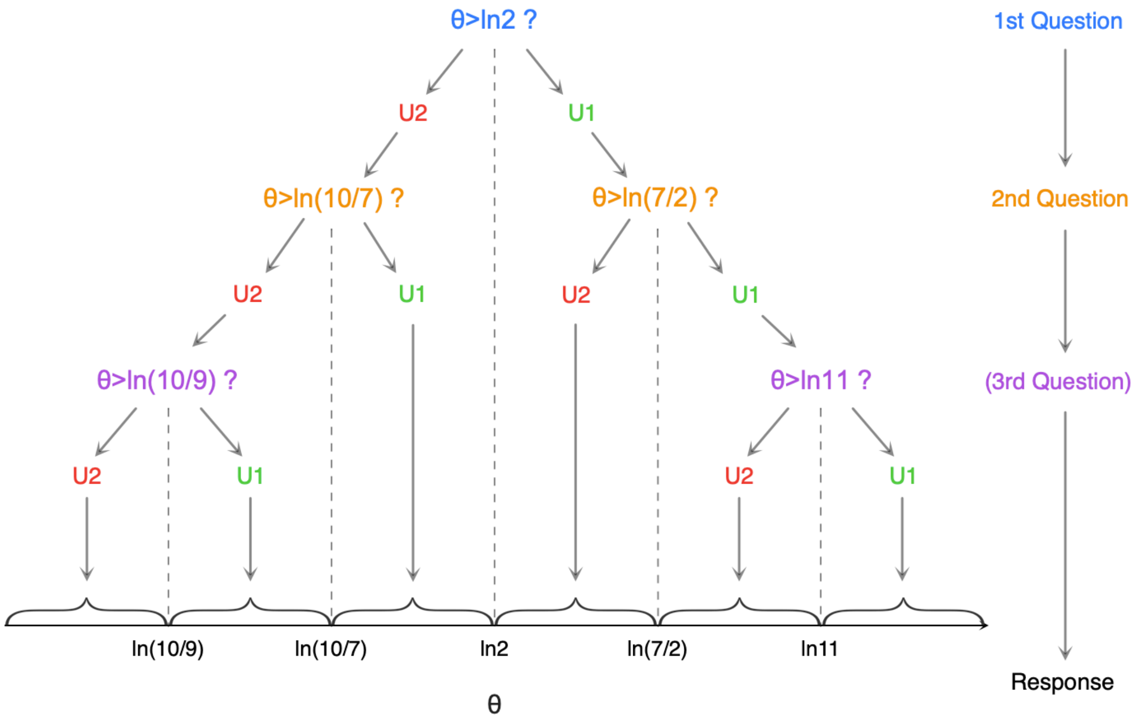


Figure 5: Unfolding brackets. U1 and U2 stand for Universe One and Universe Two, respectively.

month. In the example of Figure 2, $\Delta C_{U1} = \$2,000$ 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 $\theta < \ln 2$.[31]

The survey uses unfolding brackets to pin down a finer range of θ for each response: all respondents answered one to two follow-up questions that associate their responses with values of θ in one of the six brackets of Figure 5. For example, if a respondent chooses Universe One in the survey question corresponding to the θ threshold of $\ln 2$, the module continues with a follow-up question associated with the θ threshold of $\ln 7/2$. If Universe Two is then chosen, the module ends, and the response implies that the respondent's θ (with potential response biases and errors)[32] falls between $\ln 2$ and $\ln 7/2$.

Applying the statistical model to the responses and the response-parameter mappings leads to an estimate of about 1.07 for the habit depreciation rate, which corresponds to the

[31] 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.

[32] These remaining response biases and errors will be taken care of by the statistical model and robustness checks.

(annual) habit depreciation factor of about 0.67 (Table 4). This depreciation speed implies that about 90% of habit depends on the spending of the last two years, which is remarkably close to the finding in the psychology literature that income adaptation takes about two years.[33]

The speed at which habit depreciates is important. In simple additive habit formation models, the faster habit depreciates, the less risk averse agents of the model become because as habit adjusts faster with consumption, they fear less about not being able to meet their habitual level of spending. The probably most cited paper on habit formation, Campbell and Cochrane (1999),[34] also employs the additive habit.[35] In their model, however, the faster habit depreciates, the more risk averse model agents are. The reason is that the implied steady-state habit intensity[36] of their model is not constant but increases with the habit depreciation rate. The higher the habit intensity, the more likely 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 agents become. The net effect of a higher habit depreciation rate in their model is the sum of these two effects, which ultimately makes the agents of the model more risk averse.

Plugging the estimate of the habit depreciation speed into Campbell and Cochrane (1999) causes the agents to become so risk averse that the equity premiums they require are too high to have been observed historically (column 3 of Table 5). The time discount factor also has to be unrealistically low, 0.35 per year, to match the mean historical risk-free rate. When a more realistic annual time discount factor of 0.89 is used, which is the level Campbell and Cochrane (1999) choose, people become even more risk averse. They require an even 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 higher risk aversion drives up the motive for precautionary saving. This motive is so strong that people are willing to pay more than 92% of the principal to be able to transfer the remaining less than 8% of it to the next year. When one lowers the time discount factor or the habit depreciation factor, the model moments are

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

[34] Google Scholar reported that this paper had been cited 5,116 times as of April 26, 2020.

[35] 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.

[36] Under additive habit, the instantaneous utility is $u(C - \alpha H)$ where α is the habit intensity parameter. In Campbell and Cochrane's (1999) notation, $X = \alpha H$. After a steady state is reached, the implied steady-state habit intensity is, therefore, X/C.

Table 5: 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**	**0.59**	**0.30**
Time discount factor	-	0.89	0.35	**0.89**	0.43	0.71
Expected excess ln return	6.69%	6.71%	43.94%	101.52%	36.58%	16.51%
Std of excess ln return	15.20%	15.64%	31.78%	96.99%	29.33%	22.01%
Sharpe ratio	0.43	0.43	1.38	1.05	1.25	0.75
Mean risk-free rate	0.94%	0.94%	0.94%	-92.19%	0.94%	0.94%

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

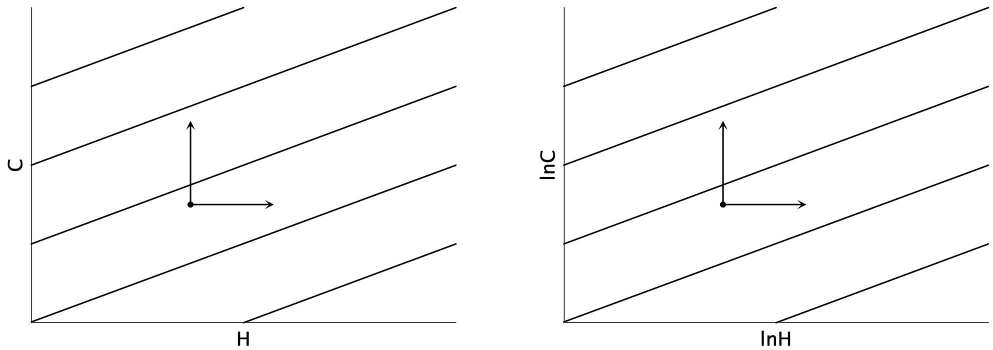
closer to reality, but the percentage differences are still at least 40% (columns 5 and 6), even when habit depreciates by only 30% each year, which is far away from the 99% HPDI of the habit depreciation factor.[37]

The survey respondents might not be representative of the marginal investors who price the assets. It is, however, unnecessary for the respondents to represent the marginal investors in every way possible. The above discussion remains valid as long as the typical habit depreciation speed of the respondents is the same or close to that of the marginal investors, which would be the case if this parameter is a deep preference parameter that does not vary significantly across demographics. Section 6.1 presents such evidence: the habit depreciation rate does not vary empirically 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.

The above discussion shows that the explanatory power of a popular habit formation model is significantly affected when the micro-based estimate of habit depreciation factor[38]

[37]Section E of the supplemental material shows that other results of Campbell and Cochrane (1999) are also sensitive to the habit depreciation speed.

[38]Note that the estimate has been aggregated for the representative agent of the model.



(a) Additive habit (b) Multiplicative habit ($(\ln C, \ln H)$ space)

Figure 6: Indifference maps for additive and multiplicative habits.

is plugged in. One must, however, take extra caution in interpreting this finding. Given that this paper's evidence supports the existence of habit formation, it is more likely that the way habit formation is modeled needs improvement (more on this in the next section) than that modeling habit formation is the wrong way to go. 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 fully explain asset pricing and other phenomena.

4.3 Testing Additive and Multiplicative Habits

Additive and multiplicative habits are used by basically all current habit formation models that have been taken to data. To see whether the micro evidence supports these two specifications, I propose and implement four tests of 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 $\frac{u_{CH}}{u_{CC}} \frac{u_C}{u_H} = 1$.*

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 slopes of the indifference curves. The two tests are the two bases spanning all such movements: increase H alone and increase C alone (Figure 6a).

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

In the space of $(\ln C, \ln H)$, the two tests of multiplicative habit have the same intuition as those of additive habit (Figure 6b).

Because the tests are functions of $-\frac{u_H}{u_C}$, $\frac{Hu_{HH}}{u_H}$, $\frac{u_{CH}}{u_{HH}}$, and $\frac{u_{CC}}{u_{HH}}$, their implementation requires eliciting these preference parameters. Due to the generality of the felicity function, elicitation of the preference parameters will be up to third-order approximations.

To elicit the slope of indifference curve, $-\frac{u_H}{u_C}$, I vary both future and past spending in the same direction to move along an indifference curve.[39]

Proposition 4. *Under the second-order approximation, $-\frac{u_H}{u_C} < \frac{(\rho+\theta)\Delta C_{future}}{\rho\Delta C_{past}+\theta\Delta C_{future}}$ if the respondent chooses Universe One over Universe Two for a better future experience in a slope of indifference curve question.*[40]

The estimate for the slope of the indifference curve is about 0.60 (Table 4). The implied positive sign of u_C is consistent with the assumption of positive monotonicity of consumption. The magnitude of this estimate implies that, to a first-order approximation, roughly 60% of utility changes resulted from consumption changes are eventually habituated, consistent with Van Praag and Frijters's (1999) finding that about 60% of the effect of income on happiness is lost with time.

To elicit $\frac{Hu_{HH}}{u_H}$, the survey presents the respondents with a trade-off between the level and fluctuation of past spending.[39] I estimate $\frac{Hu_{HH}}{u_H}$ to be about 7.46 (Table 4), which by the estimated $u_H < 0$, implies $u_{HH} < 0$.

Proposition 5. *Under the second-order approximation, $\frac{Hu_{HH}}{u_H} < \frac{2(\rho+2\theta)}{\rho+\theta}\frac{\Delta C_1/\Delta C_2-1}{(\Delta C_1/\Delta C_2)^2+1}\frac{H}{\Delta C_2}$ if the respondent chooses Universe One over Universe Two for a better future experience in a $\frac{Hu_{HH}}{u_H}$ question.*

The elicitation of $\frac{u_{CH}}{u_{HH}}$ rests on inducing fluctuations in both future and past spending at the same time.[39]

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

---

[39] The resulting monthly spending graphs are in Figures A.11 (for $-\frac{u_H}{u_C}$), A.12 (for $\frac{Hu_{HH}}{u_H}$), A.13 (for $\frac{u_{CH}}{u_{HH}}$), and A.14 (for $\frac{u_{CC}}{u_{HH}}$) of Section H of the supplemental material.

[40] ΔC_{past} and ΔC_{future} denote differences of the monthly spending in the question from the baseline monthly spending, \$5,000 per month. Knowledge about the time discount rate, ρ, is required to estimate the slope of indifference curve and some other preference parameters, the elicitation of which is relegated to Section G of the supplemental material because of this indirect interest in it.

Table 6: Statistics for Testing Additive and Multiplicative Habits

	Mode	Mean	Median	99% HPDI
$\frac{u_{CH}u_H}{u_{HH}u_C}$	0.52	0.52	0.52	[0.36, 0.70]
$\frac{u_{CH}u_C}{u_{CC}u_H}$	0.39	0.41	0.40	[0.26, 0.60]
$\frac{Hu_Hu_{CH}}{u_C(u_H+Hu_{HH})}$	0.46	0.46	0.46	[0.32, 0.62]
$\frac{Cu_Cu_{CH}}{u_H(u_C+Cu_{CC})}$	-0.25	-0.26	-0.25	[-0.36, -0.18]

The estimate of $\frac{u_{CH}}{u_{HH}}$ is about -0.88 (Table 4). Given the above estimate of $u_{HH} < 0$, $u_{CH} > 0$, consistent with the sensitization of habit formation: the higher habit is, ceteris paribus, the more valuable an additional unit of consumption is.[41]

$\frac{u_{CC}}{u_{HH}}$ is about the trade-off between two sources of fluctuations, one from future spending and the other from past spending.[39]

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

$\frac{u_{CC}}{u_{HH}}$ is estimated to be about 3.71 (Table 4), consistent with the 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 $\frac{u_{CC}}{u_{HH}}$, the left-hand-side statistics of the tests of additive and multiplicative habits can be calculated. Their point estimates are far away from one (Table 6), the right-hand side of the tests. Furthermore, one is far away from the 99% HPDIs of these statistics, implying that the micro evidence rejects both the additive and multiplicative habits with very high confidence.

It is worth emphasizing again that the evidence supports the existence of habit formation and that none of the estimates of the preference parameters rules out the possibility of an evidence-consistent habit formation preference, which might be the key to explain the model-data inconsistency discussed in the last section and other phenomena current habit formation models struggle to account for.

[41]Though not the focus of this paper, to the extent that the nonparametric felicity function of this paper nests the felicity functions of existing models of reference dependence with backward-looking averages (e.g., DellaVigna, Lindner, Reizer, and Schmieder, 2017), $u_{CH} > 0$ suggests that in the felicity function, consumption is not additively separable from habit or other backward-looking reference points, favoring habit formation models over the additively separable models of reference dependence with backward-looking averages.

4.4 Existence of External Habit Formation and Composition of Habit

The discussion so far has been holding other people's past spending constant and, therefore, has been abstracting from its potential effect on habit. This section presents evidence on whether and by how much other's past spending affects habit.

The existence of internal habit formation implies $u_H < 0$. It follows that seeing whether external habit formation exists is equivalent to seeing whether others' spending, denoted as C_{others}, affects one's own habit, H. Given the observational equivalence of linear and nonlinear habit evolutions,[42] I model the potential dependence of habit on others' spending as per Grishchenko (2010):

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

where the external habit mixture coefficient, ω, governs the contribution of others' spending on the habit. If ω equals 0, others' spending has no effect on the habit and, therefore, external habit formation does not exist. Otherwise, if ω is between 0 and 1, external habit formation exists and the value of ω reflects the importance of external habit formation. To elicit ω, the survey varies both others' and one's own past spending.[43]

Proposition 8. $\omega > \frac{\Delta C}{\Delta C + \Delta C_{others}}$ *if the respondent chooses Universe One over Universe Two for a better future experience in an external habit formation question.*

The 95% HPDI of the estimate of external habit mixture coefficient falls between 0 and 1 (Table 4), consistent with the existence of external habit formation.[44] The point estimate indicates that others' spending contributes to about 18% of one's own habit.

4.5 Relative Strength of Habit Formation and Peer Effect

To elicit the relative strength of habit formation and peer effect, I allow the possibility that other people's spending has a contemporaneous influence—peer effect—on one's own felicity function, $u(C, C_{\text{others}}, H)$, and then elicit $\frac{u_{C_{\text{others}}}}{u_H}$ by varying others' spending in both the past and the future.[45]

[42] See Section A of the online Appendix for proof.

[43] The resulting monthly spending graphs are in Figure A.15 of Section H of the supplemental material.

[44] Section F of the supplemental material elicits the existence of external habit formation without assuming any parametric habit evolution, as in the elicitation of the existence of internal habit formation. The evidence there is consistent with the evidence here—external habit formation exists.

[45] See Figure A.16 in Section H of the supplemental material for the resulting monthly spending graphs.

Proposition 9. *Under the first-order approximation,* $\frac{u_{C_{others}}}{u_H} < \frac{\omega}{\rho+\theta}\left(\rho\frac{\Delta C^{U2}_{others}}{\Delta C^{U1}_{others}} - \theta\right)$ *if the respondent chooses Universe One over Universe Two for a better future experience in a* $\frac{u_{C_{others}}}{u_H}$ *question.*

The point estimate for $\frac{u_{C_{\text{others}}}}{u_H}$ is about 1.03 (Table 4) and not significantly different from 1 at the 95% level, consistent with habit formation and peer effect having same-sized welfare impacts.

Two additional implications follow from the significant negative sign of $u_{C_{\text{others}}}$ as implied by the estimate and the previously estimated $u_H < 0$. First, peer effect exists separately from external habit. Because external habit and peer effect are accounted for separately in the elicitation, the fact that the estimate of $u_{C_{\text{others}}}$ is significantly negative means that peer effect exists after controlling for external habit. Second, peer effect is stronger than altruism. Note that the model do not restrict the sign of $u_{C_{\text{others}}}$ a priori, which can go both ways: altruism ($u_{C_{\text{others}}} > 0$) and peer effect ($u_{C_{\text{others}}} < 0$). Essentially, $u_{C_{\text{others}}}$ represents the net effect of these two phenomena. The significant negative sign of $u_{C_{\text{others}}}$, therefore, indicates that peer effect dominates altruism.

5 Explaining the Easterlin Paradox

The happiness–income paradox proposed by Easterlin states that income and happiness tend to be positively correlated in the short run and cross section but uncorrelated in the long run (Easterlin, 1973, 1974, 1995, 2017; Kaiser and Vendrik, 2019). Alternative views have been proposed: among others, that the U.S. data tend to be outliers (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 explores the explanation of the happiness–income pattern through the lens of habit formation and peer effect. To highlight the intuition of the explanation, the following discussion takes the view from a zero long-run gradient. Alternative views can be accommodated by slight changes of parameter values without changing the intuition.

Habit formation and peer effect have been the most popular potential explanations of the paradox. Recent evidence on peer effect (Luttmer, 2005; De Giorgi, Frederiksen, and Pistaferri, 2020) suggests that it is not powerful enough to fully explain the phenomenon.

To my knowledge, evidence on whether habit formation can help with the explanation is absent from the literature. Using the previous section's estimates on peer effect and habit formation of both the internal and external types, I show in this section that while each alone cannot generate the happiness–income pattern of the Easterlin paradox, together they can.

Four clarifications merit discussion before proceeding. The first is that this section focuses on the causal channel that income changes happiness. Typical life experiences and studies exploiting exogenous variations (Frijters, Haisken-DeNew, and Shields, 2004; Gardner and Oswald, 2007) support this view. Evidence aside, this causality motivated the discovery of the paradox[46] and is the most counterintuitive, interesting,[47] and policy relevant. Non-income happiness-altering factors do not help explain the paradox because they generally improve with income, making the long-run happiness–income relationship even more mysterious (Di Tella and MacCulloch, 2008). The second clarification is that, following the literature (Clark, Frijters, and Shields, 2008; Benjamin, Heffetz, Kimball, and Rees-Jones, 2012; Perez-Truglia, 2020), I assume that the potential distinction between happiness and utility is of minimal effect on the discussion below. Third, the paradox holds when income is replaced by consumption because consumption is closely related to income (Figure 7a), while happiness still has a long-term trend of about zero (Figure 7b). Because the paradox holds under either income and consumption, the following discussion uses them interchangeably.[48] 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 (Organisation for Economic Co-operation and Development, 2013), studied, and relevant to the paradox. I use instantaneous utility as a proxy for affect[49] and lifetime utility for life satisfaction.

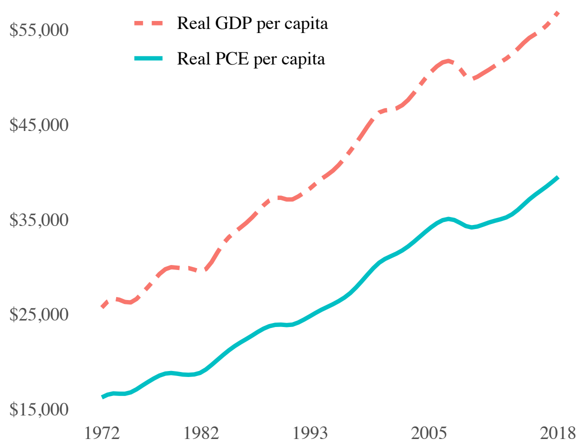
Because existing habit formation models are inconsistent with people's behavior, the credibility of any structural simulations under existing habit formation specifications is un-

[46]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).

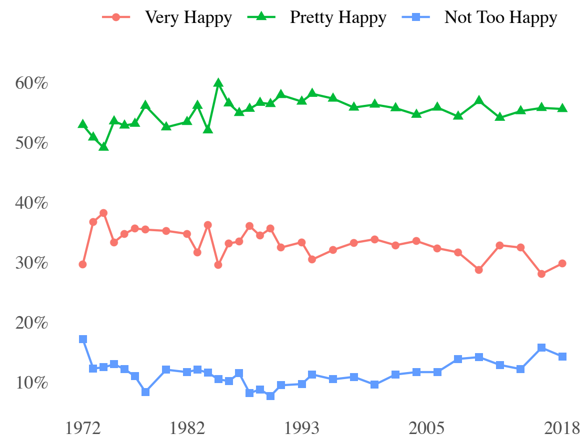
[47]This is evidenced by that the vast majority of speculative explanations of the paradox have focused on this channel.

[48]Compared with income, consumption relates more directly to human welfare, as is widely accepted in the economic literature. The relative lack of attention to the relationship between consumption and happiness is at least partly due to a relative lack of reliable micro-level panel data on total consumption.

[49]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 in survey questions measuring affect. Experiments with these two (and several other) time frames show trivial differences from no time integration.



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



(b) General happiness in the United States, 1972–2018. Survey response to the 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.

Figure 7: Income, consumption, and happiness in the United States.

tenable. To still assess the explanation of the paradox by habit formation and peer effect, I conduct semi-structural simulations based on the previous section's extensive evidence on these two phenomena. In particular, I specify that people are influenced by both internal and external habit formation as well as peer effect. Habit evolves according to equation (1) with the habit depreciation rate and the external habit mixture coefficient calibrated to their estimates, 1.07 and 0.18, respectively. Peer effect and external habit formation take effect only after others' spending changes become known to the agent, which is assumed to be k years after others' consumption changes.[50] When that happens, peer effect applies instantly, while external habit formation applies gradually in the way suggested by the micro evidence.

The effects of habit formation and peer effect on utility, to a first-order approximation, are captured by $\frac{u_H}{u_C}$ and $\frac{u_{C_{\text{others}}}}{u_C}$. The estimates of these two ratios are both greater than -1 at the 95% level (Table 4), which suggests that habit formation and peer effect, each alone, cannot fully explain the paradox.

The long-run nil happiness–income gradient dictates that

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

[50]The exact value of k does not matter for the intuition of the explanation. It only affects the speed at which the utility converges to its steady state.

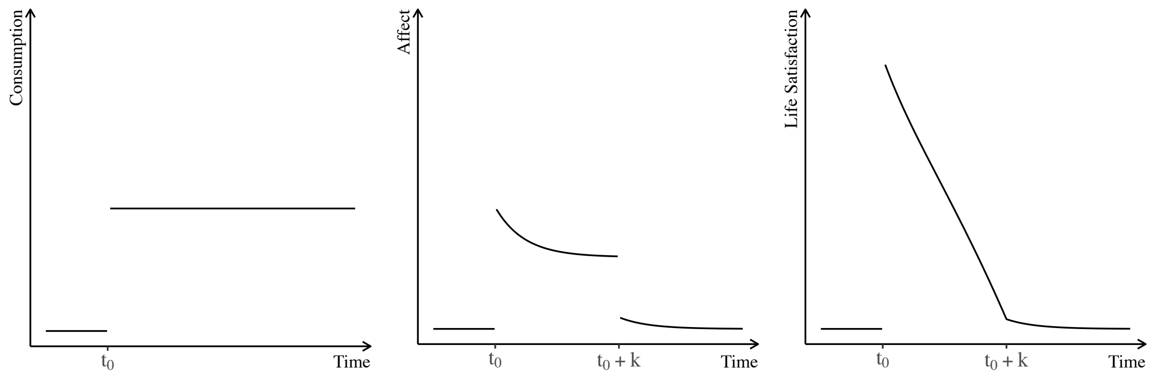
which is consistent with the estimate at the 95% level (Table 4). The point estimate of the left-hand side of the above equation is less than -1, which, aside from statistical precision considerations, provides the potential for the explanation of the paradox to be consistent with the general improvement of happiness-altering non-income factors (Di Tella and MacCulloch, 2008) and with the slightly negative long-run happiness–income slope in the United States (Stevenson and Wolfers, 2008; Firebaugh and Tach, 2012). For illustrative purposes, I focus on the scenario where the sum is -1. For concreteness, let us choose $\frac{u_H}{u_C} = \frac{u_{C_{\text{others}}}}{u_C} = -0.5$, both of which are within their respective 95% HPDIs and consistent with habit formation and peer effect having same-sized welfare impacts. 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, neither of which alters the happiness–income 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 peer effect entirely cancel the happiness effect of permanent consumption changes in the long run. To see this, imagine an economy was at a steady state where its residents were at some constant level of happiness before the instant t_0. Suppose that starting from t_0 onward the economy grows so that everyone's consumption permanently increases by a small amount of Δc (Figure 8a). As a result, to a first-order approximation,[51] the residents' happiness as measured by affect goes up by $u_C \Delta c$ at t_0. As time passes, the residents gradually get used to this higher level of self-spending, resulting in a buildup of internal habit that pulls affect down (Figure 8b). 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 peer comparison, which further pushes affect down. After that, external habit joins the play and, together with internal habit, erodes the remaining gain of affect until it completely disappears.

Integrating affect discounted by time preference,[52] one gets life satisfaction, 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 steady-state level (Figure 8c). For later reference, this pattern could be labeled as the wear-off effect: over time, habit formation and peer effect cancel out the happiness innovations brought by permanent consumption changes.

[51] This section focuses on first-order approximations because the elicitation of $u_{C_{\text{others}}}/u_C$ is under a first-order approximation. This is a good approximation when Δc is small, which is maintained here.

[52] I calibrate the time discount rate to 0.13, based on this study's estimate of this parameter. See Section G of the supplemental material for details. The value of this parameter does not affect the intuition.



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

Figure 8: One-episode growth.

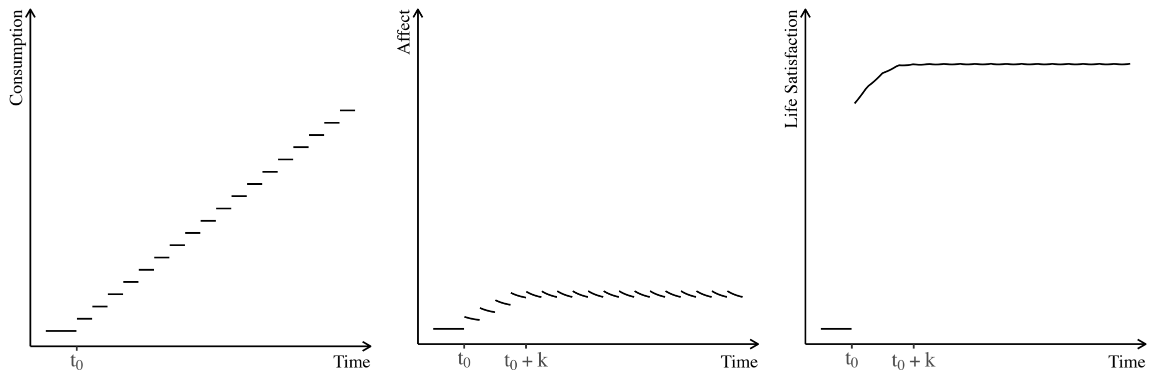
In reality, economies tend to grow over time, and, as a result, people typically earn more and consume more over time. To capture the key aspect of this phenomenon, suppose everyone's consumption increases permanently by Δc each year after t_0 (Figure 9a). Figures 9b and 9c plot the agent's happiness as time progresses. Unsurprisingly, habit formation wears off the gain of happiness within each year after t_0, as in the one-episode-growth scenario above.[53] What is new is the dynamics of happiness: instead of eventually returning to its previous steady-state level, happiness gradually builds up and then plateaus. For later reference, these two patterns of the happiness dynamics could be labeled as the transition effect and the plateau effect. The transition effect exists, contrasting with the decreasing trend of the one-episode-growth scenario, because in each year the annual growth of consumption brings a new episode of the wear-off effect whose initial happiness-enhancing phase[54] stacks onto those from previous years. Habit formation and peer effect gradually build up a happiness-reducing momentum that eventually cancels out the happiness-enhancing momentum that drives the transition effect, leading the agent to a happiness plateau. The instant when such exact cancellation first happens is precisely the moment when the wear-off effect brought by the consumption growth at t_0 is in full swing for the first time.

Because the wear-off effect is proportional to Δc,[55] the transition and plateau effects

[53]The discontinuities of the utility at the start of each year after t_0 result purely from the simplifying assumption that consumption permanently increases at the start of each year after t_0, which is inessential. When the consumption changes are smoother, the discontinuities will be reduced. All of this section's analysis carries through in such scenarios.

[54]The time interval when happiness is higher than its steady-state level in Figures 8b and 8c.

[55]This is a direct implication of 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 between the high



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

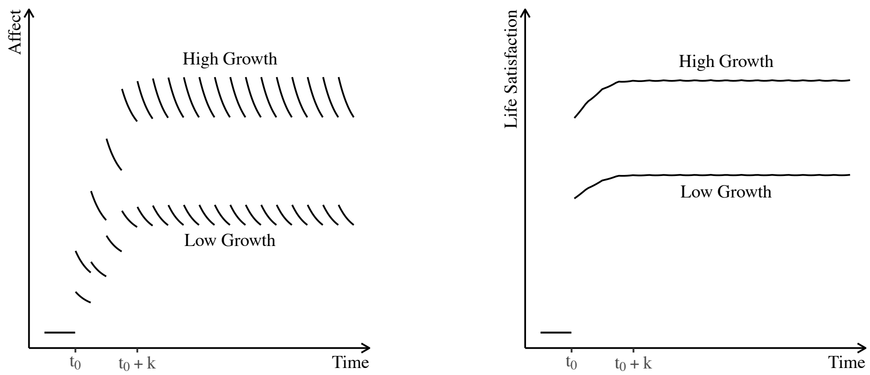
Figure 9: Multi-episode growth.

are also proportional to Δc (Figure 10). This could be labeled as the level effect—higher consumption growth leads to higher levels of happiness during both the transition and the plateau phases. The level effect predicts that faster-growing economies tend to enjoy larger increases in happiness. Frijters, Haisken-DeNew, and Shields's (2004) empirical evidence supports this prediction.

The level effect explains the positive cross-sectional correlation between income and happiness; higher income growth makes people or countries richer and places them on higher happiness curves. Economic fluctuations in 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 increase or decrease, the transition effect always implies a positive relationship between income and happiness. The plateau effect explains the long-run nil correlation between income and happiness. Even though income frequently fluctuates, it fluctuates around its trend. This trend growth determines the plateaued level of happiness, which governs the long-run trend of happiness. In other words, the long-run trend of the happiness curve flattens even though consumption and income keep growing, hence the nil correlation.

To deepen the intuition, it is helpful to look at an analogy that might be called "running 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 gradually accelerate to the speed of $\frac{u_H + u_{C_{\text{others}}}}{u_C}\Delta c = -\Delta c$ stairs per unit of time. Suppose the escalator is long enough so that it catches up to (the negative of) your

and low consumption changes will be smaller, the difference remains positive.



(a) Affect (b) Life satisfaction

Figure 10: Level effect. To highlight the effect on affect, the vertical axis in panel (a) is 15% of that in panel (b).

speed before you can reach the top. The elevation you reach represents happiness, and the (total) number of stairs you run represents consumption. The escalator symbolizes the joint effect of habit formation and peer effect.

With this analogy, it is illustrative to propose and resolve another paradox, the escalator paradox, which parallels the Easterlin paradox (Table 7). The escalator (Easterlin) paradox states that running more stairs (increasing income) raises elevation (happiness) in the cross section and short run but not in the long run. Why is this the case? In the long run, the escalator (habit formation and peer effect) eventually catches up to your running speed (consumption growth), after which the additional stairs you run (additional consumption you get) 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 canceling effect of habit formation and peer effect). 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 canceling effect of habit formation and peer effect) is larger during the transition phase, which accumulates to a higher level of elevation (happiness).

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, to the extent that people ultimately only care about happiness and that happiness eventually stops growing with economic growth, should we continue promoting economic growth after happiness plateaus? The answer implied by the explanation is yes. Happiness decreases if the econ-

Table 7: Twin Paradoxes

Dimension	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* raise **elevation**?	Transition effect (+ fluctuation)
Cross section	Why are *richer* people/countries **happier**?	Why are *faster* people **more elevated**?	Level effect

Note: Italics and boldface indicate parallelism of the twin paradoxes.

omy grows at slower speeds. In other words, economic growth initially raises 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.

6 Robustness

This section discusses the robustness of the estimates that underlie the results to demographics, time horizon, additional attention checks, and response biases and errors of nonzero and wave-varying mean.

6.1 Demographic Effects

The survey collects information on age, gender, household size, and household income of the respondents. Allowing the demographic variables to shift the means of the parameter distributions in the statistical model, I find that the demographics do not affect the estimates (Table 8). In particular, 0 is included in all of the 95% HPDIs of the estimated effects of demographic variables, except those of gender, household size and income on $\frac{Hu_{HH}}{u_H}$. After accounting for multiple hypothesis testing, these effects vanish.[56]

This result supports the view that the parameters the survey elicited are deep preference parameters that do not vary with demographic characteristics. Because the ratios of utility derivatives depend on the spending profiles in the survey, it is also reassuring that their estimates do not vary with the demographics of the respondents and, therefore, with their

[56]The adjusted probabilities of the estimates less than or equal to zero under the Holm algorithm are 0.90, 0.89, and 0.38, respectively.

Table 8: Demographic Effects on Parameter Estimates and Flexible Response Biases and Errors

	Demographic Effects					Response
	Omitted category	Age	Gender	Household size	Household income	bias and error (wave one)
Habit depreciation rate	1.20	0.00	-0.17	0.09	-0.04	-0.02
	[0.75, 1.64]	[-0.02, 0.02]	[-0.58, 0.27]	[-0.06, 0.24]	[-0.09, 0.02]	[-0.24, 0.22]
External habit mixture coefficient	0.19	0.00	0.03	-0.03	-0.01	-0.03
	[0.02, 0.36]	[-0.00, 0.01]	[-0.15, 0.20]	[-0.09, 0.04]	[-0.04, 0.01]	[-0.12, 0.07]
$-u_H/u_C$	0.64	0.00	-0.09	0.05	0.02	-0.02
	[0.41, 0.89]	[-0.01, 0.01]	[-0.32, 0.13]	[-0.03, 0.13]	[-0.01, 0.04]	[-0.11, 0.07]
Hu_{HH}/u_H	6.77	-0.01	1.39	-0.53	0.22	-0.91
	[5.40, 8.27]	[-0.06, 0.05]	[0.17, 1.50]	[-1.00, -0.06]	[0.03, 0.45]	[-3.11, 0.43]
u_{CH}/u_{HH}	-0.92	0.00	0.12	0.03	0.03	0.08
	[-1.22, -0.61]	[-0.01, 0.01]	[-0.17, 0.41]	[-0.08, 0.13]	[-0.01, 0.06]	[-0.08, 0.22]
u_{CC}/u_{HH}	4.38	0.03	-0.70	-0.36	-0.04	-0.34
	[3.05, 5.69]	[-0.02, 0.09]	[-1.50, 0.50]	[-0.87, 0.08]	[-0.23, 0.13]	[-1.16, 0.33]
$u_{C_{\text{others}}}/u_H$	0.71	-0.02	0.12	0.21	0.06	-0.11
	[-0.00, 1.50]	[-0.05, 0.01]	[-0.60, 0.84]	[-0.03, 0.48]	[-0.03, 0.15]	[-0.48, 0.28]

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

heterogeneous spending profiles in reality, for it implies that the 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 in the hypothetical situations.

6.2 Finite Horizon

The general model assumes an infinite horizon, as do almost all current habit formation models in the literature. To investigate the effect of this assumption on the results, I rederive all the elicitation propositions of the preference parameters under finite horizons and find that the changes are minimal: no change for the elicitation of some parameters and tiny changes for the rest.[57] As a result, estimation under the finite horizon[58] (column 1 of Table 9)

[57]The thresholds for the habit depreciation rate and external habit mixture coefficient are exactly the same in both time horizons. The changes to the thresholds of other parameters are simply replacing 1 with $1-e^{-\rho T}$, $1-e^{-(\rho+\theta)T}$, 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.

[58]The finite horizon is 30 years in the future, because the survey instructs the respondents: "If easier, think of ... 'Future' as the next 30 years."

Table 9: Robustness Estimates

		Additional attention checks				
	Finite horizon	No end-of-survey-quiz mistake	Consistent demographic information	Same past experience chosen	No polar response	Flexible response bias and error
	(1)	(2)	(3)	(4)	(5)	(6)
Habit depreciation rate	1.08 [0.89, 1.28]	1.10 [0.84, 1.34]	1.11 [0.83, 1.38]	1.16 [0.73, 1.57]	1.31 [0.77, 1.83]	1.11 [0.85, 1.34]
External habit mixture coefficient	0.18 [0.09, 0.27]	0.12 [0.00, 0.21]	0.13 [0.00, 0.24]	0.16 [0.00, 0.38]	0.09 [0.00, 0.34]	0.21 [0.08, 0.32]
$-u_H/u_C$	0.57 [0.46, 0.68]	0.65 [0.51, 0.80]	0.61 [0.45, 0.76]	0.51 [0.31, 0.70]	0.52 [0.29, 0.77]	0.60 [0.49, 0.72]
Hu_{HH}/u_H	7.45 [6.68, 8.48]	8.11 [6.94, 9.75]	8.47 [7.04, 10.20]	9.70 [7.51, 13.14]	9.01 [6.74, 12.72]	8.25 [6.83, 10.60]
u_{CH}/u_{HH}	-0.88 [-1.02, -0.72]	-0.82 [-1.00, -0.62]	-0.86 [-1.05, -0.67]	-0.74 [-0.94, -0.46]	-0.79 [-1.03, -0.53]	-0.91 [-1.07, -0.75]
u_{CC}/u_{HH}	3.84 [3.13, 4.48]	3.75 [2.79, 4.68]	3.39 [2.52, 4.41]	3.99 [2.48, 5.49]	3.14 [2.06, 4.14]	4.12 [3.07, 4.91]
$u_{C_{\text{others}}}/u_H$	1.15 [0.74, 1.47]	1.04 [0.60, 1.53]	0.94 [0.50, 1.45]	0.90 [0.22, 1.61]	0.46 [-0.06, 1.01]	1.12 [0.71, 1.54]
$u_{C_{\text{others}}}/u_C$	-0.64 [-0.89, -0.41]	-0.66 [-1.03, -0.36]	-0.54 [-0.93, -0.27]	-0.41 [-0.87, -0.09]	-0.22 [-0.58, 0.04]	-0.65 [-0.96, -0.40]
$u_{CH}u_H/u_{HH}u_C$	0.49 [0.37, 0.63]	0.53 [0.36, 0.71]	0.52 [0.34, 0.70]	0.36 [0.18, 0.55]	0.38 [0.19, 0.65]	0.54 [0.41, 0.69]
$u_{CH}u_C/u_{CC}u_H$	0.39 [0.29, 0.55]	0.33 [0.22, 0.51]	0.39 [0.25, 0.62]	0.34 [0.16, 0.65]	0.44 [0.21, 0.92]	0.37 [0.25, 0.53]
$Hu_Hu_{CH}/u_C\,(u_H + Hu_{HH})$	0.44 [0.33, 0.55]	0.47 [0.32, 0.63]	0.47 [0.31, 0.63]	0.32 [0.17, 0.50]	0.35 [0.17, 0.58]	0.48 [0.36, 0.62]
$Cu_Cu_{CH}/u_H\,(u_C + Cu_{CC})$	-0.25 [-0.31, -0.19]	-0.23 [-0.33, -0.16]	-0.25 [-0.38, -0.17]	-0.18 [-0.32, -0.10]	-0.25 [-0.43, -0.15]	-0.23 [-0.32, -0.17]
$\text{sgn}\,(Q_H)$		-1.00 [-1.00, -1.00]	-1.00 [-1.00, -1.00]	-1.00 [-1.00, -1.00]	-1.00 [-1.00, -1.00]	-1.00 [-1.00, -1.00]

Notes: Posterior mode above 95% HPDI/HPMI. Finite horizon does not affect the elicitation and estimate of $\text{sgn}\,(Q_H)$.

gives essentially identical estimates to the benchmark estimates under the infinite horizon.

6.3 Additional Attention Checks

In fielding the survey, explicit attention checks were used to screen out respondents who did not understand the hypothetical situation or the monthly spending graphs. In getting the sample for the benchmark estimation, the responses of those who sped through the survey, submitted duplicate responses, or were located outside of the United States are also deleted. This section makes use of implicit attention checks to see whether a potential lack of attention biases the estimates. Because they are not perfect proxies for attention, the implicit attention checks are applied them successively, from the relatively more reliable to the relatively less reliable.

Toward the end of the survey, the respondents were quizzed again on the basic hypothetical situation. There are 132 respondents in wave one and 53 in wave two who made at least one mistake in answering the five-question quiz. Deleting their responses from the sample does not significantly change the estimates (column 2 of Table 9).

The survey collected 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 wave consistency of answers to the demographic questions can serve as an implicit attention check. Applying this check eliminates another 18 responses from the remaining sample. The estimates are essentially unchanged (column 3 of Table 9).

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,[59] past spending is the same across the two universes, where the respondents should choose the same past experience. Deleting those who gave different answers shrinks the remaining sample by 97 and 13 responses in waves one and two, respectively. Even though the resulting HPDIs inflate because of the much smaller sample size, the estimates remain very close to the baseline estimates (column 4 of Table 9).

Finally, I use a measure of response consistency across the waves as an attention check. Considering that it involves more speculation, this attention check eliminates only those who gave at least one polar response—any response corresponding to the first (last) extreme range of parameter values in wave one and the last (first) in wave two. This check deletes another 34 responses from both waves, resulting in further expansion of the HPDIs, but, again, the estimates are not significantly different (column 5 of Table 9), and therefore the results remain robust.

6.4 Response Bias and Error of Nonzero and Wave-Varying Mean

The statistical model assumes a zero mean for the response bias and error across both waves of the survey. Relaxing this assumption, I arrive at a statistical model with response biases and errors of nonzero means that potentially vary across waves. Without loss of generality,[60]

[59] See Section G of the supplemental material.

[60] Only two means, one for each wave, can be identified. The specification here identifying the average and the difference of the means is equivalent to a specification that specifies the two means using two parameters,

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 biases and errors in wave one or equivalently the negative of the means of the response biases and errors in wave two are indistinguishable from zero at the 95% level (last column of Table 8). This implies that the parameter estimates are robust to the nonzero-wave-varying-mean response biases and errors, which is confirmed by column 6 of Table 9.

7 Conclusion

This paper have documented new and extensive micro evidence for habit formation through survey experiments. It finds that people's spending behavior exhibits habit formation. The majority of the habit forms internally, while a small fraction (about 18%) of the habit forms externally. This implies that in terms of micro validity, internal habit formation is a better choice than external habit formation.

Habit depreciates by about two-thirds per year. The parameter governing the speed of habit depreciation can significantly affect the performance of habit formation models.

Essentially all current habit formation models are rejected because their preference specifications fail to pass the four tests this paper have proposed. This justifies a search for habit formation preferences that match the survey evidence, which potentially can explain phenomena current habit formation models cannot.

Habit formation has a same-sized welfare impact as peer effect. Both external habit formation and peer effect exist in people's spending behavior. Peer effect dominates altruism.

Combining habit formation with peer effect can generate the happiness–income pattern highlighted by the Easterlin paradox. The mechanism suggests that happiness can increase with ever-growing income but only for a while before the wear-off effect induced by habit formation and peer effect puts happiness on a plateau. The level and transition effects explain the cross-sectional and short-run positive happiness–income gradients, whereas the plateau effect explains the long-run nil (or low) happiness–income gradient. Even though happiness eventually plateaus while income keeps growing, continued income growth is

one for each mean. If the two means are different, μ_ε should be significantly different from 0.

still necessary to maintain the plateaued level of happiness.

Future research could explore potential cross-country variations of the preference parameters, which might help explain the observed cross-country heterogeneities in happiness–income dynamics.

References

Abel, A. B. (1990): "Asset Prices Under Habit Formation and Catching Up With the Joneses," *American Economic Review*, 80(2), 38–42.

Adolfson, M., S. Laséen, J. Lindé, and M. Villani (2007): "Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through," *Journal of International Economics*, 72(2), 481–511.

Alessie, R., and F. Teppa (2010): "Saving and Habit Formation: Evidence from Dutch Panel Data," *Empirical Economics*, 38(2), 385–407.

Altig, D., L. J. Christiano, M. Eichenbaum, and J. Lindé (2011): "Firm-Specific Capital, Nominal Rigidities and the Business Cycle," *Review of Economic Dynamics*, 14(2), 225–247.

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

Ameriks, J., J. Briggs, A. Caplin, M. D. Shapiro, and C. Tonetti (2019): "Long-Term-Care Utility and Late-in-Life Saving," *Journal of Political Economy*.

Anderson, C. A., J. J. Allen, C. Plante, A. Quigley-McBride, A. Lovett, and J. N. Rokkum (2018): "The MTurkification of Social and Personality Psychology," *Personality and Social Psychology Bulletin*, 45(6), 842–850.

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

Benjamin, D. J., K. B. Cooper, O. Heffetz, and M. S. Kimball (2019): "A Well-Being Snapshot in a Changing World," *American Economic Review Papers and Proceedings*, 109, 344–49.

Benjamin, D. J., O. Heffetz, M. S. Kimball, and A. 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, D. J., O. Heffetz, M. S. Kimball, and N. Szembrot (2014): "Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference," *American Economic Review*, 104(9), 2698–2735.

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

Boldrin, M., L. J. Christiano, and J. D. Fisher (2001): "Habit Persistence, Asset Returns, and the Business Cycle," *American Economic Review*, pp. 149–166.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016): "Stereotypes," *Quarterly Journal of Economics*, 131(4), 1753–1794.

Browning, M., and M. D. Collado (2007): "Habits and Heterogeneity in Demands: A Panel Data Analysis," *Journal of Applied Econometrics*, 22(3), 625–640.

Campbell, J. Y., and J. 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, C. D., J. Overland, and D. N. Weil (2000): "Saving and Growth with Habit Formation," *American Economic Review*, pp. 341–355.

Chen, X., and S. C. Ludvigson (2009): "Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models," *Journal of Applied Econometrics*, 24(7), 1057–1093.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113(1), 1–45.

Clark, A. E., P. Frijters, and M. 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, J. H. (2017): "Macro-Finance," *Review of Finance*, 21(3), 945–985.

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

Crawford, I. (2010): "Habits Revealed," *Review of Economic Studies*, 77(4), 1382–1402.

De Giorgi, G., A. Frederiksen, and L. Pistaferri (2020): "Consumption Network Effects," *Review of Economic Studies*, 87(1), 130–163.

DellaVigna, S., A. Lindner, B. Reizer, and J. F. Schmieder (2017): "Reference-Dependent Job Search: Evidence from Hungary," *Quarterly Journal of Economics*, 132(4), 1969–2018.

Detemple, J. B., and F. Zapatero (1991): "Asset Prices in an Exchange Economy with Habit Formation," *Econometrica*, 59(6), 1633.

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

Dou, W. W., A. W. Lo, A. Muley, and H. Uhlig (2017): "Macroeconomic Models for Monetary Policy: A Critical Review from a Finance Perspective," *SSRN 2899842*.

Dynan, K. E. (2000): "Habit Formation in Consumer Preferences: Evidence from Panel Data," *American Economic Review*, pp. 391–406.

Easterlin, R. A. (1973): "Does Money Buy Happiness?," *Public Interest*, 30, 3.

——— (1974): "Does Economic Growth Improve the Human Lot? Some Empirical Evidence," in *Nations and Households in Economic Growth*, pp. 89–125. Elsevier.

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

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

Ferson, W. E., and G. M. Constantinides (1991): "Habit Persistence and Durability in Aggregate Consumption: Empirical Tests," *Journal of Financial Economics*, 29(2), 199–240.

Firebaugh, G., and L. Tach (2012): *Income, Age, and Happiness in America*pp. 267–287. Princeton University Press.

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

Frijters, P., J. P. Haisken-DeNew, and M. 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, J. C. (2000): "Habit Formation in Consumption and Its Implications for Monetary-Policy Models," *American Economic Review*, pp. 367–390.

GARDNER, J., AND A. J. OSWALD (2007): "Money and Mental Well-being: A Longitudinal Study of Medium-Sized Lottery Wins," *Journal of Health Economics*, 26(1), 49–60.

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

GUARIGLIA, A., AND M. ROSSI (2002): "Consumption, Habit Formation, and Precautionary Saving: Evidence from the British Household Panel Survey," *Oxford Economic Papers*, 54(1), 1–19.

HARA, K., A. ADAMS, K. MILLAND, S. SAVAGE, C. CALLISON-BURCH, AND J. P. BIGHAM (2018): "A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk," *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*.

HAVRANEK, T., M. RUSNAK, AND A. SOKOLOVA (2017): "Habit Formation in Consumption: A Meta-Analysis," *European Economic Review*, 95, 142–167.

HOUTHAKKER, H. S., AND L. D. TAYLOR (1970): *Consumer Demand in the United States*. Harvard University Press.

IWAMOTO, K. (2013): "Habit Formation in Household Consumption: Evidence from Japanese Panel Data," *Economics Bulletin*, 33(1), 323–333.

JOHNSON, D., AND J. RYAN (2018): "Amazon Mechanical Turk Workers Can Provide Consistent and Economically Meaningful Data," MPRA Paper 88450, University Library of Munich, Germany.

JOHNSTON, R. J., K. J. BOYLE, W. ADAMOWICZ, J. BENNETT, R. BROUWER, T. A. CAMERON, W. M. HANEMANN, N. HANLEY, M. RYAN, R. SCARPA, ET AL. (2017): "Contemporary Guidance for Stated Preference Studies," *Journal of the Association of Environmental and Resource Economists*, 4(2), 319–405.

KAHNEMAN, D., AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47(2), 263.

KAISER, C. F., AND M. C. M. VENDRIK (2019): "Different Versions of the Easterlin Paradox: New Evidence for European Countries," *Economics of Happiness*, pp. 27–55.

KAPTEYN, A., AND F. TEPPA (2003): "Hypothetical Intertemporal Consumption Choices," *Economic Journal*, 113(486), C140–C152.

KHANAL, A. R., A. K. MISHRA, AND S. NEDUMARAN (2018): "Consumption, Habit Formation, and Savings: Evidence from a Rural Household Panel Survey," *Review of Development Economics*, 23(1), 256–274.

KIMBALL, M. S., F. OHTAKE, D. H. RECK, Y. TSUTSUI, AND F. ZHANG (2015): "Diminishing Marginal Utility Revisited," *SSRN 2592935*.

KIMBALL, M. S., C. R. SAHM, AND M. D. SHAPIRO (2008): "Imputing Risk Tolerance from Survey Responses," *Journal of the American Statistical Association*, 103(483), 1028–1038.

——— (2009): "Risk Preferences in the PSID: Individual Imputations and Family Covariation," *American Economic Review Papers and Proceedings*, 99(2), 363–368.

KIMBALL, M. S., AND M. D. SHAPIRO (2008): "Labor Supply: Are the Income and Substitution Effects Both Large or Both Small?," Working Paper 14208, National Bureau of Economic Research.

KIMBALL, M. S., AND R. WILLIS (2006): "Happiness and Utility," Unpublished.

Korniotis, G. M. (2010): "Estimating Panel Models with Internal and External Habit Formation," *Journal of Business and Economic Statistics*, 28(1), 145–158.

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

Ljungqvist, L., and H. Uhlig (2000): "Tax Policy and Aggregate Demand Management Under Catching Up with the Joneses," *American Economic Review*, 90(3), 356–366.

——— (2015): "Comment on the Campbell-Cochrane Habit Model," *Journal of Political Economy*, 123(5), 1201–1213.

Lubik, T. A., and F. Schorfheide (2004): "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, 94(1), 190–217.

Lupton, J. P. (2002): "The Habit Liability in Life-Cycle Consumption and Portfolio Choice," *Dissertation, University of Michigan.*

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

Naik, N. Y., and M. J. Moore (1996): "Habit Formation and Intertemporal Substitution in Individual Food Consumption," *Review of Economics and Statistics*, 78(2), 321–328.

Organisation for Economic Co-operation and Development (2013): *OECD Guidelines on Measuring Subjective Well-being*. OECD Publishing.

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

Perez-Truglia, R. (2020): "The Effects of Income Transparency on Well-Being: Evidence from a Natural Experiment," *American Economic Review*, 110(4), 1019–54.

Ravina, E. (2019): "Habit Formation and Keeping Up With the Joneses: Evidence From Micro Data," *SSRN 928248*.

Ravn, M., S. Schmitt-Grohé, and M. Uribe (2006): "Deep Habits," *Review of Economic Studies*, 73(1), 195–218.

Rayo, L., and G. S. Becker (2007): "Habits, Peers, and Happiness: An Evolutionary Perspective," *American Economic Review*, 97(2), 487–491.

Rhee, W. (2004): "Habit Formation and Precautionary Saving: Evidence from the Korean Household Panel Studies," *Journal of Economic Development*, 29(2), 1–19.

Sacks, D. W., B. Stevenson, and J. Wolfers (2012): "The New Stylized Facts About Income and Subjective Well-Being.," *Emotion*, 12(6), 1181.

Sahm, C. R. (2007): "Stability of Risk Preference," Finance and Economics Discussion Series 2007-66, Board of Governors of the Federal Reserve System (U.S.).

Samuelson, P. A. (1950): "The Problem of Integrability in Utility Theory," *Economica*, 17(68), 355.

Smets, F., and R. Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1(5), 1123–1175.

——— (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97(3), 586–606.

Stevenson, B., and J. Wolfers (2008): "Economic Growth and Subjective Well-Being: Reassessing the

Easterlin Paradox," *Brookings Papers on Economic Activity*, 2008(1), 1–87.

Stock, J. H., and J. H. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68(5), 1055–1096.

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

Van Praag, B. M. S., and P. Frijters (1999): *The Measurement of Welfare and Well-Being: The Leyden Approach*pp. 413–433. Russell Sage Foundation.

Online Appendix for "Survey Evidence on Habit Formation"

Jiannan Zhou

This appendix contains the proof of the observational equivalence of linear and nonlinear habit evolutions under general habit formation preferences, aggregation of the preference parameters, response distribution, and proofs of the elicitation propositions.

A Observational Equivalence of Linear and Nonlinear Habit Evolutions Under General Habit Formation Preferences

This section shows that the model with the linear habit evolution (Model L below) and the models with nonlinear habit evolutions (Model N below) are observationally equivalent (in the sense of Definition 1 below) by a monotonic transformation of the scale on which habit is measured.

- Model L: $\mathbb{E}_0 \int_0^\infty e^{-\rho t} u(C, H)\, dt$ s.t. $\dot{H} = \theta (C - H)$
- Model N: $\mathbb{E}_0 \int_0^\infty e^{-\rho t} v(C, \mathcal{H})\, dt$ s.t. $\dot{\mathcal{H}} = f(C, \mathcal{H})$, where f can be a nonlinear function of C and $\mathcal{H}$.

Note that $H_t = h(C_0, H_0, t)$ if $C_t = \bar{C}$ for $t \geq 0$ where the subscripts index time. Similarly, $\mathcal{H}_t = k(C_0, \mathcal{H}_0, t)$ if $C_t = \bar{C}$ for $t \geq 0$. That is, if consumption does not change for $t \geq 0$, 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) ceteris paribus, when one function is monotonic in the variable, the other function is also monotonic in the variable.

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 are entangled with respect to time.[1]

[1] That is, $\dot{H}_t \cdot \dot{\mathcal{H}}_t$ does not change its sign. It is possible that $\dot{H}_t \cdot \dot{\mathcal{H}}_t = 0$ in some time intervals, but $\dot{H}_t \cdot \dot{\mathcal{H}}_t$ does not change its sign around the intervals.

Proposition 10. *Model L and Model N are observationally equivalent if the monotonicities of H and $\mathcal{H}$ are entangled with respect to time.*

Proof. Suppose that consumption changes at instant 0 and stays at that level afterwards: $C_t = C_{t+\varepsilon} \neq C_{-\varepsilon}\ \forall\ t \geq 0$ and $\varepsilon > 0$. Without loss of generality, suppose also that habit reaches its new steady state at instant T. Because H and $\mathcal{H}$ are entangled monotonically with respect to time, H and $\mathcal{H}$ are monotonic from instant 0 to instant T and flat afterward (i.e., remain at constant levels), say at levels $\bar{H}$ and $\bar{\mathcal{H}}$. That is, $H_t = a\left(t|C_0, H_0\right)$ 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 $t > T$.

Because

$$
\begin{aligned}
\mathcal{H}_t &= b\left(t|C_0, \mathcal{H}_0\right) \\
&= b\left(a^{-1}\left(a\left(t|C_0, H_0\right)\right)|C_0, \mathcal{H}_0\right) \\
&= b\left(a^{-1}\left(H_t|C_0, H_0\right)|C_0, \mathcal{H}_0\right)
\end{aligned}
$$

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 bijective function G that maps H_t into $\mathcal{H}_t$:

$$
\mathcal{H}_t = G\left(H_t\right) \equiv \begin{cases} b\left(a^{-1}\left(H_t|C_0, H_0\right)|C_0, \mathcal{H}_0\right) & 0 \leq t \leq T \\ \frac{\bar{\mathcal{H}}}{\bar{H}} H_t & t > T. \end{cases}
$$

Because $v\left(C, \mathcal{H}\right) = v\left(C, G\left(H\right)\right) \equiv u\left(C, H\right)$, Model N gives the same utility as Model L for any consumption path that is constant for $t \geq 0$.

When the consumption path is not constant for $t \geq 0$, the utilities from the two models remain equal. To see this, start from the instant 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 instant onward. Then go back to the instant 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 instant of interest.

Because the utilities from the two models are the same, the consumption choices generated from these two models coincide. Proof by contradiction: suppose that the two models lead to different optimal consumption paths—$\left\{C_L^*\right\} \neq \left\{C_N^*\right\}$ 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 $U\left(\left\{C_L^*\right\}, H_0\right) \neq V\left(\left\{C_N^*\right\}, \mathcal{H}_0\right)$, at least one of the two consumption paths is not maximizing lifetime utility, contradicting that they are optimal solutions to the respective models. If $U\left(\left\{C_L^*\right\}, H_0\right) = V\left(\left\{C_N^*\right\}, \mathcal{H}_0\right)$, the consumption path $\left\{C_L^*\right\}$ is also a solution to the Model N while $\left\{C_N^*\right\}$ also a solution to Model L. Therefore, $\left\{C_H^*\right\}$ and $\left\{C_{\mathcal{H}}^*\right\}$ 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. $Q.E.D.$

Because the monotonicities of H and $\mathcal{H}$ are entangled with respect to time, by Proposition 10, Model L and Model N are observationally equivalent.

One can easily allow external habit formation and peer effect in the models; the proof of the equivalence result in these situations is straightforward because other people's spending is exogenous to the agent.

B Aggregation

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

To aggregate individuals, their welfare 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 $\bar{C}_i = \bar{H}_i = \varphi \; \forall i$, spending an extra dollar[2] while holding habit constant, brings the same marginal utility to every individual: $u_{i,C}\left(\bar{C}_i, \bar{H}_i\right) = u_{j,C}\left(\bar{C}_j, \bar{H}_j\right) = \bar{u} \; \forall i, j$.

With the comparability of the individuals' utilities, the representativeness of the representative agent (R) implies that $N u_R\left(C_R, H_R\right) = \sum_i u_i\left(C_i, H_i\right)$ when $C_R = C_i = A$ and

[2] An epsilon dollar, to be exact.

$H_R = H_i = B \; \forall i$ 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-agent model behaves, and hence the equality of $N u_R (C_R, H_R) = \sum_i u_i (C_i, H_i)$ under $C_R = C_i$, $H_R = H_i$, and $u_R = u_i \; \forall i$. Now, allowing individuals to be heterogeneous along the dimension of utility function after the normalization of the comparability condition, the representativeness condition simply requires that the representative agent represents the individuals along the welfare dimension, after controlling for consumption and habit.

B.1 Aggregation of Habit Depreciation Rate

Even though habit depreciation rate (θ) and habit (H) are one-to-one mapped at each instant of time[3] for any given consumption profile, there are infinitely many pairs of them that satisfy at the same time the representative agent's habit evolution ($\dot{H}_R = \theta_R (C_R - H_R)$), the individuals' habit evolutions ($\dot{H}_i = \theta_i (C_i - H_i)$), the comparability condition, and the representativeness condition. The intuition is that while habit depends on the habit depreciation rate, its steady-state level does not. In other words, a different value of θ leads to a different value of H at all the instants before the steady state, and this difference vanishes after the steady state is reached. The representativeness condition eliminates such indeterminacy and pins down a unique θ_R for individuals' given θ_i's.

To find the mapping between the aggregate habit depreciation rate (θ_R) and the individuals' habit depreciation rates (θ_i's), imagine that everyone starts at the homogenous steady state and increases their 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 resulting changes to the utilities are

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

[3] Before a steady state is reached.

and

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

where $u_{i,X} \equiv \partial u_i / \partial X$.

By $N \Delta u_R\left(C_R, H_R\right) = \sum_i \Delta u_i\left(C_i, H_i\right)$, as implied by the representativeness condition,

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

Because $\Delta C_R = \Delta C_i$ and $u_{R,C}\left(C_R, H_R\right) = u_{i,C}\left(C_i, H_i\right)\ \forall i$ (see Section B.2),

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

where the second equality holds because of the independence between slope of indifference 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 B.2) and $u_{R,C}\left(C_R, H_R\right) = u_{i,C}\left(C_i, H_i\right)\ \forall i$, it follows that

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

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

B.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 homogenous steady state ($\bar{C}_R = \bar{H}_R = \bar{C}_i = \bar{H}_i\ \forall i$).

Because $N u_R\left(C_R, H_R\right) = \sum_i u_i\left(C_i, H_i\right)$ 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 are the average of the utility derivatives of the individuals:

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

where X denotes the variable and order of differentiation 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 steady state.

1. Under the normalization of $u_{R,C}\left(\bar{C}_R, \bar{H}_R\right) = u_{i,C}\left(\bar{C}_i, \bar{H}_i\right) = \bar{u}\ \forall i$,

$$\frac{1}{N}\sum_i u_{i,H} = -\frac{1}{N}\sum_i\left(-\frac{u_{i,H}}{u_{i,C}}\cdot u_{i,C}\right) = -\left[\frac{1}{N}\sum_i\left(-\frac{u_{i,H}}{u_{i,C}}\right)\right]\cdot u_{i,C} \equiv -\mu_{-\frac{u_{i,H}}{u_{i,C}}}\cdot\bar{u},$$

   where $\mu_{-\frac{u_{i,H}}{u_{i,C}}} \equiv \frac{1}{N}\sum_i\left(-\frac{u_{i,H}}{u_{i,C}}\right)$. Similar notations are used hereafter.

2. Because the parameters are independent,

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

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

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

and

$$\frac{1}{N}\sum_i u_{i,C_{\text{others}}} = \frac{1}{N}\sum_i \frac{u_{i,C_{\text{others}}}}{u_{i,H}} \cdot u_{i,H}$$
$$= \frac{1}{N}\sum_i \frac{u_{i,C_{\text{others}}}}{u_{i,H}} \cdot \frac{1}{N}\sum_i u_{i,H}$$
$$= \mu_{\frac{u_{i,C_{\text{others}}}}{u_{i,H}}} \cdot \left(-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\right) \cdot \bar{u}.$$

3. With these, the representative agent's parameters can be calculated:

$$-\frac{u_{R,H}}{u_{R,C}} = -\frac{\frac{1}{N}\sum_i u_{i,H}}{\frac{1}{N}\sum_i u_{i,C}} = -\frac{\frac{1}{N}\sum_i u_{i,H}}{\bar{u}} = \mu_{-\frac{u_{i,H}}{u_{i,C}}},$$

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

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

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

and

$$\frac{u_{R,C_{\text{others}}}}{u_{R,H}} = \frac{\frac{1}{N}\sum_i u_{i,C_{\text{others}}}}{\frac{1}{N}\sum_i u_{i,H}} = \frac{\mu_{\frac{u_{i,C_{\text{others}}}}{u_{i,C}}} \cdot \left(-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\right) \cdot \bar{u}}{\left(-\mu_{-\frac{u_{i,H}}{u_{i,C}}}\right) \cdot \bar{u}} = \mu_{\frac{u_{i,C_{\text{others}}}}{u_{i,H}}}.$$

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

B.3 Aggregation of External Habit Mixture Coefficient

Imagine that everyone's consumption increases by the same iota amount so that the representative agent also increases her consumption by this same amount. The representativeness condition implies that the changes in utilities satisfy

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

Using 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 $\Delta H_R / \Delta H_i = \theta_R \left(1-\omega_R\right) / \left[\theta_i \left(1-\omega_i\right)\right]$,

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

where the second equality holds because of the independence between the preference parameters. Finally, by

$$
\frac{u_{R,H}\left(C_R, H_R\right)}{u_{R,C}\left(C_R, H_R\right)} = \frac{1}{N} \sum_i \frac{u_{i,H}\left(C_i, H_i\right)}{u_{i,C}\left(C_i, H_i\right)}
$$

and $\theta_R = \frac{1}{N} \sum_i \theta_i$,

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

In words, the representative agent's external habit mixture coefficient equals the average of individuals' external habit mixture coefficients.

C Response Distribution

Table A.1 summarizes the distributions of responses to survey questions eliciting the preference parameters identifiable to scale. Response distribution of parameters identifiable only to sign are reported at the places where their estimates are reported.

Table A.1: Response Distributions (Percentage) for Preference Parameters Identifiable to Scale

Question	Wave	Response					
		U1U1U1	U1U1U2	U1U2	U2U1	U2U2U1	U2U2U2
Habit depreciation rate	1	28	9	17	11	6	28
	2	29	10	14	11	6	30
External habit mixture coefficient	1	17	5	9	14	9	46
	2	24	5	16	6	4	46
$-u_H/u_C$	1	33	4	7	7	12	38
	2	32	4	6	2	20	36
Hu_{HH}/u_H	1	14	5	8	11	3	59
	2	24	2	6	5	1	61
u_{CH}/u_{HH}	1	7	4	12	9	28	40
	2	9	3	10	9	34	35
u_{CC}/u_{HH}	1	23	30	11	10	5	21
	2	24	19	8	10	9	30
$u_{C_{\text{others}}}/u_H$	1	26	18	10	8	3	36
	2	24	19	8	6	3	40

Notes: U1 and U2 stand for Universe One and Universe Two, respectively. U1U1U2 denotes the response sequence of first choosing U1, then U1 again in the first follow-up question, and finally U2 in the second follow-up question. Similar notations are used to denote other response sequences.

D Proofs of Elicitation Propositions

Before proving the elicitation propositions, this section first proves three lemmas and then derives three quantities, which will be used repeatedly in the proofs of the propositions.

D.1 Lemmas

Lemma 1. *For $a, b, c \in \mathbb{R}$, if $a(a+b) > 0$ and $0 \le c \le 1$, $a(a+cb) > 0$.*

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 $b \ge 0$, $(c-1)b \le 0$ and thus $a+cb \le a+b < 0$. If $b < 0$, by $a < 0$ and $c \ge 0$, $a+cb < 0$. Therefore, $a(a+cb) > 0$.

Suppose $a > 0$ and $a+b > 0$. If $b \le 0$, $(c-1)b \ge 0$ and, therefore, $a+cb \ge a+b > 0$. If $b > 0$, by $a > 0$ and $c \ge 0$, $a+cb > 0$. In both cases, $a(a+cb) > 0$. *Q.E.D.*

In words, the lemma states that if a and $a+b$ share a sign and $0 \le c \le 1$, $a+cb$ share its sign with a and $a+b$.

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

Proof. $\sum_{s=0}^{n} (\Delta e)^s (\Delta f)^{n-s} / M^n$ is decreasing in $n \in \mathbb{N}^+$ if $\forall n \in \mathbb{N}^+$,

$$\frac{\sum_{s=0}^{n} (\Delta e)^s (\Delta f)^{n-s}}{M^n} - \frac{\sum_{s=0}^{n+1} (\Delta e)^s (\Delta f)^{n+1-s}}{M^{n+1}} > 0.$$

Now,

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

where the first inequality holds because $M - \Delta e - \Delta f \geq 0$ and $\Delta e > 0$. *Q.E.D.*

Lemma 3. *For $\Delta e, \Delta f, M, \theta, t \in \mathbb{R}^+$, if $M - \Delta e - \Delta f \geq 0$, $u(C, H)$ is analytic with $u_H < 0 \,\forall H$, and $\partial^{i_k} u / \partial H^{i_k} \leq \Lambda_k \,\forall i_k \in I \equiv \left\{i \middle| \partial^i u / \partial H^i > 0, i \in \mathbb{N}^+\right\}$ with $\partial^{i_1} u / \partial H^{i_1} < \Lambda_1$, where i_k is the k-th smallest element of I,*

$$\Lambda_k \equiv -\sum_{n=1}^{i_k} e^{(i_k - n)\theta t} \frac{i_k! \sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s}}{n! \sum_{s=0}^{i_k-1} (\Delta e)^s (\Delta f)^{i_k-1-s}} \left[\frac{\partial^n u}{\partial H^n} 1(n \notin I) + \Lambda_{K(n)} 1(n \in I)\right],$$

and $K(n) = k$ *if* $n = i_k$, *then*

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

Proof. By $u_H < 0 \; \forall H$, analyticity of $u(C, H)$, and $Me^{-\theta t} > 0$,

$$u\left(C, H + Me^{-\theta t}\right) - u(C, H) = \sum_{n=1}^{\infty} \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(Me^{-\theta t}\right)^n < 0. \tag{A.1}$$

Because Δe, Δf, $M > 0$, and $M - \Delta e - \Delta f \geq 0$, $0 < \frac{\Delta e + \Delta f}{M} \leq 1$.

By $u_H < 0$, inequality (A.1), and Lemma 1,

$$u_H Me^{-\theta t} + \frac{\Delta e + \Delta f}{M} \sum_{n=2}^{\infty} \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(Me^{-\theta t}\right)^n < 0. \tag{A.2}$$

By Lemma 2,

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

If $u_{HH} \leq 0$, apply Lemma 1 to inequality (A.2) to get

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

This process of successive applications of Lemmas 1 and 2 can continue until $\partial^{i_1} u / \partial H^{i_1} > 0$. Note that $i_1 \geq 2$ since $u_H < 0$.

In general, when $\partial^i u / \partial H^i > 0$ for $i \in I$, it is necessary to bound the $\partial^i u / \partial H^i$'s from above, so that

$$\sum_{n=1}^{i} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} < 0,$$

enabling the continued applications of Lemmas 1 and 2. Next, I show that the bounds of $\Lambda_{\{k\}}$ achieve this goal.[4]

Suppose, for some $i_k \in I$,

$$\sum_{n=1}^{z-1} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) < 0 \tag{A.4}$$

$\forall\, z \leq i_k$.

By inequalities (A.1) and (A.4), Lemmas 1 and 2 can be applied to all $n \leq i_k$ in the fashion inequality (A.3) is derived from inequality (A.1), which gives

$$\begin{aligned} &\sum_{n=1}^{i_k-1} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) \\ &+ \frac{\frac{\sum_{s=0}^{i_k-1} (\Delta e)^s (\Delta f)^{i_k-1-s}}{M^{i_k-1}}}{\frac{\sum_{s=0}^{i_k-2} (\Delta e)^s (\Delta f)^{i_k-2-s}}{M^{i_k-2}}} \frac{\sum_{s=0}^{i_k-2} (\Delta e)^s (\Delta f)^{i_k-2-s}}{M^{i_k-2}} \sum_{i=i_k}^{\infty} \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(M e^{-\theta t} \right)^n \\ &< 0. \end{aligned} \tag{A.5}$$

Since

$$0 \leq \frac{\frac{\sum_{s=0}^{i_k-1} (\Delta e)^s (\Delta f)^{i_k-1-s}}{M^{i_k-1}}}{\frac{\sum_{s=0}^{i_k-2} (\Delta e)^s (\Delta f)^{i_k-2-s}}{M^{i_k-2}}} \leq 1,$$

as implied by Lemma 2, and inequality (A.4) with $z = i_k$, Lemma 1 can be applied to inequality (A.5) for getting

$$\begin{aligned} &\sum_{n=1}^{i_k-1} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) \\ &+ \frac{\frac{\sum_{s=0}^{i_k-1} (\Delta e)^s (\Delta f)^{i_k-1-s}}{M^{i_k-1}}}{\frac{\sum_{s=0}^{i_k-2} (\Delta e)^s (\Delta f)^{i_k-2-s}}{M^{i_k-2}}} \frac{\sum_{s=0}^{i_k-2} (\Delta e)^s (\Delta f)^{i_k-2-s}}{M^{i_k-2}} \sum_{i=i_k}^{\infty} \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(M e^{-\theta t} \right)^n \end{aligned}$$

[4]As mentioned in the paper, the ubiquitous additive and multiplicative habits with power utility satisfy these bounds under common parameter values.

$$
\begin{aligned}
&= \sum_{n=1}^{i_k} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) \\
&\quad + \frac{\sum_{s=0}^{i_k-1} (\Delta e)^s (\Delta f)^{i_k-1-s}}{M^{i_k-1}} \sum_{n=i_k+1}^{\infty} \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(M e^{-\theta t} \right)^n \\
&< 0.
\end{aligned}
$$

Now,

$$
\begin{aligned}
&\sum_{n=1}^{i_k} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) \\
&= \sum_{n=1}^{i_k-1} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) + e^{-i_k\theta t} M \frac{1}{i_k!} \frac{\partial^{i_k} u}{\partial H^{i_k}} \left(\sum_{s=0}^{i_k-1} (\Delta e)^s (\Delta f)^{i_k-1-s} \right) \\
&\leq \sum_{n=1}^{i_k-1} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) + e^{-i_k\theta t} M \frac{1}{i_k!} \left(\sum_{s=0}^{i_k} (\Delta e)^s (\Delta f)^{i_k-1-s} \right) \\
&\cdot \left(- \sum_{n=1}^{i_k} e^{(i_k-n)\theta t} \frac{i_k! \sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s}}{n! \sum_{s=0}^{i_k-1} (\Delta e)^s (\Delta f)^{i_k-1-s}} \left[\frac{\partial^n u}{\partial H^n} 1\,(n \notin I) + \Lambda_{K(n)} 1\,(n \in I) \right] \right) \\
&= \sum_{n=1}^{i_k-1} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) \\
&\quad - \sum_{n=1}^{i_k} e^{-n\theta t} M \frac{1}{n!} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) \left[\frac{\partial^n u}{\partial H^n} 1\,(n \notin I) + \Lambda_{K(n)} 1\,(n \in I) \right] \\
&= \sum_{n=1}^{i_k-1} 1\,(n \in I)\, e^{-n\theta t} M \frac{1}{n!} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) \left(\frac{\partial^n u}{\partial H^n} - \Lambda_{K(n)} \right) \\
&< 0
\end{aligned} \tag{A.6}
$$

where the first inequality holds because $\partial^{i_k} u/\partial H^{i_k} \leq \Lambda_k$ and the last inequality because $\partial^i u/\partial H^i \leq \Lambda_{K(i)} \forall i \in I$ and $\partial^{i_1} u/\partial H^{i_1} < \Lambda_1$. That is, inequality (A.4) holds for $z = i_k + 1$.

Since $\partial^n u/\partial H^n < 0$ for $i_k < n < i_{k+1}$, inequality (A.4) also holds $\forall z \leq i_{k+1}$. In words, if inequality (A.4) holds $\forall z \leq i_k$, it also holds for $z \leq i_{k+1}$. Since it is trivially true that inequality (A.4) holds $\forall z \leq i_1$ (note $\partial^n u/\partial H^n < 0$ for $n < i_1$), inequality (A.4) holds $\forall z \leq i \;\forall i \in I$.

In particular, it holds for the largest element of I ($i_{|I|}$[5]). Replacing k by $|I|$ in the step

[5] $|I|$ denotes the cardinality of set I.

of inequality (A.4) and following the subsequent derivation to inequality (A.6) leads to

$$\sum_{n=1}^{i_{|I|}} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) < 0. \tag{A.7}$$

Because $\partial^n u / \partial H^n < 0 \ \forall n > i_{|I|}$,

$$\sum_{n=i_{|I|}+1}^{\infty} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) < 0. \tag{A.8}$$

Finally,

$$\begin{aligned}
&\sum_{n=1}^{\infty} e^{-n\theta t} \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) \\
=&\frac{1}{M} \left[\sum_{n=1}^{i_{|I|}} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) \right. \\
&\left. + \sum_{n=i_{|I|}+1}^{\infty} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) \right] \\
<&0,
\end{aligned}$$

where the inequality follows from inequalities (A.7) and (A.8). *Q.E.D.*

D.2 Three Quantities

This section presents the derivations of three quantities regarding the utility differences between spending paths, which will be used repeatedly in proving the propositions.

First, note that $\dot{H} = \theta (C - H)$ leads to

$$H(t) = e^{-\theta(t-t_0)} H(t_0) + \int_{t_0}^{t} e^{-\theta(t-s)} \theta C(s)\, ds.$$

Second, denote a steady state by $C(t) = H(t) = \bar{C} = \bar{H} \ \forall t$. Then

$$\begin{aligned}
\Delta H(t) &\equiv H(t) - \bar{H} \\
&= e^{-\theta(t-t_0)} \left(H(t_0) - \bar{H} \right) + \int_{t_0}^{t} e^{-\theta(t-s)} \theta \left(C(s) - \bar{H} \right) ds
\end{aligned}$$

$$
\begin{aligned}
&= e^{-\theta(t-t_0)}\left(H\left(t_0\right)-\bar{H}\right)+\int_{t_0}^{t} e^{-\theta(t-s)}\theta\left(C\left(s\right)-\bar{C}\right)ds \\
&\equiv e^{-\theta(t-t_0)}\Delta H\left(t_0\right)+\int_{t_0}^{t} e^{-\theta(t-s)}\theta\Delta C\left(s\right)ds
\end{aligned}
$$

where $\Delta C\left(t\right)\equiv C\left(t\right)-\bar{C}$.

The first quantity is for the class of spending paths with $\Delta C\left(t\right)=A\cdot 1\left(t\leq 0\right)+B\cdot 1\left(t>0\right)$ (that is, $\Delta C\left(t\right)=A$ for $t\leq 0$ and $\Delta C\left(t\right)=B$ for $t>0$). Under such spending paths, for $t\geq 0$,

$$
\begin{aligned}
\Delta H\left(t\right) &= e^{-\theta\infty}\Delta H\left(-\infty\right)+\int_{-\infty}^{0} e^{-\theta(t-s)}\theta A ds+\int_{0}^{t} e^{-\theta(t-s)}\theta B ds \\
&= e^{-\theta t}A+\left(1-e^{-\theta t}\right)B. \qquad \text{(A.9)}
\end{aligned}
$$

For the second equality, note that $\Delta H\left(-\infty\right)$ is finite by the definition of habit.

The difference between lifetime utilities starting now ($t=0$), $\int_0^\infty e^{-\rho t}u\left(C\left(t\right),H\left(t\right)\right)dt$, under this class of spending paths and under the steady state $\left(\bar{C},\bar{H}\right)$ is

$$
\begin{aligned}
&\Psi\left(A,B\right) \\
&\equiv \int_0^\infty e^{-\rho t}u\left(C\left(t\right),H\left(t\right)\right)dt-\int_0^\infty e^{-\rho t}u\left(\bar{C},\bar{H}\right)dt \\
&= \int_0^\infty e^{-\rho t}\left[u\left(C\left(t\right),H\left(t\right)\right)-u\left(\bar{C},\bar{H}\right)\right]dt \\
&= \int_0^\infty e^{-\rho t}\left[u\left(\bar{C}+\Delta C\left(t\right),\bar{H}+\Delta H\left(t\right)\right)-u\left(\bar{C},\bar{H}\right)\right]dt \\
&= \int_0^\infty e^{-\rho t}\left[u\left(\bar{C}+B,\bar{H}+e^{-\theta t}A+\left(1-e^{-\theta t}\right)B\right)-u\left(\bar{C},\bar{H}\right)\right]dt \\
&= \int_0^\infty e^{-\rho t}\left\{u_C B+\frac{1}{2}u_{CC}B^2+u_{CH}B\left(e^{-\theta t}A+\left(1-e^{-\theta t}\right)B\right)+\cdots\right. \\
&\quad \left.+\left[u_H\left(e^{-\theta t}A+\left(1-e^{-\theta t}\right)B\right)+\frac{1}{2}u_{HH}\left(e^{-\theta t}A+\left(1-e^{-\theta t}\right)B\right)^2+\cdots\right]\right\}dt
\end{aligned}
$$

where the last equality holds because u is analytic, and all the utility derivatives are evaluated at the steady state.

For example, under the second-order approximation,

$$
\Psi\left(A,B\right)
$$

$$
\begin{aligned}
=&\frac{1}{\rho}\Bigg[u_C B + u_H \frac{\rho A + \theta B}{\rho + \theta} \\
&+ \frac{1}{2}\left(u_{CC} B^2 + 2u_{CH} \frac{\rho A B + \theta B^2}{\rho + \theta} + u_{HH} \frac{\rho\left(\rho + \theta\right) A^2 + 2\rho\theta A B + 2\theta^2 B^2}{\left(\rho + \theta\right)\left(\rho + 2\theta\right)}\right)\Bigg]
\end{aligned}
$$

The second quantity is related to the class of spending paths with

$$
\Delta C\left(t\right) = A \cdot 1\left(-1 < t \le 0\right) + B \cdot 1\left(0 < t \le 1\right) + 0 \cdot 1\left(t \le -1 \text{ or } t > 1\right).
$$

It follows that, for $t \ge 0$,

$$
\begin{aligned}
\Delta H\left(t\right) =& e^{-\theta\infty}\Delta H\left(-\infty\right) + \int_{-\infty}^{-1} e^{-\theta(t-s)}\theta 0 ds + \int_{-1}^{0} e^{-\theta(t-s)}\theta A ds \\
&+ \int_{0}^{\min\{1,t\}} e^{-\theta(t-s)}\theta B ds + \int_{\min\{1,t\}}^{t} e^{-\theta(t-s)}\theta 0 ds \\
=& e^{-\theta t}\left(1 - e^{-\theta}\right) A + \left(e^{-\theta(t-\min\{1,t\})} - e^{-\theta t}\right) B \\
=& e^{-\theta t}\left[\left(1 - e^{-\theta}\right) A + \left(e^{\theta\min\{1,t\}} - 1\right) B\right].
\end{aligned}
$$

The difference between lifetime utilities starting now, $\int_0^\infty e^{-\rho t} u\left(C\left(t\right), H\left(t\right)\right) dt$, under this class of spending paths and under the steady state $\left(\bar{C}, \bar{H}\right)$ is

$$
\begin{aligned}
&\Upsilon\left(A, B\right) \\
\equiv& \int_0^\infty e^{-\rho t} u\left(C\left(t\right), H\left(t\right)\right) dt - \int_0^\infty e^{-\rho t} u\left(\bar{C}, \bar{H}\right) dt \\
=& \int_0^\infty e^{-\rho t}\left[u\left(C\left(t\right), H\left(t\right)\right) - u\left(\bar{C}, \bar{H}\right)\right] dt \\
=& \int_0^\infty e^{-\rho t}\Big[u\left(\bar{C} + B \cdot 1\left(0 < t \le 1\right), \bar{H} + e^{-\theta t}\left(\left(1 - e^{-\theta}\right) A + \left(e^{\theta\min\{1,t\}} - 1\right) B\right)\right) \\
&-u\left(\bar{C}, \bar{H}\right)\Big] dt \\
=& \int_0^1 e^{-\rho t}\left[u\left(\bar{C} + B, \bar{H} + e^{-\theta t}\left(\left(1 - e^{-\theta}\right) A + \left(e^{\theta t} - 1\right) B\right)\right) - u\left(\bar{C}, \bar{H}\right)\right] dt \\
&+ \int_1^\infty e^{-\rho t}\left[u\left(\bar{C}, \bar{H} + e^{-\theta t}\left(1 - e^{-\theta}\right)\left(A + B e^{\theta}\right)\right) - u\left(\bar{C}, \bar{H}\right)\right] dt \\
=& \int_0^1 e^{-\rho t}\Big\{u_C B + \frac{1}{2} u_{CC} B^2 + u_{CH} B e^{-\theta t}\left(\left(1 - e^{-\theta}\right) A + \left(e^{\theta t} - 1\right) B\right) + \cdots \\
&+ \Big[u_H e^{-\theta t}\left(\left(1 - e^{-\theta}\right) A + \left(e^{\theta t} - 1\right) B\right) + \frac{1}{2} u_{HH}\left(e^{-\theta t}\left(\left(1 - e^{-\theta}\right) A + \left(e^{\theta t} - 1\right) B\right)\right)^2
\end{aligned}
$$

$$
\begin{aligned}
&+\cdots\Big]\Big\}\,dt+\int_1^\infty e^{-\rho t}\Big[u_H e^{-\theta t}\left(1-e^{-\theta}\right)\left(A+Be^{\theta}\right)\\
&+\frac{1}{2}u_{HH}\left(e^{-\theta t}\left(1-e^{-\theta}\right)\left(A+Be^{\theta}\right)\right)^2+\cdots\Big]\,dt.
\end{aligned}
$$

When others' spending varies, it is necessary to consider peer effect and external habit: $u\left(C, C_{\text{others}}, H\right)$ with $\dot{H}=\theta\left((1-\omega)C+\omega C_{\text{others}}-H\right)$. Under this habit evolution,

$$
\Delta H\left(t\right)=e^{-\theta(t-t_0)}H\left(t_0\right)+\int_{t_0}^{t}e^{-\theta(t-s)}\theta\left[(1-\omega)\,\Delta C\left(s\right)+\omega\Delta C_{\text{others}}\left(s\right)\right]ds,
$$

where $\Delta C_{\text{others}}\left(t\right)\equiv C_{\text{others}}\left(t\right)-\bar{C}_{\text{others}}$.

For the class of spending paths with $\Delta C\left(t\right)=A\cdot 1\left(t\le 0\right)+B\cdot 1\left(t>0\right)$ and $\Delta C_{\text{others}}\left(t\right)=D\cdot 1\left(t\le 0\right)+E\cdot 1\left(t>0\right)$, similar to equation (A.9),

$$
\Delta H\left(t\right)=e^{-\theta t}\left((1-\omega)\,A+\omega D\right)+\left(1-e^{-\theta t}\right)\left((1-\omega)\,B+\omega E\right).
$$

The difference between lifetime utilities starting now, $\int_0^\infty e^{-\rho t}u\left(C\left(t\right), C_{\text{others}}\left(t\right), H\left(t\right)\right)dt$, under this class of spending paths and under the steady state $\left(\bar{C}, \bar{C}_{\text{others}}, \bar{H}\right)$, the third quantity, is

$$
\begin{aligned}
&\Phi\left(A, B, D, E\right)\\
&\equiv\int_0^\infty e^{-\rho t}u\left(C\left(t\right), C_{\text{others}}\left(t\right), H\left(t\right)\right)dt-\int_0^\infty e^{-\rho t}u\left(\bar{C}, \bar{C}_{\text{others}}, \bar{H}\right)dt\\
&=\int_0^\infty e^{-\rho t}\left[u\left(C\left(t\right), C_{\text{others}}\left(t\right), H\left(t\right)\right)-u\left(\bar{C}, \bar{C}_{\text{others}}, \bar{H}\right)\right]dt\\
&=\int_0^\infty e^{-\rho t}\Big[u\Big(\bar{C}+B, \bar{C}_{\text{others}}+E,\\
&\quad\bar{H}+e^{-\theta t}\left((1-\omega)\,A+\omega D\right)+\left(1-e^{-\theta t}\right)\left((1-\omega)\,B+\omega E\right)\Big)-u\left(\bar{C}, \bar{C}_{\text{others}}, \bar{H}\right)\Big]\,dt\\
&=\int_0^\infty e^{-\rho t}\Big\{u_C B+u_{C_{\text{others}}}E+u_H\Big[e^{-\theta t}\left((1-\omega)\,A+\omega D\right)\\
&\quad+\left(1-e^{-\theta t}\right)\left((1-\omega)\,B+\omega E\right)\Big]+\frac{1}{2}u_{CC}B^2+\frac{1}{2}u_{C_{\text{others}}C_{\text{others}}}E^2\\
&\quad+\frac{1}{2}u_{HH}\left[e^{-\theta t}\left((1-\omega)\,A+\omega D\right)+\left(1-e^{-\theta t}\right)\left((1-\omega)\,B+\omega E\right)\right]^2\\
&\quad+u_{CC_{\text{others}}}BE+u_{CH}B\left[e^{-\theta t}\left((1-\omega)\,A+\omega D\right)+\left(1-e^{-\theta t}\right)\left((1-\omega)\,B+\omega E\right)\right]\\
&\quad+u_{C_{\text{others}}H}E\left[e^{-\theta t}\left((1-\omega)\,A+\omega D\right)+\left(1-e^{-\theta t}\right)\left((1-\omega)\,B+\omega E\right)\right]+\cdots\Big\}\,dt.
\end{aligned}
$$

D.3 Proof of Proposition 1

Proof. That θ is habit depreciation rate implies $\theta \in \mathbb{R}^{+}$. Taking $M = 5000$ gives $M - \Delta C_{U1} - \left(1 - e^{-\theta}\right) \Delta C_{U2} > 0$ in all the questions for habit depreciation rate.[6]

A respondent preferring Universe One for a better future experience (U) in a habit depreciation rate question implies

$$
\begin{aligned}
& U\,(\text{Universe One}) - U\,(\text{Universe Two}) \\
=& \left[U\,(\text{Universe One}) - U\,(\text{Baseline})\right] - \left[U\,(\text{Universe Two}) - U\,(\text{Baseline})\right] \\
=& \Psi\left(\Delta C_{U1}, 0\right) - \Upsilon\left(\Delta C_{U2}, 0\right) \\
=& \int_0^\infty e^{-\rho t} \left[u_H e^{-\theta t} \Delta C_{U1} + \frac{1}{2} u_{HH} \left(e^{-\theta t} \Delta C_{U1}\right)^2 + \cdots\right] dt \\
& - \int_0^\infty e^{-\rho t} \left[u_H e^{-\theta t} \left(1 - e^{-\theta}\right) \Delta C_{U2} + \frac{1}{2} u_{HH} \left(e^{-\theta t} \left(1 - e^{-\theta}\right) \Delta C_{U2}\right)^2 + \cdots\right] dt \\
=& \int_0^\infty e^{-\rho t} \left\{u_H e^{-\theta t} \left[\Delta C_{U1} - \left(1 - e^{-\theta}\right) \Delta C_{U2}\right]\right. \\
& \left. + \frac{1}{2} u_{HH} e^{-2\theta t} \left[\left(\Delta C_{U1}\right)^2 - \left(\left(1 - e^{-\theta}\right) \Delta C_{U2}\right)^2\right] + \cdots\right\} dt \\
=& \left[\Delta C_{U1} - \left(1 - e^{-\theta}\right) \Delta C_{U2}\right] \int_0^\infty e^{-\rho t} \left\{u_H e^{-\theta t}\right. \\
& \left. + \frac{1}{2} u_{HH} e^{-2\theta t} \left[\Delta C_{U1} + \left(1 - e^{-\theta}\right) \Delta C_{U2}\right] + \cdots\right\} dt \\
=& \left[\Delta C_{U1} - \left(1 - e^{-\theta}\right) \Delta C_{U2}\right] \\
& \cdot \int_0^\infty e^{-\rho t} \sum_{n=1}^{\infty} e^{-n\theta t} \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left[\sum_{s=0}^{n-1} \left(\Delta C_{U1}\right)^s \left(\left(1 - e^{-\theta}\right) \Delta C_{U2}\right)^{n-1-s}\right] dt \\
>& 0.
\end{aligned}
$$

The inequality, by Lemma 3, implies $\Delta C_{U1} - \left(1 - e^{-\theta}\right) \Delta C_{U2} < 0$ or equivalently

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

Q.E.D.

It is worth noting that when others' spending does not vary, the elicitation propositions in

[6] See Figure 2 of the paper for the initial monthly spending graphs and Table A.2 for all the values of ΔC_{U1} and ΔC_{U2} for this survey module.

here and the following, which are based on $\dot{H} = \theta (C - H)$, give exactly the same thresholds for the preference parameters of interest as under $\dot{H} = \theta ((1 - \omega) C + \omega C_{\text{others}} - H)$, and therefore lead to precisely the same estimates.

D.4 Proof of Proposition 2

Proof.

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

and

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

Q.E.D.

D.5 Proof of Proposition 3

Proof.

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

and

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

Q.E.D.

D.6 Proof of Proposition 4

Proof. A respondent preferring Universe One for a better future experience (U) in a slope of indifference curve question[7] implies

$$\begin{aligned}
& U \text{ (Universe One)} - U \text{ (Universe Two)} \\
=& [U \text{ (Universe One)} - U \text{ (Baseline)}] - [U \text{ (Universe Two)} - U \text{ (Baseline)}] \\
=& \Psi \left(\Delta C_{\text{past}}, \Delta C_{\text{future}}\right) - \Psi \left(-\Delta C_{\text{past}}, -\Delta C_{\text{future}}\right) \\
=& \frac{1}{\rho} \left\{ u_C \Delta C_{\text{future}} + u_H \frac{\rho \Delta C_{\text{past}} + \theta \Delta C_{\text{future}}}{\rho + \theta} \right.
\end{aligned}$$

[7] See Figure A.11 of the supplemental material for the initial monthly spending graphs and Table A.2 for all the values of ΔC_{past} and ΔC_{future} for this survey module.

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

where the third equality holds under the second-order approximation.

The inequality, by $u_C > 0$ and $\rho > 0$,[8] implies

$$
-\frac{u_H}{u_C} < \frac{\left(\rho+\theta\right)\Delta C_{\text{future}}}{\rho\Delta C_{\text{past}}+\theta\Delta C_{\text{future}}}.
$$

Q.E.D.

D.7 Proof of Proposition 5

Proof. A respondent preferring Universe One for a better future experience (U) in a $\frac{Hu_{HH}}{u_H}$ question[9] implies

$$
\begin{aligned}
&U\text{ (Universe One)}-U\text{ (Universe Two)}\\
=&U\text{ (Universe One)}-U\text{ (Baseline)}\\
=&\frac{1}{2}\Psi\left(-\Delta C_1,0\right)+\frac{1}{2}\Psi\left(\Delta C_2,0\right)\\
=&\frac{1}{2}\left[u_H\frac{-\Delta C_1}{\rho+\theta}+\frac{1}{2}u_{HH}\frac{\left(-\Delta C_1\right)^2}{\rho+2\theta}\right]+\frac{1}{2}\left[u_H\frac{\Delta C_2}{\rho+\theta}+\frac{1}{2}u_{HH}\frac{\left(\Delta C_2\right)^2}{\rho+2\theta}\right]
\end{aligned}
$$

[8] The sign of ρ is elicited in the time discount rate question.

[9] See Figure A.12 of the supplemental material for the initial monthly spending graphs and Table A.2 for all the values of ΔC_1 and ΔC_2 for this survey module.

$$
\begin{aligned}
=&\frac{1}{2}\left[u_H \frac{\Delta C_2-\Delta C_1}{\rho+\theta}+\frac{1}{2}u_{HH}\frac{(\Delta C_1)^2+(\Delta C_2)^2}{\rho+2\theta}\right] \\
>&0,
\end{aligned}
$$

where the third equality holds under the second-order approximation.

The inequality, by $u_H < 0$[10] and $H > 0$, implies

$$
\frac{Hu_{HH}}{u_H} < \frac{2(\rho+2\theta)}{\rho+\theta}\frac{\Delta C_1/\Delta C_2-1}{(\Delta C_1/\Delta C_2)^2+1}\frac{H}{\Delta C_2}.
$$

Q.E.D.

D.8 Proof of Proposition 6

Proof. A respondent preferring Universe One for a better future experience (U) in a $\frac{u_{CH}}{u_{HH}}$ question[11] implies

$$
\begin{aligned}
&U\,(\text{Universe One})-U\,(\text{Universe Two}) \\
=&\left[U\,(\text{Universe One})-U\,(\text{Baseline})\right]-\left[U\,(\text{Universe Two})-U\,(\text{Baseline})\right] \\
=&\frac{1}{2}\Psi\left(\Delta C_{\text{past}},\Delta C_{\text{future}}\right)+\frac{1}{2}\Psi\left(-\Delta C_{\text{past}},-\Delta C_{\text{future}}\right)-\frac{1}{2}\Psi\left(0,\Delta C_{\text{future}}\right)-\frac{1}{2}\Psi\left(0,-\Delta C_{\text{future}}\right) \\
=&\frac{1}{2}\frac{1}{\rho}\left\{\left[u_C\Delta C_{\text{future}}+u_H\frac{\rho\Delta C_{\text{past}}+\theta\Delta C_{\text{future}}}{\rho+\theta}+\frac{1}{2}\left(u_{CC}\left(\Delta C_{\text{future}}\right)^2\right.\right.\right. \\
&+2u_{CH}\frac{\rho\Delta C_{\text{past}}\Delta C_{\text{future}}+\theta\left(\Delta C_{\text{future}}\right)^2}{\rho+\theta} \\
&\left.\left.+u_{HH}\frac{\rho(\rho+\theta)\left(\Delta C_{\text{past}}\right)^2+2\rho\theta\Delta C_{\text{past}}\Delta C_{\text{future}}+2\theta^2\left(\Delta C_{\text{future}}\right)^2}{(\rho+\theta)(\rho+2\theta)}\right)\right] \\
&+\left[-u_C\Delta C_{\text{future}}-u_H\frac{\rho\Delta C_{\text{past}}+\theta\Delta C_{\text{future}}}{\rho+\theta}+\frac{1}{2}\left(u_{CC}\left(\Delta C_{\text{future}}\right)^2\right.\right. \\
&+2u_{CH}\frac{\rho\Delta C_{\text{past}}\Delta C_{\text{future}}+\theta\left(\Delta C_{\text{future}}\right)^2}{\rho+\theta} \\
&\left.\left.+u_{HH}\frac{\rho(\rho+\theta)\left(\Delta C_{\text{past}}\right)^2+2\rho\theta\Delta C_{\text{past}}\Delta C_{\text{future}}+2\theta^2\left(\Delta C_{\text{future}}\right)^2}{(\rho+\theta)(\rho+2\theta)}\right)\right]
\end{aligned}
$$

[10] This sign is elicited in the existence of (internal) habit formation question.

[11] See Figure A.13 of the supplemental material for the initial monthly spending graphs and Table A.2 for all the values of ΔC_{past} and ΔC_{future} for this survey module.

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

where the third equality holds under the third-order approximation.[12]

The inequality, by $u_{HH}<0$ and $\rho>0$,[13] implies

$$
\frac{u_{CH}}{u_{HH}}<-\frac{(\rho+\theta)\Delta C_{\text{past}}+2\theta\Delta C_{\text{future}}}{2(\rho+2\theta)\Delta C_{\text{future}}}.
$$

Q.E.D.

D.9 Proof of Proposition 7

Proof. A respondent preferring Universe One for a better future experience (U) in a $\frac{u_{CC}}{u_{HH}}$ question[14] implies

$$
\begin{aligned}
&U\,(\text{Universe One})-U\,(\text{Universe Two})\\
=&\left[U\,(\text{Universe One})-U\,(\text{Baseline})\right]-\left[U\,(\text{Universe Two})-U\,(\text{Baseline})\right]\\
=&\frac{1}{2}\Psi(0,\Delta C_{\text{future}})+\frac{1}{2}\Psi(0,-\Delta C_{\text{future}})-\frac{1}{2}\Psi\left(\Delta C_{\text{past}},0\right)-\frac{1}{2}\Psi\left(-\Delta C_{\text{past}},0\right)\\
=&\frac{1}{2}\frac{1}{\rho}\Bigg\{\left[\left(u_C+\frac{u_H\theta}{\rho+\theta}\right)\Delta C_{\text{future}}+\frac{1}{2}\left(u_{CC}+\frac{2u_{CH}\theta}{\rho+\theta}+\frac{2u_{HH}\theta^2}{(\rho+\theta)(\rho+2\theta)}\right)(\Delta C_{\text{future}})^2\right]
\end{aligned}
$$

[12] The third-order terms cancel each other and are omitted for space consideration.

[13] The signs are elicited in the Hu_{HH}/u_H and time discount rate questions.

[14] See Figure A.14 of the supplemental material for the initial monthly spending graphs and Table A.2 for all the values of ΔC_{past} and ΔC_{future} for this survey module.

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

where the second equality holds under the third-order approximation.[15]

The inequality, by $u_{HH} < 0$ and $\rho > 0$,[16] implies

$$
\frac{u_{CC}}{u_{HH}} < \frac{\rho}{\rho+2\theta}\left(\frac{\Delta C_{\text{past}}}{\Delta C_{\text{future}}}\right)^2-\frac{2\theta}{\rho+\theta}\frac{u_{CH}}{u_{HH}}-\frac{2\theta^2}{(\rho+\theta)(\rho+2\theta)}.
$$

Q.E.D.

D.10 Proof of Proposition 8

Proof. Taking $M = 5000$ gives $M-(1-\omega)\Delta C-\omega\Delta C_{\text{others}} > 0$ in all the questions for external habit formation.[17]

A respondent preferring Universe One for a better future experience (U) in an external habit formation question implies

$$
\begin{aligned}
&U\,(\text{Universe One})-U\,(\text{Universe Two})\\
=&[U\,(\text{Universe One})-U\,(\text{Baseline})]-[U\,(\text{Universe Two})-U\,(\text{Baseline})]\\
=&\Phi\,(\Delta C,0,0,0)-\Phi\,(0,0,\Delta C_{\text{others}},0)\\
=&\int_0^\infty e^{-\rho t}\left[u_He^{-\theta t}(1-\omega)\Delta C+\frac{1}{2}u_{HH}\left(e^{-\theta t}(1-\omega)\Delta C\right)^2+\cdots\right]dt\\
&-\int_0^\infty e^{-\rho t}\left[u_He^{-\theta t}\omega\Delta C_{\text{others}}+\frac{1}{2}u_{HH}\left(e^{-\theta t}\omega\Delta C_{\text{others}}\right)^2+\cdots\right]dt\\
=&[(1-\omega)\Delta C-\omega\Delta C_{\text{others}}]\int_0^\infty e^{-\rho t}\Bigg[u_He^{-\theta t}\\
&+\frac{1}{2}u_{HH}\left(e^{-\theta t}\right)^2[(1-\omega)\Delta C+\omega\Delta C_{\text{others}}]+\cdots\Bigg]dt
\end{aligned}
$$

[15] The third-order terms cancel each other and are omitted for space consideration.

[16] The signs are elicited in the Hu_{HH}/u_H and time discount rate questions.

[17] See Figure A.15 of the supplemental material for the initial monthly spending graphs and Table A.2 for all the values of ΔC and ΔC_{others} for this survey module.

$$
\begin{aligned}
&= \left[(1-\omega)\,\Delta C - \omega \Delta C_{\text{others}}\right] \\
&\quad \cdot \int_0^{\infty} e^{-\rho t} \sum_{n=1}^{\infty} e^{-n\theta t} \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} \left((1-\omega)\,\Delta C\right)^s \left(\omega \Delta C_{\text{others}}\right)^{n-1-s} \right) dt \\
&> 0.
\end{aligned}
$$

The inequality, by Lemma 3, implies $(1-\omega)\,\Delta C - \omega \Delta C_{\text{others}} < 0$ or equivalently

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

Q.E.D.

D.11 Proof of Proposition 9

Proof. A respondent preferring Universe One for a better future experience (U) in a $\frac{u_{C_{\text{others}}}}{u_H}$ question[18] implies

$$
\begin{aligned}
& U\text{ (Universe One)} - U\text{ (Universe Two)} \\
={}& \left[U\text{ (Universe One)} - U\text{ (Baseline)}\right] - \left[U\text{ (Universe Two)} - U\text{ (Baseline)}\right] \\
={}& \Phi\left(0,0,0,\Delta C^{U1}_{\text{others}}\right) - \Phi\left(0,0,\Delta C^{U2}_{\text{others}},0\right) \\
={}& \int_0^{\infty} e^{-\rho t} \left[u_{C_{\text{others}}} \Delta C^{U1}_{\text{others}} + u_H \left(1 - e^{-\theta t}\right) \omega \Delta C^{U1}_{\text{others}} \right] dt - \int_0^{\infty} e^{-\rho t} u_H e^{-\theta t} \omega \Delta C^{U2}_{\text{others}} dt \\
={}& \frac{1}{\rho} u_{C_{\text{others}}} \Delta C^{U1}_{\text{others}} + u_H \frac{\omega}{\rho(\rho+\theta)} \left(\theta \Delta C^{U1}_{\text{others}} - \rho \Delta C^{U2}_{\text{others}}\right) \\
>{}& 0,
\end{aligned}
$$

where the third equality holds under the first-order approximation.

The last inequality, by $u_H < 0$ and $\rho > 0$,[19] implies

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

Q.E.D.

[18] See Figure A.16 of the supplemental material for the initial monthly spending graphs and Table A.2 for all the values of $\Delta C^{U1}_{\text{others}}$ and $\Delta C^{U2}_{\text{others}}$ for this survey module.

[19] The signs are elicited in the existence of (internal) habit formation and time discount rate questions.

Table A.2: 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 speed	ΔC_{U1}	400	1200	2000	2000	2000
	ΔC_{U2}	4000	4000	4000	2800	2200
External habit mixture coefficient	ΔC	4500	1200	500	500	500
	ΔC_{others}	500	500	500	1200	4500
$-\frac{u_H}{u_C}$	ΔC_{past}	2000	2000	2000	2000	2000
	ΔC_{future}	20	80	200	400	1000
$\frac{H u_{HH}}{u_H}$	ΔC_1	540	600	650	700	800
	ΔC_2	500	500	500	500	500
$\frac{u_{CH}}{u_{HH}}$	ΔC_{past}	2000	1600	1000	600	100
	ΔC_{future}	200	200	200	200	200
$\frac{u_{CC}}{u_{HH}}$	ΔC_{past}	500	1500	2200	3000	3500
	ΔC_{future}	500	500	500	500	500
$\frac{u_{C_{\text{others}}}}{u_H}$	$\Delta C_{\text{others}}^{U1}$	3000	600	300	150	100
	$\Delta C_{\text{others}}^{U2}$	3000	3000	3000	3000	3000

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 a module at the end of the 1st follow-up question (cf. Figure 5 in the paper). All amounts are in U.S. dollars.