Survey Evidence on Habit Formation

Jiannan Zhou*

March 31, 2020

Habit formation has explained many important economic phenomena but is surrounded by controversies over its existence, specifications, and implications. To fill this gap, I document an extensive set of micro evidence for habit formation, through designing and fielding a survey eliciting ten preference parameters informative about habit formation. Habit forms both internally and externally, depreciates by 66% annually, and has equisized welfare impact as peer effect. I propose and implement four tests of additive and multiplicative habits and find that these ubiquitous preferences are rejected. Finally, I show that combining habit formation with peer effect could explain the Easterlin paradox.

JEL: E21, G12, D60

Keywords: habit formation, micro evidence, Easterlin paradox

Habit formation captures the phenomenon of response decrement to repetitive stimulation.[1] It 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

* 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, and many seminar and conference participants for their helpful comments.

[1]This notion of habit formation is what the current economic literature on habit formation is trying to model and is consistent with the biology literature. It differs from the day-to-day notion of the cue-routine-reward habit. Because it is already a widely accepted term in economics, I will continue referring to this phenomenon as habit formation. This definition of habit formation sets it apart from path dependence, with which it is sometimes confused. It is also worth noting that habit formation 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; Campbell

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 this question, I use and extend Barsky et al.'s (1997) method of direct survey measurement of structural preferences to document a new and extensive set of micro evidence for 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 behavior) and external habit formation (habit based on other people's past behavior). Literature investigating evidence of habit formation has mostly focused on the existence of internal habit formation and has drawn mixed conclusions (see column 1 of Table 1). Studies based on macrodata tend to find significant evidence of internal habit formation, whereas those based on microdata find it to be relatively less significant. Evidence of external habit formation has received much less attention (column 2 of Table 1), although numerous theoretical papers assume its existence.[3] These two types of habit formation can have dramatically different implications for optimal tax policy and welfare analysis (Ljungqvist and Uhlig, 2000, 2015). Designing and implementing thought experiments that differentiate between the two types of habit formation, I document evidence for the existence of both internal and external habit formation. Few authors have studied the composition of internal and external habit formation (column 3 of Table 1). Allowing habit to form both internally and externally as per Grishchenko (2010), I estimate that external habit formation accounts for a small portion (about 19%) 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.

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 and abstracts from habits based on individual varieties of consumption (e.g., Ravn, Schmitt-Grohé and Uribe, 2006). Total consumption habit formation accounts for the overwhelming majority of current economic literature on habit formation. Following the literature, I do not add qualifiers like *total* or *consumption* to the word *habit* when referring to this habit.

[3] See, e.g., Abel, 1990; Campbell and Cochrane, 1999; Smets and Wouters, 2007; Dou et al., 2017.

[4] I focus 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.

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. Two potential reasons for the current state of the literature are a lack of recognition of the importance of the habit depreciation rate parameter and the massive disagreement between its few existing micro- and macro-estimates (column 5 of Table 1). To illustrate this parameter's importance, I show that simply changing its value can significantly affect the performance of habit formation models. My estimate of this parameter is based on microdata and aggregated for a representative agent. It is much closer to existing macrodata-based estimates than existing microdata-based estimates.

Third, neither the additive nor multiplicative habit formation 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 tests and estimates with specifications of questionable validity. In a general utility function naturally nesting these two formulations, I propose and implement four tests of the validity of the preference specifications. To the best of my knowledge, this is the first time such tests have been done in the literature. The tests are based on the mapping between the preferences and their indifference curves. The test results imply that both types of habit utility function 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 a general preference are consistent with the idea of habit formation,[5] suggesting that specifications of habit formation preference consistent with the survey evidence could be found.[6]

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

[5]Specifically, $u_H < 0$ and $u_{CH} > 0$.
[6]I leave this direction to future research.

(Luttmer, 2005; De Giorgi, Frederiksen and Pistaferri, 2020) but have disagreed on the strength of its welfare impact relative to habit formation. Alvarez-Cuadrado, Casado and Labeaga (2015) estimate internal habit formation to be as strong as peer effect, while Ravina (2019) finds internal habit formation to be about 70% stronger than peer effect. Allowing both internal and external habit formation, I provide an estimate of the relative strength of the two phenomena 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).[7] 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 existing habit formation models are rejected, the credibility of structural simulations under any existing specification is unknown. Using my aforementioned extensive evidence on habit formation that is free from such specification errors, I conduct a semi-structural simulation and find that, when coupled with peer effect, habit 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 I call "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 total stairs you run and the elevation you reach repre-

[7]There is an ongoing debate on whether the long-run gradient is exactly zero or slightly positive. I intend not to participate in the debate, as I supply no new evidence on happiness measures, and my discussion here very easily accommodates both views. My goal is to show how habit formation (and peer effect) affects the relationship between income (or consumption) and happiness.

sent 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, your elevation increases, implying a positive correlation between total stairs run and elevation, just like the positive happiness–income gradient in the short run. After the escalator catches up with you, your elevation stops changing even though you keep running, implying a zero correlation between total stairs run and elevation, just like the long-run nil (or low) happiness–income gradient. People who run faster plateau at higher elevations, implying a positive correlation between total stairs run and elevation, just like the cross-section positive happiness–income gradient. This analysis implies that even if happiness eventually stops growing with income, continued income growth is still necessary to maintain happiness. In the language of the analogy, keeping running is necessary to maintain elevation.

Providing this new and extensive set of micro evidence for habit formation requires overcoming two main challenges. The first is to find 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 reducing and uncovering specification errors that potentially exist in all current habit formation models. Two recent papers have investigated habit formation models with relatively limited generality of this kind. 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 relaxes the joint concavity of the nonparametric felicity function 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 preference parameters[8] that

[8]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.

are instrumental to addressing the controversies over habit formation.

The second main challenge is that the estimation of all the preference parameters of interest requires variations that, to the best of my knowledge, do not exist in reality. Some of the variations are even impossible to be approximated using conventional methods under the required generality of the framework.[9] In particular, revealed preference approach does not allow testing of multiplicative habit, one of the only two common habit formation preferences in current literature. This is because, under utility maximization in competitive markets, revealed preference approach cannot reveal non-concave or non-quasiconcave preferences[10] (Samuelson, 1950) 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 part(s) of the indifference curves. Consequently, tests of non-quasiconcave preferences, like the multiplicative habit,[11] necessarily require a different approach. To deal with this challenge, I employ and extend the stated preference survey method for direct measurement of structural preferences of Barsky et al. (1997). Specifically, I designed thought experiments to elicit the ten preference parameters of interest, implemented the thought experiments in a survey, and fielded two waves of the survey on Amazon Mechanical Turk (MTurk). Applications of stated preference method in economics can be traced back to Thurstone (1931) and are commonly seen in the elicitation of reduced-form preference parameters in, among other fields, environmental economics, experimental economics, and health economics.[12] Barsky et al. (1997) extends the methods to the elicitation of structural preference parameters, which has since been used in a growing number of papers.[13]

Though subject to potential response biases and errors, direct survey measurement

[9]The required variation for estimating each parameter is different from that required for another parameter. See Section III for details.

[10]The inability to distinguish concave and quasiconcave preferences under finite data is another well-known limitation of the revealed preference approach. See also Diewert (1973).

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

[12]See, e.g., Johnston et al. (2017); Ameriks et al. (2019); Jha and Shayo (2019).

[13]See, e.g., Kapteyn and Teppa (2003); Sahm (2007); Kimball, Sahm and Shapiro (2008); Kimball and Shapiro (2008); Kimball, Sahm and Shapiro (2009); Benjamin et al. (2014); Kimball et al. (2015); Benjamin et al. (2019).

of preferences is designed to overcome the problems of conventional methods—such as weak and other identification issues, and measurement error and other data problems (Kimball and Shapiro, 2008). This method is especially indispensable in the context of habit formation where a lack of required variations in conventional data 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. The survey evidence is important, not just for the preference parameters conventional methods can never illuminate but also for the preference parameters conventional methods can illuminate. This is because the limitations of the conventional methods tend to be orthogonal to the limitations of the survey method, implying that the survey evidence and the evidence from conventional methods are complementary. Response biases and errors can be and have been carefully studied and dealt with through defining, measuring, and devising ways to address them. This paper deals with potential response biases and errors through the design and implementation of the thought experiments, survey, estimation, and robustness checks.[14]

In addition to providing the new and extensive micro evidence on habit formation, this paper also contributes to methods of preference elicitation. Most existing research elicits structural preference parameters in preferences with fully specified functional forms. To be immune to misspecification errors, some authors have dispensed with the functional form assumptions and used first-order approximations in eliciting structural parameters in preferences of unspecified forms (e.g., Benjamin et al., 2014). In this paper, I show that this approach is a special case of a general approach in which one could use any, even the infinitieth, order of approximation, allowing for potentially drastic reductions, including elimination, of approximation errors and improvement of the accuracy of preference elicitation.

This paper proceeds as follows. Section I presents the general model and survey instrument. Section II summarizes the data and statistical model. Section III contains the elicitation, estimate, and implication of each preference parameter of interest. Section IV explores the explanation of the Easterlin paradox. Section V concludes.[15]

[14]The robustness checks confirm the results of this paper and are reported in Section F of the

Table 1—Estimates of Habit Parameters in Selected Literature

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

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

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.

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, Moore (1996)	Y	*(N)*	*(0)*	0.08	*(Y)*	*(A)*
Dynan (2000)	N	*(N)*	*(0)*	-0.04	*(Y)*	*(A)*
Guariglia, Rossi (2002)	N	*(N)*	*(0)*	-0.27	*(Y)*	*(A)*
Lupton (2002)	Y	*(N)*	*(0)*	0.23	9.2%/Y	*(A)*
Kapteyn, Teppa (2003)	Y	*(N)*	*(0)*	0.78	[f]	*(M)*
Rhee (2004)	Y, N[b]	*(N)*	*(0)*	0.61, 0.62	*(Y)*	*(A)*
Browning, Collado (2007)	Y, N[b]	*(N)*	*(0)*	0.01-0.14	*(Q)*	*(A)*
Alessie, Teppa (2010)	Y	*(N)*	*(0)*	0.21	*(Y)*	*(A)*
Iwamoto (2013)	N	*(N)*	*(0)*	-0.38	*(Y)*	*(A)*
Khanal et al. (2018)	Y	*(N)*	*(0)*	0.55	*(Y)*	*(A)*
Ravina (2019)	Y	N	0.03[c]	0.50	*(Q)*	*(A, M)*
Panel B. Macrodata						
Ferson, 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, Wright (2000)	Y, N[b]	*(N)*	*(0)*	[d]	*(M, Y)*	*(A)*
Smets, Wouters (2003)	*(N)*	Y	*(1)*	0.57	*(Q)*	*(A)*
Lubik, Schorfheide (2004)	Y	*(N)*	*(0)*	0.57	*(Q)*	*(M)*
Christiano et al. (2005)	Y	*(N)*	*(0)*	0.65	*(Q)*	*(A)*
Adolfson et al. (2007)	Y	*(N)*	*(0)*	0.69	*(Q)*	*(A)*
Smets, 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 et al. (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 parenthesis (and italics for further distinction) are assumed parameter values of the studies. [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. The online Appendix contains the reference list of the studies.

online Appendix.

[15]Please see the online Appendix for aggregation of the preference parameters, observational equivalence of linear and nonlinear habit evolutions, additional effects of habit depreciation speed, nonparametric elicitation of existence of external habit formation, elicitation of time discount rate, response frequency, additional details of the survey, proofs, and additional figures and tables.

I. Methodology

A. 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, I maintain expected utility and exponential time discounting.[16] 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 is providing. In particular, I leave open whether and how habit affects utility.[17] I allow the respondent's utility to 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), I will explicitly show the additional variable(s) in the utility function.

Habit evolves according to $\dot{H} = \theta\left(C - H\right)$, where θ is the habit depreciation rate. I choose this specification 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 slightly 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

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

[17]As a technical note, to maintain the generality of the utility function under the infinitieth-order approximation, if habit exists, it is necessary and therefore I assume 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.

habit evolution.[18] Second, this habit evolution has an intuitive unit, 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 set of micro evidence for habit formation, I need information on whether habit affects utility and, if it does, the values of the preference parameters governing effects of habit on the utility: θ, ρ, ratios of utility derivatives up to the second order ($-\frac{u_H}{u_C}$, $\frac{Hu_{HH}}{u_H}$, $\frac{u_{CH}}{u_{HH}}$, and $\frac{u_{CC}}{u_{HH}}$), external habit mixture coefficient, and strength of habit formation relative to peer effect.[19] Eliciting, estimating, and using these preference parameters to shed light on the controversies surround habit formation is the primary subject of the rest of the paper.

B. Survey

To elicit the preference parameters, I design simple thought 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 thought experiments is to compare the welfare implications of different spending paths.[20] The exact thought experiments vary from one parameter to another and will be discussed in detail in Section III. This section sets the stage for that discussion by presenting the survey design.[21]

As discussed earlier, the elicitation of all the preference parameters of interest requires placing the comparison of spending paths in hypothetical or stated-preference 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.

[18]See Section B of the online Appendix for the proof. The basic intuition of the proof is that the nonlinearity can be "transported" to the utility function that is agnostic about how habit affects it.

[19]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 III.D and III.E.

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

[21]See Section G of the online Appendix for additional details of the survey.

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 traditional 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.

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

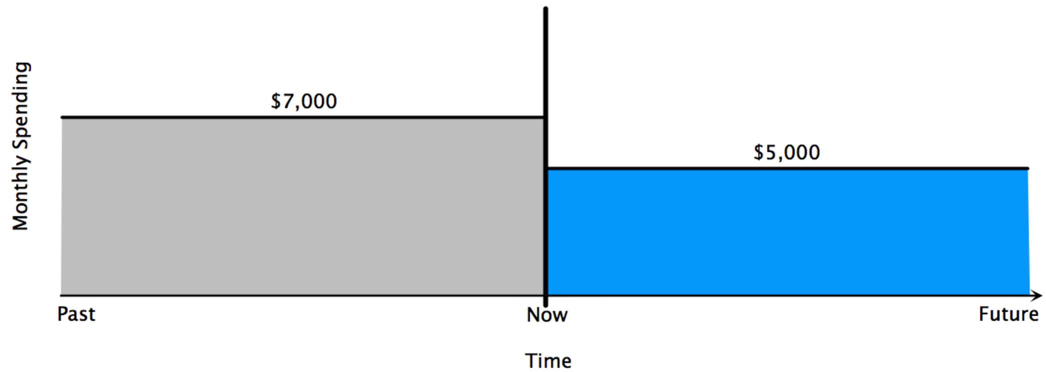


FIGURE 1. A TYPICAL MONTHLY SPENDING GRAPH

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 for a typical survey question.

Despite their advantages making possible the documentation of the extensive set of micro evidence for habit formation, surveys can be subject to response biases and errors. The most common concerns are sample representativeness, which I address in the next section, and incentive compatibility, which I address now. The survey is incentive-compatible 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 elicitations, 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 hypothetical nature of the questions and because none of the 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 mitigate these concerns, the thought 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, I reiterate it at the start of

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

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

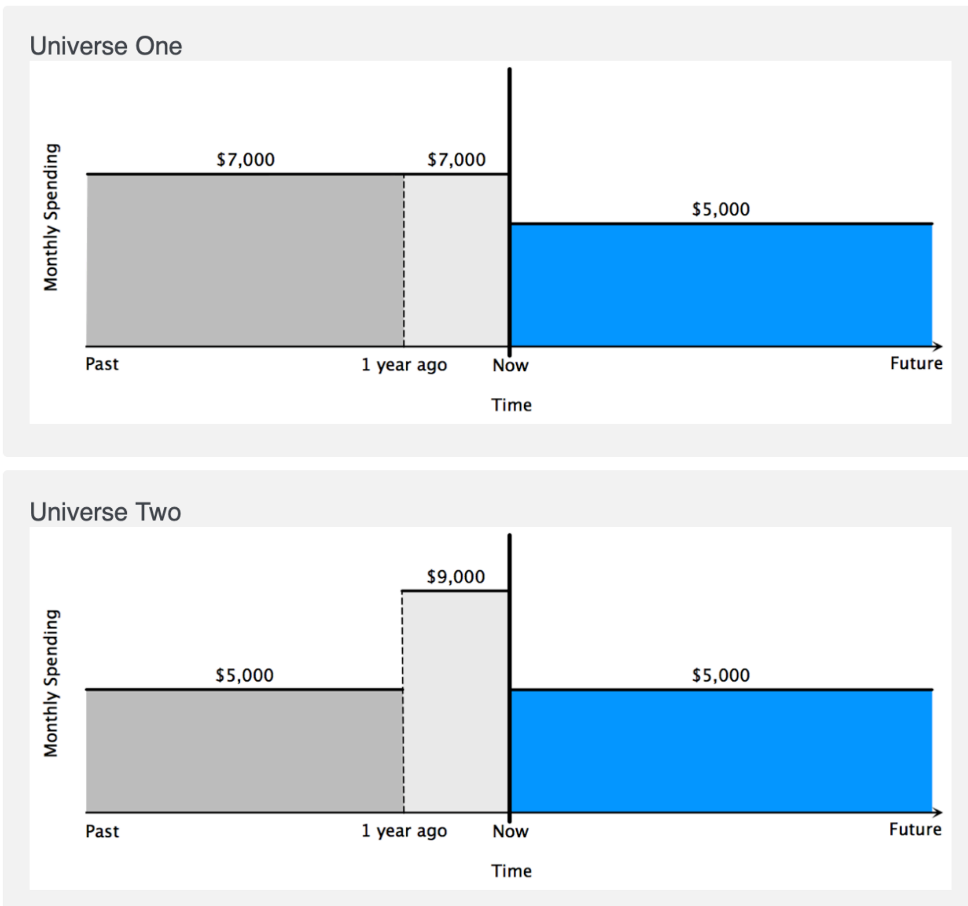


FIGURE 2. A TYPICAL SURVEY QUESTION

every core module. To help the respondents compare graphical spending paths, the survey questions also tell them in words in what time horizon the paths differ. To help them distinguish past experience from future experience, they are asked to express views on both experiences. I also repeat 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 avoid having respondents 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 across 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 the 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.[22] A series of more speculative checks on attention are also conducted.[23]

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

II. Data and Statistical Model

A. *Data*

I fielded the two waves of the survey on MTurk, an online crowdsourcing platform for human intelligence tasks. Many economic studies have used this platform.[24]

MTurk workers voluntarily participated in the study.[25] To avoid the potential influences of cultural differences, I restrict my respondents to U.S. residents. From

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

[23]See Section F.3 of the online Appendix.

[24]See, e.g., Oster, Shoulson and Dorsey, 2013; Kuziemko et al., 2015; Bordalo et al., 2016; Martínez-Marquina, Niederle and Vespa, 2019; Benjamin et al., 2019. Similar online platforms, like oDesk, have also been used in the literature (e.g., Pallais, 2014).

[25]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 et al., 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”	

Notes: Household income is annual. The two waves of the survey were fielded in July and August of 2018.
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).

the 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 my 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

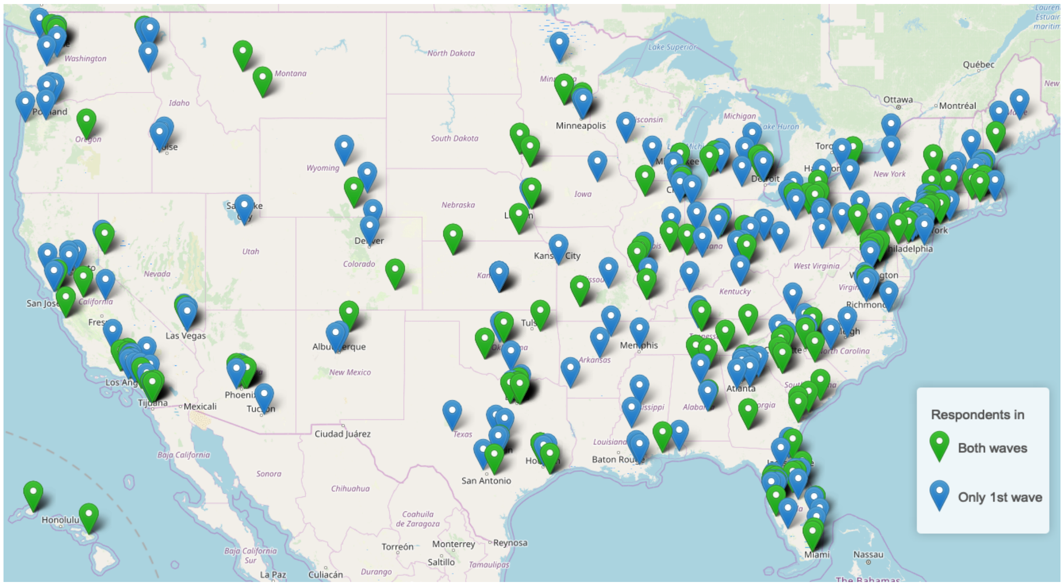


Figure 3. Locations of Respondents

for accuracy. As a result of this fact and potential response biases and errors that need to be taken care of by the statistical model, response frequencies for individual parameters alone are not particularly informative and are reported in Table A.14 of the online Appendix. Response frequencies of parameters identifiable to sign only are reported at the places where their estimates are reported.

B. Statistical Model

The statistical model underlying the estimation addresses response biases and errors not addressed by the design and implementation of the thought experiments and survey or by the elimination of invalid responses. Potential remaining response biases were dealt with through robustness checks in Section F of the online Appendix.

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. In a robustness check, I allow the means of the response biases and errors to be nonzero and to vary across waves and find 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, I assume that the parameters are independent within a respondent. Because my respondents spread across the United States (Figure 3) and most likely did not know each other, I also assume that responses are 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:

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

Given that almost all current habit formation models assume a representative agent, I will focus on the implication of my estimates for the representative-agent models with habit formation. In Section A 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.

For accuracy, I jointly estimate all the parameters identifiable to scale. To deal with the computational burden of the resulting high dimensional estimation, I utilize a Bayesian method to bypass the 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.

In implementing Hamiltonian Monte Carlo, I use uniform priors, not only to let data speak as much as possible but also to establish the equivalence between the maximum a posteriori (MAP) estimates and the maximum likelihood estimates.[26] 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[27] and the highest posterior density or mass interval (HPDI or HPMI).

[26]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.

[27]MAP together with posterior mean and median.

TABLE 3—RESPONSE FREQUENCY (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

III. Elicitation, Estimation, and Implication

A. 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 $Q_H \equiv u(C, H + \Delta h) - u(C, H) < 0$ but not with $Q_H \geq 0$, for $\Delta h > 0$ and $\Delta h \propto$ 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 over one with more past spending 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 Universe One (Table 3), consistent with the existence of internal habit formation for the representative agent. My estimate of $\text{sgn}(Q_H)$ confirms this (Table 4).

As a clarification, the fact that habit formation exists does not mean 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. My evidence of people's spending behavior shows that it exhibits response decrement to repetitive stimulation, matching the definition of habit formation. Therefore, habit formation exists (in people's spending behavior). The existence of habit formation, 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

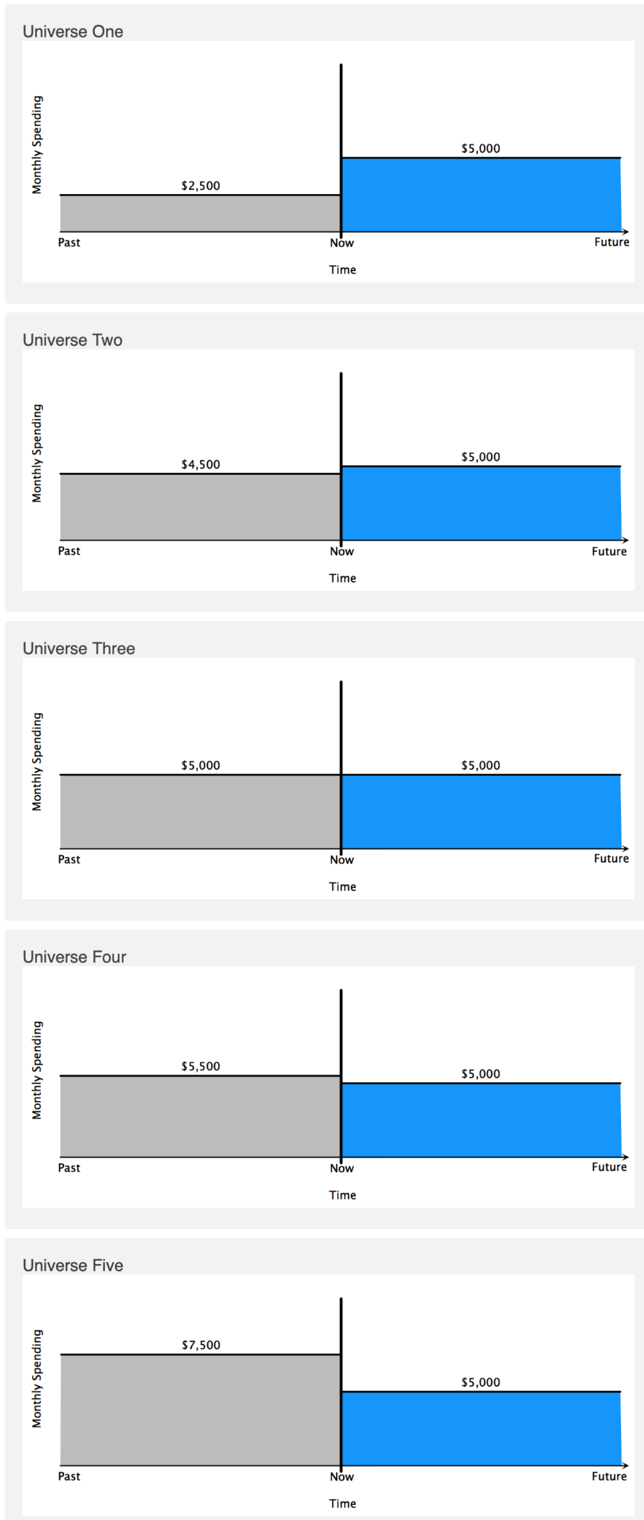


Figure 4. Monthly Spending Graphs: Existence of Internal Habit Formation

TABLE 4—ESTIMATES OF PREFERENCE PARAMETERS

	MAP	Mean	Median	95% HPDI/HPMI
sgn (Q_H)	-1.00	-1.00	-1.00	[-1.00, -1.00]
Habit depreciation rate	1.09	1.08	1.08	[0.89, 1.29]
Habit depreciation factor, annual	0.66	0.66	0.66	[0.59, 0.73]
$-u_H/u_C$	0.57	0.57	0.57	[0.46, 0.68]
Hu_{HH}/u_H	7.50	7.53	7.51	[6.63, 8.46]
u_{CH}/u_{HH}	-0.88	-0.88	-0.88	[-1.02, -0.73]
u_{CC}/u_{HH}	3.87	3.80	3.81	[3.10, 4.49]
External habit mixture coefficient	0.19	0.18	0.18	[0.09, 0.27]
$u_{C_{\text{others}}}/u_H$	1.12	1.12	1.12	[0.77, 1.48]
$u_{C_{\text{others}}}/u_C$	-0.62	-0.64	-0.63	[-0.88, -0.41]
$u_H/u_C + u_{C_{\text{others}}}/u_C$	-1.20	-1.21	-1.21	[-1.52, -0.90]

various behaviors across both humans and animals, including the amoeba, an organism without a neural system (Folger, 1926), seems to suggest the existence of deeper, and possibly universal, explanations for habit formation.[28] 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.

B. 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: *Under exact elicitation,* $\theta > -\ln\left(1 - \frac{\Delta C_{U1}}{\Delta C_{U2}}\right)$ *if the respondent chooses Universe One over Universe Two for a better future experience in the habit depreciation rate question.*[29]

PROOF: See Section I.2 of the online Appendix.

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

[29]Based on the accuracy of the elicitation, I distinguish different orders of elicitation. First-order elicitation means that the approximation error of the elicitation will be up to the remainder of a first-order Taylor expansion, while exact elicitation means zero approximation error. In eliciting some of the preference parameters under the general model, approximation errors are inevitable. Response biases and errors that might cause violations of the propositions are taken care of by the statistical model in Section II.B and robustness checks. 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.

The intuition of the proposition is that choosing the spending path with more persistent past spending (Universe One) for a better future experience requires habit to depreciate fast. In the proposition, ΔC_{U1} (ΔC_{U2}) denotes the difference between the monthly spending in Universe One (Universe Two) and the baseline monthly spending, \$5,000 per month. In the example of Figure 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 $\theta < \ln 2$.[30]

I use unfolding brackets to pin down a finer range of θ for each response: all respondents answered one to two follow-up questions that associate her response 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) falls between $\ln 2$ and $\ln 7/2$.

Applying the statistical model to the responses and the mappings between the responses and the parameters, I get an estimate of 1.09 for the habit depreciation rate, which corresponds to a (annual) habit depreciation factor of 0.66 (Table 4). This annual 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.[31]

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 for not being able to meet their habitual level of spending. The most cited paper on habit formation, Campbell and Cochrane (1999),[32] also employs the additive habit.[33] In their model, however, the faster habit depreciates, the more risk averse model agents are. The

[30]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.

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

[32]Google Scholar reports that this paper has been cited 5,094 times as of March 31, 2020.

[33]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.

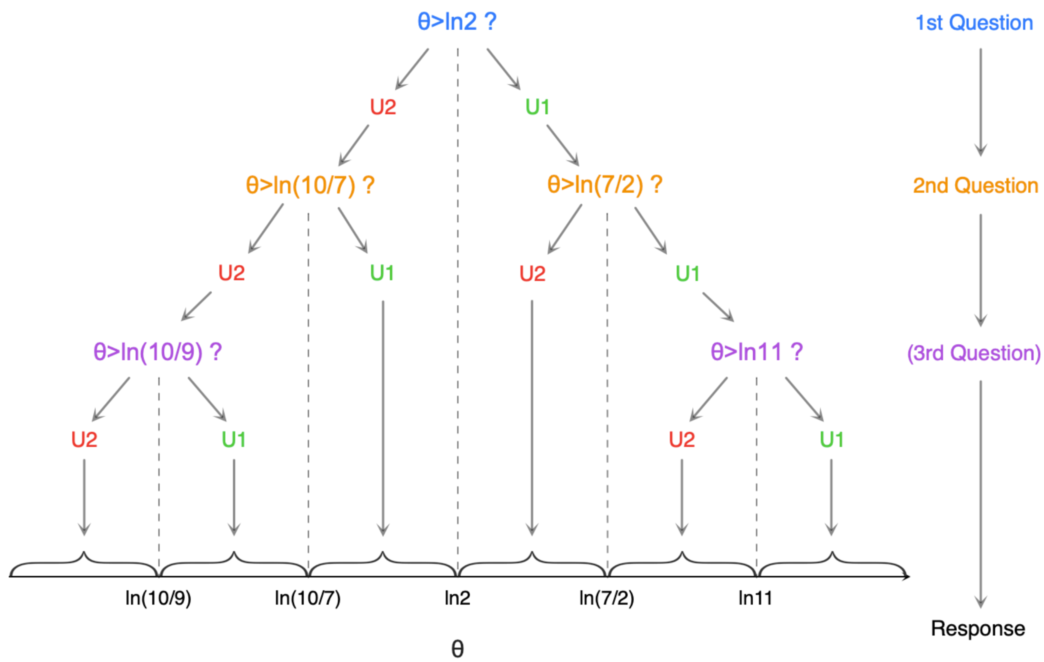


FIGURE 5. UNFOLDING BRACKETS
Note: U1 and U2 stand for Universe One and Universe Two, respectively.

reason is that the implied steady-state habit intensity[34] 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 agent becomes. The net effect of a higher habit depreciation rate in their model is the sum of these two effects, which ultimately makes the agent in the model more risk averse.

Plugging my estimate of the habit depreciation speed into Campbell and Cochrane (1999) causes the agents to become so risk averse that 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, 0.36 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

[34]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, $X_t = \alpha_t H_t$. After a steady state is reached, the implied steady-state habit intensity is, therefore, X_t / C_t.

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	**0.66**	**0.66**	**0.59**	**0.30**
Time discount factor	-	0.89	0.36	**0.89**	0.43	0.71
Expected excess log return	6.69%	6.71%	43.07%	98.80%	36.58%	16.51%
Std of excess log return	15.20%	15.64%	30.87%	95.49%	29.33%	22.01%
Sharpe ratio	0.43	0.43	1.40	1.03	1.25	0.75
Mean of risk-free rate	0.94%	0.94%	0.94%	-89.47%	0.94%	0.94%

Notes: All annualized values. Boldface denotes my 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 my estimate of habit depreciation factor; column 4 is based on the model sample under my 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 my estimate of habit depreciation factor; and column 6 is based on the model sample under a habit depreciation factor far smaller than the lower bound of the 99% HPDI of my estimate of it. The power coefficient of the constant relative risk aversion utility function is 2, as in Campbell and Cochrane (1999).

are willing to accept a hugely negative interest rate, -89.47% 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 almost 90% of the principal to be able to transfer the remaining about 10% of it to the next year. When one lowers the time discount factor or the habit depreciation factor, the model moments are closer to reality, but the percentage differences are still at least 40% (columns 5 and 6), even when habit depreciates by only 30% each year, which is far from the 99% HPDI of the habit depreciation factor. Section D of the online Appendix shows that other results of Campbell and Cochrane (1999) are also significantly affected by the habit depreciation speed.

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 my respondents is the same or close to that of the typical marginal investor, which would be the case if this parameter is a deep preference parameter that does not vary significantly across demographics. Section F of the online Appendix 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[35] is plugged in. One must, however, take extra caution in interpreting this finding. Given that my 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.

C. Testing Additive and Multiplicative Habits

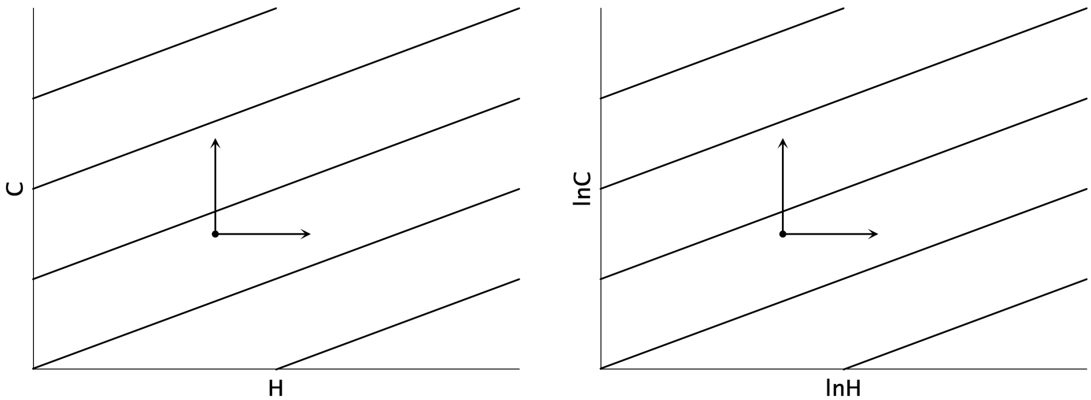
Additive and multiplicative habits are used by basically all current habit formation models that have been taken to data. The literature adopts additive habit relatively more often for its time-varying risk aversion.[36] To see whether the micro evidence supports this theoretical choice and, on a more basic level, 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* $\frac{u_{CH}}{u_{CC}} \frac{u_C}{u_H} = 1$.

PROOF: See Section I.3 of the online Appendix.

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).

[35]Note that the estimate has been aggregated for the representative agent of the model.
[36]Multiplicative habit, in its typical forms, is not able to generate such a pattern of risk attitude.



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

FIGURE 6. INDIFFERENCE MAPS

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* $\frac{C u_C u_{CH}}{u_C u_H + C u_H u_{CC}} = 1$.[37]

PROOF: See Section I.4 of the online Appendix.

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 $-\frac{u_H}{u_C}$, $\frac{H u_{HH}}{u_H}$, $\frac{u_{CH}}{u_{HH}}$, 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 the third order.

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.[38]

PROPOSITION 4: *Under second-order elicitation,* $-\frac{u_H}{u_C} < \frac{(\rho+\theta)\Delta f}{\rho \Delta e + \theta \Delta f}$ *if the respondent chooses Universe One over Universe Two for a better future experience in the slope of indifference curve question.*[39]

PROOF: See Section I.5 of the online Appendix.

[37]One can derive a class of tests based on the homotheticity of multiplicative habit. The two tests here imply this class of tests.

[38]The resulting monthly spending graphs are in Figures A.15 (for $-\frac{u_H}{u_C}$), A.16 (for $\frac{H u_{HH}}{u_H}$), A.17 (for $\frac{u_{CH}}{u_{HH}}$), and A.18 (for $\frac{u_{CC}}{u_{HH}}$) of the online Appendix.

[39]Δe and Δf 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 E of the online Appendix because of this indirect interest in it.

My estimate for the slope of the indifference curve is 0.57 (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, about 57% of utility changes resulted from consumption changes is 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 $\frac{Hu_{HH}}{u_H}$, I present the respondents with a trade-off between the level and fluctuation of past spending.[38] I estimate $\frac{Hu_{HH}}{u_H}$ to be 7.50 (Table 4), which by the estimated $u_H < 0$, implies $u_{HH} < 0$.

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

PROOF: See Section I.6 of the online Appendix.

The elicitation of $\frac{u_{CH}}{u_{HH}}$ rests on inducing fluctuations in both future and past spending at the same time.[38]

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

PROOF: See Section I.7 of the online Appendix.

My estimate of $\frac{u_{CH}}{u_{HH}}$ is -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.

$\frac{u_{CC}}{u_{HH}}$ is about the trade-off between two sources of fluctuations, one from future spending and the other from past spending.[38]

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

PROOF: See Section I.8 of the online Appendix.

$\frac{u_{CC}}{u_{HH}}$ is estimated to be 3.87 (Table 4), consistent with the assumption of $u_{CC} < 0$.

TABLE 6—STATISTICS FOR TESTING ADDITIVE AND MULTIPLICATIVE HABITS

	MAP	Mean	Median	99% HPDI
$\frac{u_{CH}u_H}{u_{HH}u_C}$	0.51	0.50	0.50	[0.34, 0.68]
$\frac{u_{CH}u_C}{u_{CC}u_H}$	0.40	0.41	0.41	[0.26, 0.62]
$\frac{Hu_Hu_{CH}}{u_C(u_H+Hu_{HH})}$	0.44	0.44	0.44	[0.30, 0.60]
$\frac{Cu_Cu_{CH}}{u_H(u_C+Cu_{CC})}$	-0.24	-0.25	-0.25	[-0.35, -0.17]

With the estimates of $-\frac{u_H}{u_C}$, $\frac{Hu_{HH}}{u_H}$, $\frac{u_{CH}}{u_{HH}}$, and $\frac{u_{CC}}{u_{HH}}$, I calculate the left-hand-side statistics of the tests of additive and multiplicative habits. 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 survey evidence rejects both the additive and multiplicative habits with very high confidence.

It is worth emphasizing again that the evidence supports 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.

D. 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 one's own habit. In this section, I vary others' spending to see whether and by how much it affects habit.

The previous evidence for 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,[40] 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,

[40]See Section B of the online Appendix for proof.

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, I vary both others' and one's own past spending.[41]

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

PROOF: See Section I.9 of the online Appendix.

The 95% HPDI of the estimate of external habit mixture coefficient falls between 0 and 1, consistent with the existence of external habit formation.[42] The point estimate indicates that others' spending contributes to about 19% of one's own habit (Table 4).

E. 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 contemporaneous influence—peer effect—on one's own felicity function, $u(C, H, C_{\text{others}})$, and then elicit the ratio of $\frac{u_{C_{\text{others}}}}{u_H}$ by varying others' spending in both the past and the future.[43]

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

PROOF: See Section I.10 of the online Appendix.

My point estimate for $\frac{u_{C_{\text{others}}}}{u_H}$ is 1.12 (Table 4) and not significantly different from 1 at the 95% level, consistent with habit formation and peer effect having same-sized welfare impacts.

I draw two additional implications based on the significant negative sign of $u_{C_{\text{others}}}$ as implied from 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

[41]The resulting monthly spending graphs are in Figure A.19 of the online Appendix.

[42]In Section C of the online Appendix, I elicit 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.

[43]See Figure A.20 in the online Appendix for the resulting monthly spending graphs.

significantly negative means that peer effect exists after controlling for external habit. Second, peer effect is stronger than altruism. I 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.

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

Four clarifications merit discussion before proceeding. The first is that I focus on the causal channel that income changes happiness. Typical life experiences and studies exploiting exogenous variations (Frijters, Haisken-DeNew and Shields, 2004; Gardner and Oswald, 2007) support this view. Evidence aside, this causality motivated the discovery of the paradox[44] and is the most counterintuitive, interesting,[45] 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 et al., 2012; Perez-Truglia, 2020), I assume that the potential distinction between happiness and utility is of minimal effect on my 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 these two terms interchangeably.[46] 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[47] and lifetime utility for life satisfaction.

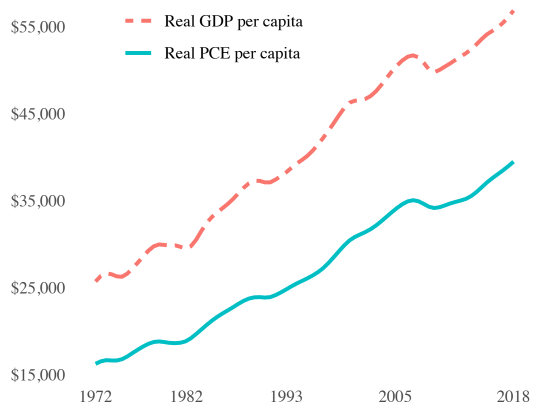
Because existing habit formation models are inconsistent with people's behavior, the credibility of any structural simulations under existing habit formation specifications is untenable. To still assess the explanation of the paradox by habit formation and peer effect, I conduct a semi-structural simulation based on my extensive survey evidence on these two phenomena. In particular, I specify that people are influenced by both internal and external habits as well as peer effect. Habit evolves according to

[44]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).

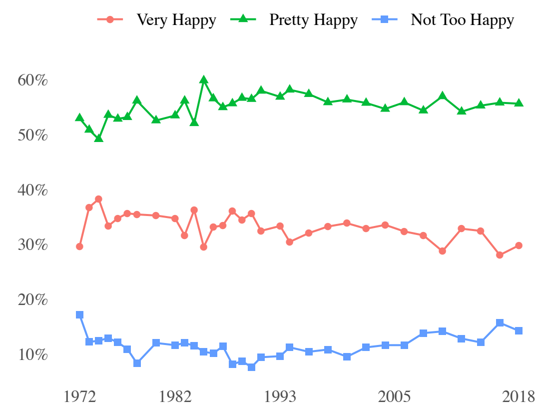
[45]This is evidenced by that the vast majority of speculative explanations of the paradox have focused on this channel.

[46]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.

[47]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

equation (1) with the habit depreciation rate and the external habit mixture coefficient calibrated to my estimates, 1.09 and 0.19, 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.[48] 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 $\frac{u_{C_{\text{others}}}}{u_C}$. My 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

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

which is consistent with my estimate at the 95% level (Table 4). The point estimate

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

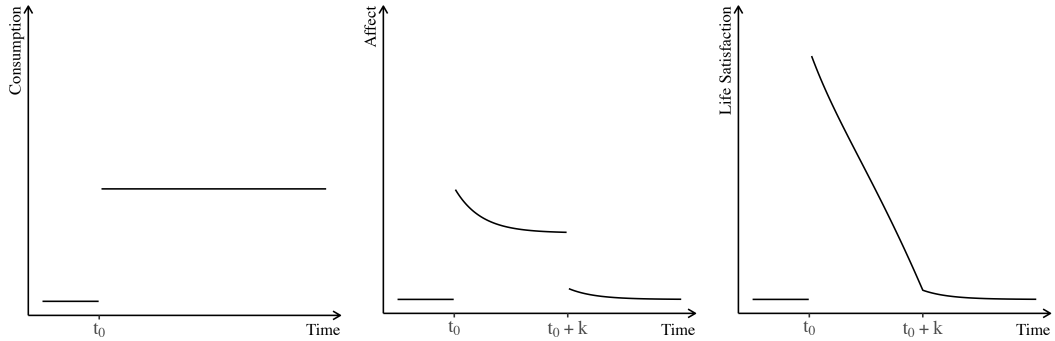
of the left-hand side of the above equation is less than -1, which, aside from statistical precision considerations, provides the potential for the explanation of the paradox to be consistent with the general improvement of happiness-altering non-income factors (Di Tella and MacCulloch, 2008) and with 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. As long as their sum is -1, the exact values of the two ratios only slightly affect the steady-state level of happiness and the convergence speed to the steady states, both of which do not alter the 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,[49] 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,[50] 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, let me label this

[49]Throughout the analysis of this section, I focus on first-order approximations, because there is no point going beyond it when the elicitation of $u_{C_{\text{others}}}/u_C$ is first-order. This is a good approximation when Δc is small, which I maintain here.

[50]I calibrate the time discount rate to 0.13, based on my estimate of this parameter. See Section E of the online Appendix for details. The value of this parameter does not affect the intuition.



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

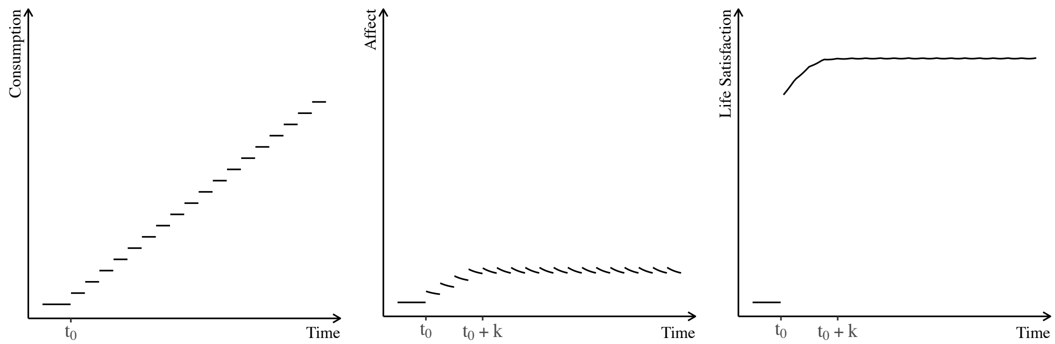
FIGURE 8. ONE-EPISODE GROWTH

pattern the wear-off effect: over time, habit formation and peer effect cancel out the happiness innovations brought by permanent changes of consumption.

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.[51] What is new is the dynamics of happiness: instead of eventually returning to its previous level, happiness gradually builds up and then plateaus. Again, for later reference, let me label these two patterns of the happiness dynamics 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[52] 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

[51]The discontinuities of the utility at the start of each year after t_0 are purely based on the simplifying assumption that consumption permanently increases at the start of each year after t_0, which is inessential. When the consumption changes are smoother, the discontinuities will be reduced. All my analysis carries through in such scenarios.

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



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

FIGURE 9. MULTI-EPISODE GROWTH

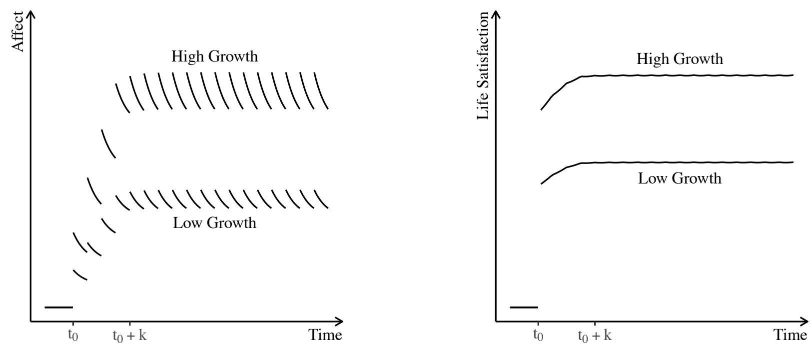
effect brought by the consumption growth at t_0 is in full swing for the first time.

Because the wear-off effect is proportional to Δc,[53] the transition and plateau effects 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. A prediction of the level effect is 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.[54] 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

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

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



(a) Affect (b) Life Satisfaction

FIGURE 10. LEVEL EFFECT

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

though income frequently fluctuates, it fluctuates around its trend. This trend growth determines the plateaued level of happiness, which underpins 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, let me draw an analogy—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 speed before you can reach the top. The elevation you reach represents happiness, and the number of stairs you run represents consumption. The escalator symbolizes the joint effect of habit formation and peer effect.

With this analogy, allow me 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

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

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 economy 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.

V. Conclusion

This paper provides a new and extensive set of micro evidence for habit formation through survey experiments. I find that people's spending behavior exhibits habit formation. The majority of the habit forms internally, while a small fraction (19%) of the habit forms externally. This implies that in terms of micro validity, internal habit formation is a better choice than external habit formation. Better still, a composition of internal and external habit formation with the estimated weights could potentially deliver a superior match with spending behavior.

Habit depreciates by 66% per year. The value of this parameter can significantly affect the performance of habit formation models. Future habit formation models should report model performance under this speed of habit depreciation.

Essentially all current habit formation models are rejected because their preference specifications fail to pass the four validity tests. 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 same-sized welfare impact as peer effect. Both external habit 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 eventually 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 still necessary to maintain the plateaued level of happiness.

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

REFERENCES

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

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

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

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

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

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

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

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

Boldrin, Michele, Lawrence J Christiano, and Jonas DM Fisher. 2001. "Habit Persistence, Asset Returns, and the Business Cycle." *American Economic Review*, 149–166.

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

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

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

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

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

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

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

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

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

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

Diewert, W. E. 1973. "Afriat and Revealed Preference Theory." *Review of Economic Studies*, 40(3): 419.

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

Dou, Winston Wei, Andrew W. Lo, Ameya Muley, and Harald Uhlig. 2017. "Macroeconomic Models for Monetary Policy: A Critical Review from a Finance Perspective." *SSRN 2899842*.

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

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

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

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

Firebaugh, Glenn, and Laura Tach. 2012. "Income, Age, and Happiness in America." *Social Trends in American Life*, 267–287. Princeton University Press.

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

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

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

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

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

Hara, Kotaro, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey 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, Tomas, Marek Rusnak, and Anna Sokolova. 2017. "Habit Formation in Consumption: A Meta-Analysis." *European Economic Review*, 95: 142–167.

Houthakker, Hendrik S, and Lester D Taylor. 1970. *Consumer Demand in the United States.* Harvard University Press.

Jha, Saumitra, and Moses Shayo. 2019. "Valuing Peace: The Effects of Financial Market Exposure on Votes and Political Attitudes." *Econometrica*, 87(5): 1561–1588.

Johnson, David, and John Ryan. 2018. "Amazon Mechanical Turk Workers Can Provide Consistent and Economically Meaningful Data." University Library of Munich, Germany MPRA Paper 88450.

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

Kaiser, Caspar F., and Maarten C. M. Vendrik. 2019. "Different Versions of the Easterlin Paradox: New Evidence for European Countries." *Economics of Happiness*, 27–55.

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

Kimball, Miles S, and Matthew D Shapiro. 2008. "Labor Supply: Are the Income and Substitution Effects Both Large or Both Small?" National Bureau of Economic Research Working Paper 14208.

Kimball, Miles S, and Robert Willis. 2006. "Happiness and Utility." Unpublished.

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

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

Kimball, Miles S., Fumio Ohtake, Daniel H. Reck, Yoshiro Tsutsui, and Fudong Zhang. 2015.

"Diminishing Marginal Utility Revisited." *SSRN 2592935*.

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

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

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

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

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

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

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

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

Pallais, Amanda. 2014. "Inefficient Hiring in Entry-Level Labor Markets." *American Economic Review*, 104(11): 3565–3599.

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

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

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

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

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

Sahm, Claudia R. 2007. "Stability of risk preference." Board of Governors of the Federal Reserve System (U.S.) Finance and Economics Discussion Series 2007-66.

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

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

Stevenson, Betsey, and Justin Wolfers. 2008. "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox." *Brookings Papers on Economic Activity*, 2008(1): 1–87.

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

Van Praag, Bernard M. S., and Paul Frijters. 1999. "The Measurement of Welfare and Well-Being: The Leyden Approach." *Well-Being: Foundations of Hedonic Psychology*, 413–433. Russell Sage Foundation.

Online Appendix for "Survey Evidence on Habit Formation"

Jiannan Zhou

March 31, 2020

Contents

List of Figures

List of Tables

A Aggregation

What do the elicited preference parameters of individual respondents tell us about the preference parameters of the representative agent? This question is of particular interest because almost all current models with habit formation assume a representative agent. In this section, I show that, given its existence, the representative agent'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 $\bar{C}_i = \bar{H}_i = \$5,000\ \forall i$, spending an extra dollar[1] while holding habit constant, brings the same marginal utility to every individual: $u_{i,C}\left(\bar{C}_i, \bar{H}_i\right) = u_{j,C}\left(\bar{C}_j, \bar{H}_j\right)\ \forall i, j$.

With the comparability of the individuals' utilities, the representativeness of the representative agent means 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 $H_R = H_i = B\ \forall i$ 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 does the homogenized heterogeneous-agent model, and hence the equality of $N u_R\left(C_R, H_R\right) = \sum_i u_i\left(C_i, H_i\right)$ 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.

[1] An epsilon dollar, to be exact.

A.1 Aggregation of Habit Depreciation Rate

Even though habit depreciation rate (θ) and habit (H) are mapped one to one at each instant of time[2] for any given consumption profile, there are infinitely many pairs of them that satisfy the representative agent's habit evolution ($\dot{H}_R = \theta_R (C_R - H_R)$), 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, different θ leads to different H before a steady state is reached, 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 aggregate habit depreciation rate (θ_R) and individual habit depreciation rate (θ_i), 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

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

and

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

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

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

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

[2]Before a steady state is reached.

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

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

where the second equality holds because of the independence between slope of indifference curve and habit depreciation rate. With $u_{R,H}(C_R, H_R) = \frac{1}{N}\sum_i u_{i,H}(C_i, H_i)$ (see Section A.2) and $u_{R,C}(C_R, H_R) = u_{i,C}(C_i, H_i)$ $\forall i$, I arrive at

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

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

A.2 Aggregation of Ratios of Utility Derivatives

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

Because $N u_R(C_R, H_R) = \sum_i u_i(C_i, H_i)$ 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:

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

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 baseline steady state.

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

bution of $-\frac{u_{i,H}(\bar{C}_i,\bar{H}_i)}{u_{i,C}(\bar{C}_i,\bar{H}_i)}$ is simply the distribution of $-u_{i,H}\left(\bar{C}_i,\bar{H}_i\right)$ scaled by the inverse of $\bar{u}$. Thus, $\mathbb{E}\left(u_{i,H}\right)$ can be calculated from

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

where $\mu_{-\frac{u_{i,H}}{u_{i,C}}}$ denotes the mean of the preference parameter $-\frac{u_{i,H}}{u_{i,C}}$ across individuals. Similar notations are used hereafter.

2. The distribution of $u_{i,HH}$ is the distribution of $\frac{Hu_{i,HH}}{u_{i,H}}$ multiplied by $u_{i,H}$ and scaled by the inverse of H. Because the parameters are independent,

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

3. Similarly,

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

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

and

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

4. With these I can calculate the representative agent's parameters:

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

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

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

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

and

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

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

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

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

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

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

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

B Observational Equivalence of Linear and Nonlinear Habit Evolutions

In this section, I show that the model with the linear habit evolution (Model L below) and models with nonlinear habit evolutions (Model N below) are observationally equivalent (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.[3] By Definition 2, their monotonicities are entangled with respect to time.[4]

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, $H_t = a\,(t|C_0, H_0)$ and $\mathcal{H}_t = b\,(t|C_0, \mathcal{H}_0)$, where $a\,(\cdot)$ and $b\,(\cdot)$ are monotonic functions of t for $0 \leq t \leq T$ and flat for $t > T$.

Because

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

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 $\mathcal{H}_t$:

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

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

[3]For example, the (nonlinear) geometric habit evolution, as in Kozicki and Tinsley (2002), synchronizes with the linear habit evolution of Model L for any given spending path.

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

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. To see this, 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

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

and

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

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 the optimal solution, 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. $\square$

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 model; the proof of the equivalence result in these situations is straightforward because other people's spending is exogenous to the agent.

C Nonparametric Elicitation of Existence of External Habit Formation

The paper elicits the existence of external habit formation with a parametric habit evolution. In this section, I achieve such elicitation nonparametrically.

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

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

for $\Delta\mathbf{C}_{\text{others}} > 0$, where the boldface denotes a historical path.

To elicit the sign of Q_{EH}, I can use the same survey question that elicits the external habit mixture coefficient.[5]

Proposition 11. *Given the existence of internal habit formation, $Q_{EH} < 0$ if the respondent chooses Universe One over Universe Two for a better future experience in the external habit formation question.*

Proof. See Section I.12. □

Choosing Universe Two over Universe One does not necessarily imply $Q_{EH} \geq 0$. As long as external habit formation is weak enough relative to internal habit formation, including the case that external habit formation does not exist, a respondent would always prefer Universe Two for a better future experience. In other words, when a respondent chooses Universe Two, the survey question does not contain extra information that allows for separating the different possibilities for the sign of Q_{EH}. Therefore, I check every possibility. If always choosing Universe Two means

[5] The monthly spending graphs are in Figure A.19.

Table A.1: Response Frequency (Percentage) for External Habit Formation Question

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

Table A.2: Estimate of Existence of External Habit Formation

	MAP	Mean	Median	95% HPMI	99% HPMI
$\text{sgn}(Q_{EH})$ (strict)	-1.00	-0.94	-1.00	[-1.00, -1.00]	[-1.00, 1.00]
$\text{sgn}(Q_{EH})$ (lenient)	-1.00	-1.00	-1.00	[-1.00, -1.00]	[-1.00, -1.00]

$Q_{EH} < 0$ (most lenient), all the responses imply the existence of external habit formation. If always choosing Universe Two means $Q_{EH} > 0$ (strict), it is least likely to find evidence for external habit formation, whereas if always choosing Universe Two means $Q_{EH} = 0$ (lenient), it is relatively more likely to find external habit formation.

The response frequency in Table A.1 indicates that the response of the average respondent agrees with the existence of external habit formation. My estimate under the strict case shows that $Q_{EH} < 0$ at the 95% level, whereas the opposite sign emerges at 99% or higher levels (Table A.2). The lenient scenario finds evidence of a negative sign for Q_{EH} at the 99% level. Overall, external habit formation exists at least at the 95% level, consistent with the estimate of the external habit mixture coefficient in Section III.D of the paper.

D Effects of Habit Depreciation Speed on Other Results of Campbell and Cochrane (1999)

Table A.3 replicates part of Table 5 of Campbell and Cochrane (1999) and compares their model's performance on the predictability of dividend yield on expected return under different habit depreciation speeds. Under my estimated annual habit depreciation factor, the regression coefficients and R^2s are larger than their postwar counterparts. Interestingly, R^2s start to decrease after the second year.

Table A.3: Effect of Habit Depreciation Speed in Campbell and Cochrane (1999): Long-Run Regression

			$\theta = 0.11$				$\theta = 0.66$			
	Postwar sample		Consumption claim		Dividend claim		Consumption claim		Dividend claim	
Horizon (years)	10 × coefficient	R^2	10 × coefficient	R^2	10 × coefficient	R^2	10 × coefficient	R^2	10 × coefficient	R^2
1	-2.6	.18	-2.0	.13	-1.9	.08	-9.5	.59	-9.6	.52
2	-4.3	.27	-3.7	.23	-3.6	.14	-12.9	.71	-12.9	.60
3	-5.4	.37	-5.1	.32	-5.0	.19	-14.0	.68	-13.9	.56
5	-9.0	.55	-7.5	.46	-7.3	.26	-14.7	.56	-14.7	.45
7	-12.1	.65	-9.4	.55	-9.2	.30	-14.4	.45	-14.3	.35

Notes: Long-horizon regressions of log excess stock returns on the log price/dividend ratio in historical and simulated data. θ is the annual habit depreciation factor. $\theta = 0.11$ is Campbell and Cochrane (1999)'s calibration, and $\theta = 0.66$ my estimate.

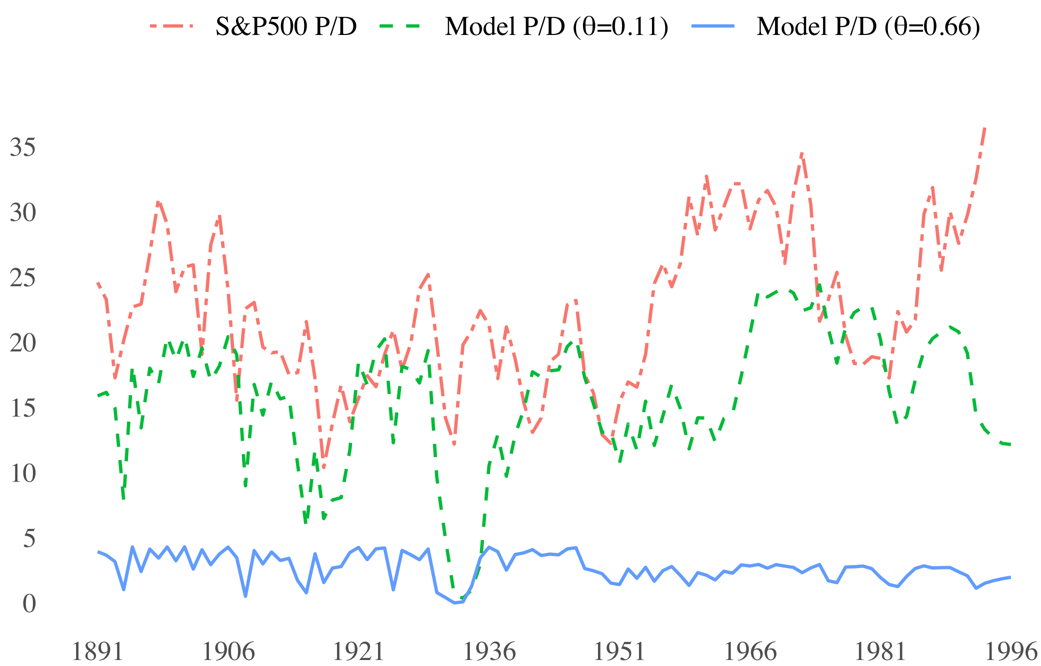


Figure A.1: Historical Price-Dividend Ratio and Model Predictions Based on Consumption History

Figure A.1 replicates Figure 9 of Campbell and Cochrane (1999), with an additional curve showing the model-predicted price-dividend ratio under my estimated habit depreciation factor of 0.66. Even though the fluctuations retain about the same pattern, a higher habit depreciation speed makes the price-dividend ratio much smaller. This is consistent with the fact that a higher value of θ induces the model agent to be more risk averse and, therefore, to require a higher return to take on the same level of consumption risk.

E Elicitation of Time Discount Rate

To elicit the time discount rate, I increase spending in the next year and the year after next year. The resulting survey question has monthly spending graphs as in Figure A.2.

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

Proof. See Section I.13. □

Estimation pins down a value of 0.13 for the time discount rate (Table A.4).

Table A.4: Estimate of Time Discount Rate

	MAP	Mean	Median	95% HPDI
ρ	0.13	0.13	0.13	[0.03, 0.23]

F Robustness Checks

In this section, I discuss the robustness of my estimates in the paper to departures from certain underlying assumptions.

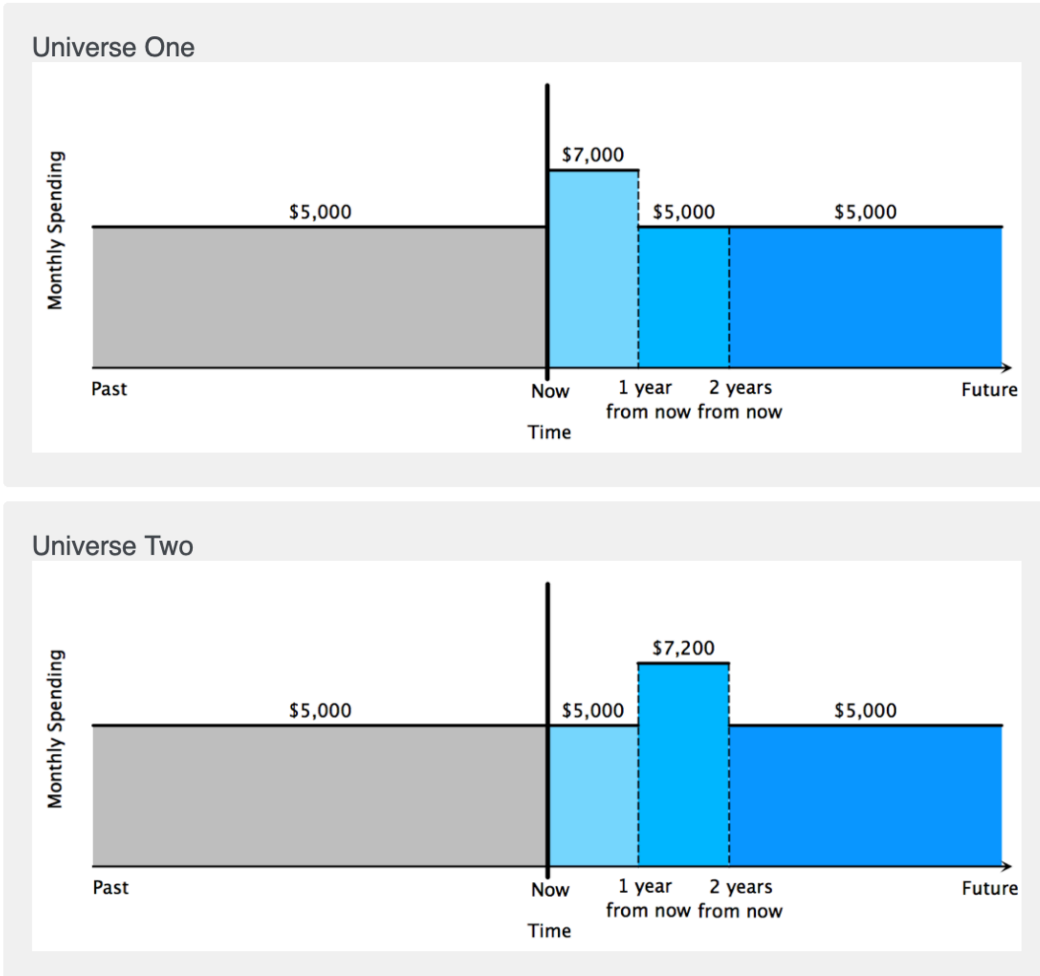


Figure A.2: Monthly Spending Graphs—Time Discount Rate

F.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 my estimates (Table A.5). 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.[6]

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 reassuring that the estimates do not vary with the demographics of the respondents and, therefore, with their 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.

Table A.5: Demographic Effects on Parameter Estimates

	Omitted category	Age	Gender	Household size	Household income
Habit depreciation rate	1.21	0.00	-0.17	0.09	-0.04
	[0.76, 1.63]	[-0.02, 0.02]	[-0.60, 0.25]	[-0.05, 0.25]	[-0.09, 0.02]
External habit mixture coefficient	0.20	0.00	0.02	-0.02	-0.02
	[0.01, 0.36]	[-0.01, 0.01]	[-0.15, 0.20]	[-0.09, 0.04]	[-0.04, 0.01]
$-u_H/u_C$	0.65	0.00	-0.09	0.05	0.01
	[0.41, 0.89]	[-0.01, 0.01]	[-0.32, 0.13]	[-0.03, 0.13]	[-0.01, 0.05]
Hu_{HH}/u_H	6.72	0.00	1.39	-0.51	0.22
	[5.45, 8.34]	[-0.06, 0.06]	[0.17, 1.50]	[-0.97, -0.04]	[0.03, 0.44]
u_{CH}/u_{HH}	-0.91	0.00	0.11	0.02	0.03
	[-1.23, -0.60]	[-0.01, 0.01]	[-0.18, 0.42]	[-0.08, 0.13]	[-0.01, 0.06]
u_{CC}/u_{HH}	4.43	0.04	-0.78	-0.37	-0.06
	[3.05, 5.74]	[-0.02, 0.09]	[-1.50, 0.50]	[-0.86, 0.08]	[-0.23, 0.12]
$u_{C_{\text{others}}}/u_H$	0.75	-0.01	0.09	0.21	0.06
	[0.02, 1.52]	[-0.05, 0.01]	[-0.64, 0.82]	[-0.04, 0.48]	[-0.02, 0.16]
Time discount rate	0.06	0.00	0.06	0.01	0.00
	[-0.19, 0.25]	[-0.01, 0.00]	[-0.14, 0.28]	[-0.06, 0.09]	[-0.03, 0.03]

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

[6] The adjusted probability of the estimates being zero or less probable than zero under the Holm algorithm are 0.90, 0.89, and 0.38, respectively.

F.2 Finite Horizon

I assume infinite horizon in the general model, as do almost all current habit formation models in the literature. To investigate the effect of this assumption on my results, I rederive all the elicitations of the preference parameters under finite horizon and find that the changes are minimal: no change for the elicitations of some parameters and tiny changes for the rest.[7] As a result, estimation under the finite-horizon elicitations gives almost identical estimates to those based on the infinite horizon (Table A.6).

TABLE A.6: FINITE-HORIZON ESTIMATES

	MAP	Mean	Median	HPDI
Habit depreciation rate	1.09	1.08	1.08	[0.89, 1.29]
External habit mixture coefficient	0.19	0.18	0.18	[0.09, 0.27]
$-u_H/u_C$	0.57	0.57	0.57	[0.46, 0.68]
Hu_{HH}/u_H	7.50	7.53	7.51	[6.63, 8.46]
u_{CH}/u_{HH}	-0.88	-0.88	-0.88	[-1.02, -0.73]
u_{CC}/u_{HH}	3.87	3.80	3.81	[3.10, 4.49]
$u_{C_{\text{others}}}/u_H$	1.12	1.12	1.12	[0.77, 1.48]
$u_{C_{\text{others}}}/u_C$	-0.62	-0.64	-0.63	[-0.88, -0.41]
$u_{CH}u_H/u_{HH}u_C$	0.51	0.50	0.50	[0.34, 0.68]
$u_{CH}u_C/u_{CC}u_H$	0.40	0.41	0.41	[0.26, 0.62]
$Hu_Hu_{CH}/u_C\,(u_H + Hu_{HH})$	0.44	0.44	0.44	[0.30, 0.60]
$Cu_Cu_{CH}/u_H\,(u_C + Cu_{CC})$	-0.24	-0.25	-0.25	[-0.35, -0.17]
Time discount rate	0.13	0.13	0.13	[0.03, 0.23]

Note: The estimates in this table are based on a time horizon of 30 years in the future, because the survey instructs the respondents: "If easier, think of ... 'Future' as the next 30 years." 99% HPDIs are reported for the four statistics for testing additive and multiplicative habits. 95% HPDIs are reported for other parameters.

[7] 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.

F.3 Additional Attention Checks

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

At the end of the survey, I quizzed the respondents again on the basic hypothetical situation of the survey. There were 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 my estimates (Table A.7), except that external habit formation no longer exists in the strict case, although it still exists in the lenient case.

Table A.7: Estimates Using Subsample of No End-of-Survey Quiz Mistake

	MAP	Mean	Median	HPDI/HPMI
Habit depreciation rate	1.09	1.10	1.10	[0.84, 1.35]
External habit mixture coefficient	0.10	0.11	0.11	[0.00, 0.21]
$-u_H/u_C$	0.65	0.65	0.65	[0.51, 0.80]
Hu_{HH}/u_H	8.15	8.26	8.23	[6.96, 9.59]
u_{CH}/u_{HH}	-0.82	-0.82	-0.82	[-1.01, -0.63]
u_{CC}/u_{HH}	3.70	3.71	3.71	[2.79, 4.68]
$u_{C_{\text{others}}}/u_H$	1.06	1.06	1.05	[0.60, 1.50]
$u_{C_{\text{others}}}/u_C$	-0.65	-0.69	-0.68	[-1.03, -0.37]
$u_{CH}u_H/u_{HH}u_C$	0.53	0.53	0.53	[0.32, 0.77]
$u_{CH}u_C/u_{CC}u_H$	0.33	0.35	0.34	[0.19, 0.59]
$Hu_Hu_{CH}/u_C\left(u_H + Hu_{HH}\right)$	0.47	0.47	0.47	[0.28, 0.68]
$Cu_Cu_{CH}/u_H\left(u_C + Cu_{CC}\right)$	-0.22	-0.24	-0.23	[-0.37, -0.14]
$\operatorname{sgn}\left(Q_H\right)$	-1.00	-1.00	-1.00	[-1.00, -1.00]
$\operatorname{sgn}\left(Q_{EH}\right)$ (strict)	-1.00	-0.01	-1.00	0.49
$\operatorname{sgn}\left(Q_{EH}\right)$ (lenient)	-1.00	-1.00	-1.00	[-1.00, -1.00]
Time discount rate	0.09	0.09	0.09	[-0.04, 0.22]

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

I asked demographic questions in both waves. Within the relatively moderate amount of time that separated the two waves, the demographics should not have changed. In other words, the 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. My estimates are essentially unchanged (Table A.8) except, again, that the evidence for the existence of external habit formation is weaker in the strict case.

Table A.8: Estimates Using Subsample of No Demographic Mistake

	MAP	Mean	Median	HPDI/HPMI
Habit depreciation rate	1.09	1.10	1.10	[0.83, 1.37]
External habit mixture coefficient	0.13	0.13	0.13	[0.00, 0.24]
$-u_H/u_C$	0.59	0.61	0.61	[0.45, 0.76]
Hu_{HH}/u_H	8.45	8.48	8.47	[7.30, 9.92]
u_{CH}/u_{HH}	-0.86	-0.86	-0.86	[-1.05, -0.67]
u_{CC}/u_{HH}	3.46	3.47	3.47	[2.55, 4.41]
$u_{C_{\text{others}}}/u_H$	0.97	0.96	0.96	[0.50, 1.43]
$u_{C_{\text{others}}}/u_C$	-0.59	-0.58	-0.58	[-0.90, -0.26]
$u_{CH}u_H/u_{HH}u_C$	0.51	0.52	0.52	[0.31, 0.77]
$u_{CH}u_C/u_{CC}u_H$	0.38	0.42	0.41	[0.22, 0.73]
$Hu_Hu_{CH}/u_C\,(u_H + Hu_{HH})$	0.46	0.47	0.47	[0.27, 0.68]
$Cu_Cu_{CH}/u_H\,(u_C + Cu_{CC})$	-0.25	-0.27	-0.26	[-0.43, -0.15]
sgn (Q_H)	-1.00	-1.00	-1.00	[-1.00, -1.00]
sgn (Q_{EH}) (strict)	-1.00	-0.44	-1.00	0.28
sgn (Q_{EH}) (lenient)	-1.00	-1.00	-1.00	[-1.00, -1.00]
Time discount rate	0.11	0.10	0.10	[-0.04, 0.22]

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

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,[8] 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 credible intervals inflate because of the much smaller sample size, the estimates remain in line with my baseline estimates (Table A.9).

[8] See Section E of this appendix.

Table A.9: Estimates Using Subsample of Same Past Experience

	MAP	Mean	Median	HPDI/HPMI
Habit depreciation rate	1.17	1.15	1.15	[0.72, 1.58]
External habit mixture coefficient	0.15	0.18	0.17	[0.00, 0.38]
$-u_H/u_C$	0.51	0.51	0.51	[0.31, 0.70]
Hu_{HH}/u_H	9.77	8.98	9.08	[7.67, 10.00]
u_{CH}/u_{HH}	-0.73	-0.71	-0.71	[-0.95, -0.46]
u_{CC}/u_{HH}	4.04	3.99	3.99	[2.51, 5.54]
$u_{C_{\text{others}}}/u_H$	0.92	0.91	0.91	[0.22, 1.63]
$u_{C_{\text{others}}}/u_C$	-0.42	-0.46	-0.45	[-0.87, -0.08]
$u_{CH}u_H/u_{HH}u_C$	0.34	0.36	0.36	[0.14, 0.63]
$u_{CH}u_C/u_{CC}u_H$	0.31	0.38	0.36	[0.12, 0.85]
$Hu_Hu_{CH}/u_C\,(u_H + Hu_{HH})$	0.30	0.32	0.32	[0.12, 0.56]
$Cu_Cu_{CH}/u_H\,(u_C + Cu_{CC})$	-0.18	-0.20	-0.19	[-0.40, -0.08]
sgn (Q_H)	-1.00	-1.00	-1.00	[-1.00, -1.00]
sgn (Q_{EH}) (strict)	-1.00	-0.52	-1.00	0.24
sgn (Q_{EH}) (lenient)	-1.00	-1.00	-1.00	[-1.00, -1.00]
Time discount rate	0.04	0.02	0.03	[-0.20, 0.25]

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

Finally, I use a measure of response consistency across the waves as an attention check. Considering that this attention check involves more speculation, I eliminate 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 credible intervals, but, again, the estimates are not significantly different (Table A.10), and my results remain robust.

Table A.10: Estimates Using Subsample of No Polar Response

	MAP	Mean	Median	HPDI/HPMI
Habit depreciation rate	1.26	1.29	1.29	[0.78, 1.84]
External habit mixture coefficient	0.06	0.15	0.13	[0.00, 0.34]
$-u_H/u_C$	0.53	0.53	0.53	[0.29, 0.77]
Hu_{HH}/u_H	8.92	8.64	8.72	[7.15, 10.00]
u_{CH}/u_{HH}	-0.78	-0.78	-0.78	[-1.02, -0.54]
u_{CC}/u_{HH}	3.07	3.11	3.11	[2.04, 4.11]
$u_{C_{\text{others}}}/u_H$	0.51	0.49	0.49	[-0.04, 1.02]
$u_{C_{\text{others}}}/u_C$	-0.22	-0.26	-0.24	[-0.58, 0.03]
$u_{CH}u_H/u_{HH}u_C$	0.41	0.41	0.41	[0.13, 0.74]
$u_{CH}u_C/u_{CC}u_H$	0.46	0.52	0.48	[0.17, 1.31]
$Hu_Hu_{CH}/u_C\,(u_H + Hu_{HH})$	0.36	0.37	0.36	[0.12, 0.67]
$Cu_Cu_{CH}/u_H\,(u_C + Cu_{CC})$	-0.26	-0.28	-0.27	[-0.53, -0.12]
$\text{sgn}\,(Q_H)$	-1.00	-1.00	-1.00	[-1.00, -1.00]
$\text{sgn}\,(Q_{EH})$ (strict)	1.00	0.21	1.00	0.60
$\text{sgn}\,(Q_{EH})$ (lenient)	-1.00	-1.00	-1.00	[-1.00, -1.00]
Time discount rate	0.12	0.08	0.09	[-0.19, 0.34]

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

F.4 Response Bias and Error of Nonzero and Wave-Varying Mean

In the statistical model, I assume 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,[9] the joint distribution of parameter $\tilde{x}$ for individual i in both waves becomes

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

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

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 (Table A.11). This implies that my estimates are robust to the nonzero-wave-varying-mean response biases and errors, which is confirmed by Table A.12.

Table A.11: Means of Response Biases and Errors in Wave One

μ_ε of	MAP	Mean	Median	95% HPDI
Habit depreciation rate	-0.01	-0.01	-0.01	[-0.23, 0.22]
External habit mixture coefficient	-0.04	-0.03	-0.03	[-0.12, 0.07]
$-u_H/u_C$	-0.02	-0.02	-0.02	[-0.11, 0.06]
Hu_{HH}/u_H	-0.84	-1.05	-1.02	[-2.51, 0.29]
u_{CH}/u_{HH}	0.07	0.07	0.07	[-0.07, 0.23]
u_{CC}/u_{HH}	-0.39	-0.39	-0.39	[-1.11, 0.36]
$u_{C_{\text{others}}}/u_H$	-0.09	-0.11	-0.11	[-0.49, 0.27]
Time discount rate	-0.05	-0.04	-0.04	[-0.16, 0.07]

Note: μ_ε denotes mean of response error in wave one.

TABLE A.12: ESTIMATES UNDER NONZERO AND WAVE-VARYING MEANS OF RESPONSE BIASES AND ERRORS

	MAP	Mean	Median	HPDI/HPMI
Habit depreciation rate	1.09	1.09	1.09	[0.85, 1.33]
External habit mixture coefficient	0.21	0.20	0.20	[0.08, 0.32]
$-u_H/u_C$	0.60	0.60	0.60	[0.49, 0.73]
Hu_{HH}/u_H	8.38	8.44	8.44	[7.20, 9.99]
u_{CH}/u_{HH}	-0.89	-0.91	-0.91	[-1.07, -0.75]
u_{CC}/u_{HH}	4.05	4.01	4.01	[3.11, 4.93]
$u_{C_{\text{others}}}/u_H$	1.12	1.11	1.11	[0.69, 1.53]
$u_{C_{\text{others}}}/u_C$	-0.64	-0.67	-0.66	[-0.97, -0.39]
$u_{CH}u_H/u_{HH}u_C$	0.54	0.55	0.55	[0.37, 0.75]
$u_{CH}u_C/u_{CC}u_H$	0.36	0.39	0.38	[0.23, 0.60]
$Hu_Hu_{CH}/u_C\,(u_H + Hu_{HH})$	0.48	0.49	0.49	[0.33, 0.67]
$Cu_Cu_{CH}/u_H\,(u_C + Cu_{CC})$	-0.23	-0.24	-0.24	[-0.36, -0.16]
sgn (Q_H)	-1.00	-1.00	-1.00	[-1.00, -1.00]
sgn (Q_{EH}) (strict)	-1.00	-0.93	-1.00	95%: [-1.00, -1.00] 99%: [-1.00, 1.00]
sgn (Q_{EH}) (lenient)	-1.00	-1.00	-1.00	[-1.00, -1.00]
Time discount rate	0.17	0.17	0.17	[0.03, 0.31]

Notes: 99% HPDIs or HPMIs are reported for the seven statistics for testing additive and multiplicative habits and existence of habit formation. 95% HPDIs are reported for other parameters. In one case, an additional HPMI of relevant level is shown to help determine the significance level.

G Additional Survey Details

The survey starts with a consent form detailing the purpose, procedure, potential benefits, payment for participation, confidentiality, withdrawal procedure, and investigator contact information. Consent was obtained from all the respondents.

G.1 Instructions

Instruction One

After consent is obtained, the survey continues with an instruction on the basic hypothetical situation discussed in Section I.B of the paper. Two practice questions follow to test the respondents' understanding of the hypothetical situation:

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

If the respondent makes a mistake in the practice questions, they need to go over the instruction again and redo the practice questions. A maximum of three attempts of the practice questions is allowed.

Instruction Two

After the practice questions, the survey continues with instructions on reading the monthly spending graphs:

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

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

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

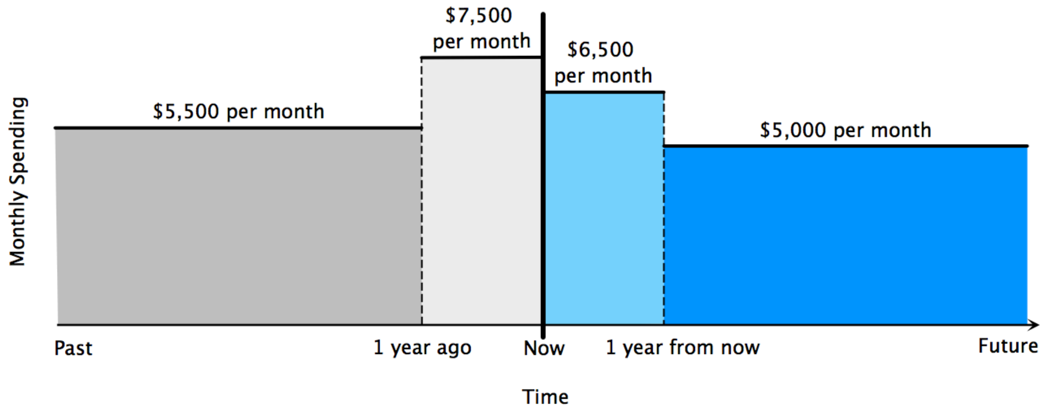


FIGURE A.3: INSTRUCTION—MORE DETAILED MONTHLY SPENDING GRAPH

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

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

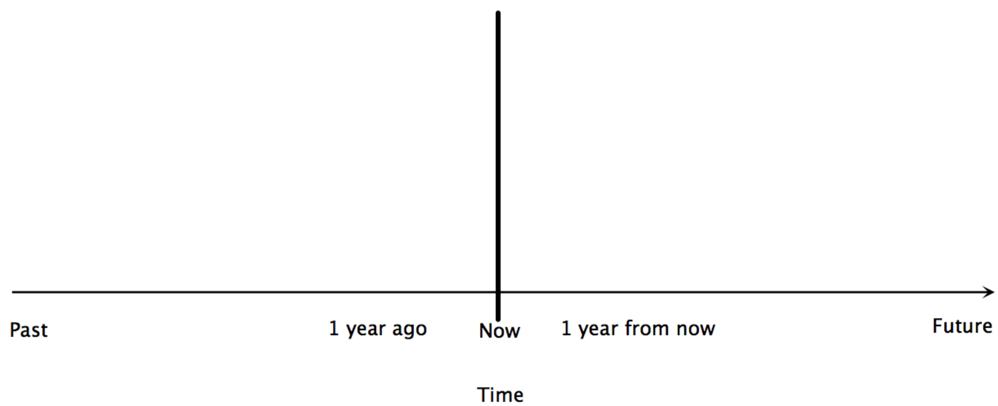


Figure A.4: Instruction—Timeline

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

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

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

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

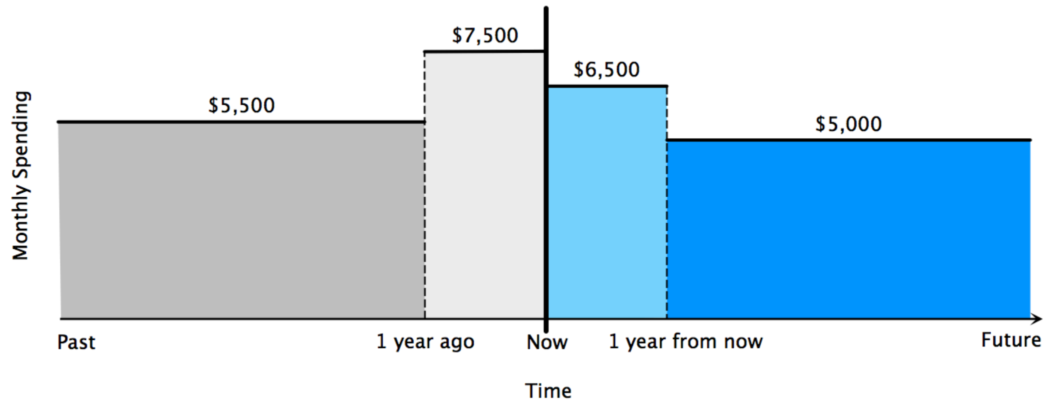


FIGURE A.5: INSTRUCTION—MONTHLY SPENDING GRAPH 2

you spent/spend

- $5,500 per month in the 'past' until '1 year ago';
- $7,500 per month from '1 year ago' until 'now' (or in the 'past year');
- $6,500 per month from 'now' to '1 year from now' (or in the 'next year');
- $5,000 per month from '1 year from now' onward.

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

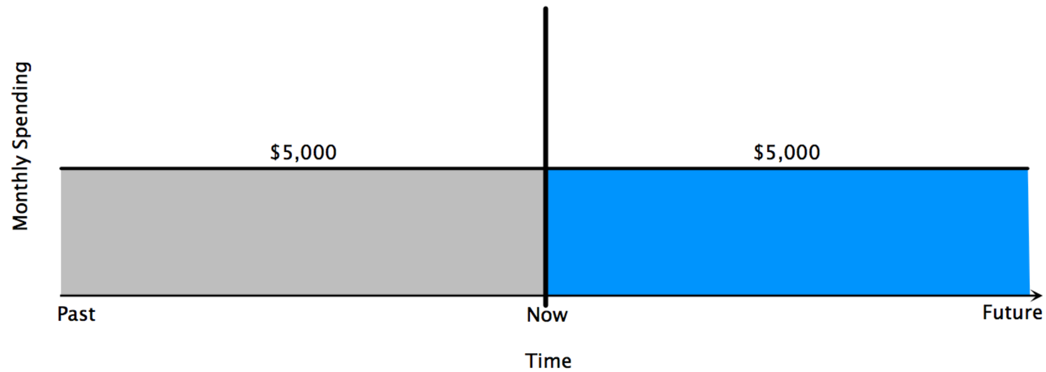


Figure A.6: Instruction—Monthly Spending Graph 3

while in Universe Two your monthly spending is

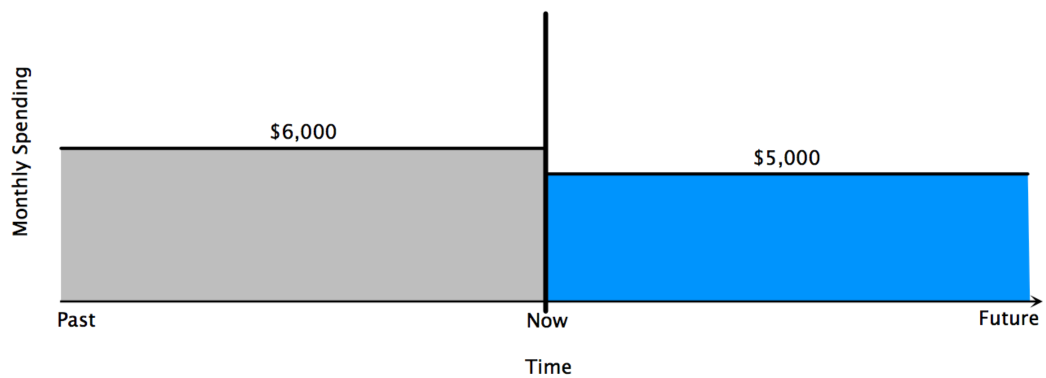


Figure A.7: Instruction—Monthly Spending Graph 4

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

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

Then, the respondents received a set of practice questions testing their understanding of the instructions.

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

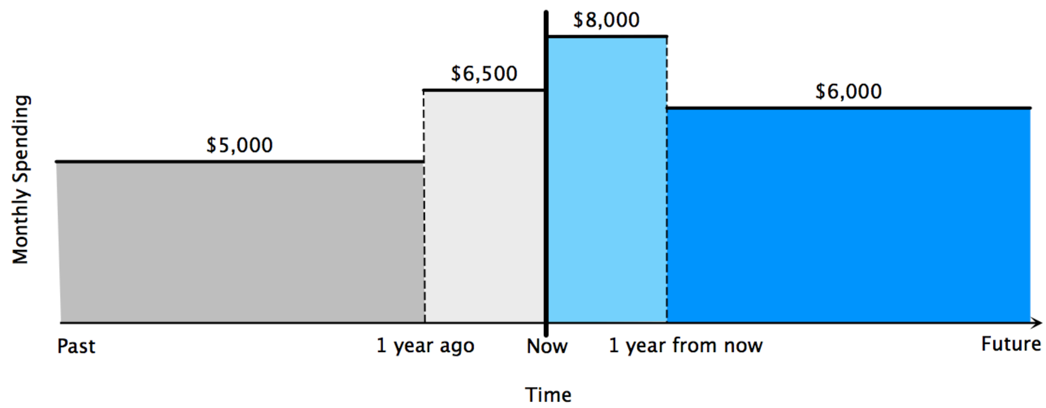


Figure A.8: Instruction—Monthly Spending Graph 5

> How much will you spend per month in the next year?
>
> - $5,000
> - $6,000
> - $6,500
> - $8,000
>
> How much did you spend per month from 'as far back as you can remember in the past' until '1 year ago'?
>
> - $5,000
> - $6,000
> - $6,500
> - $8,000

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

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

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

- Universe One

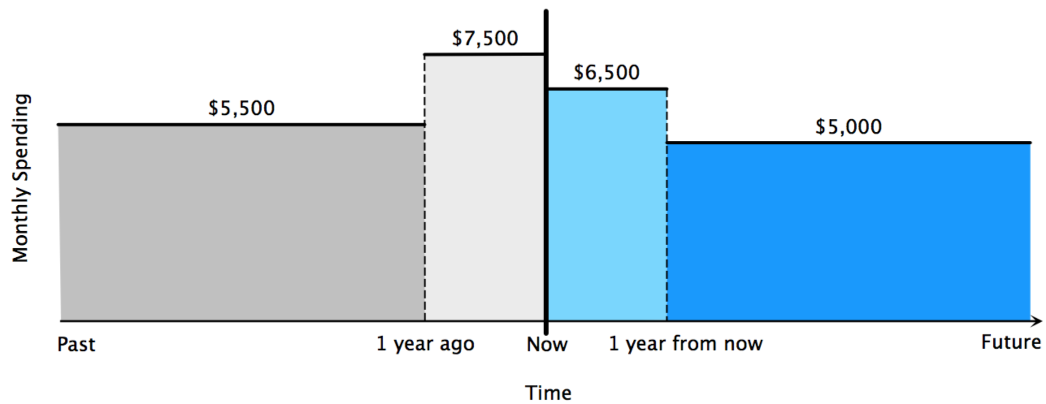


FIGURE A.5: INSTRUCTION—MONTHLY SPENDING GRAPH 2 (REPEATED FROM PAGE 30)

- Universe Two

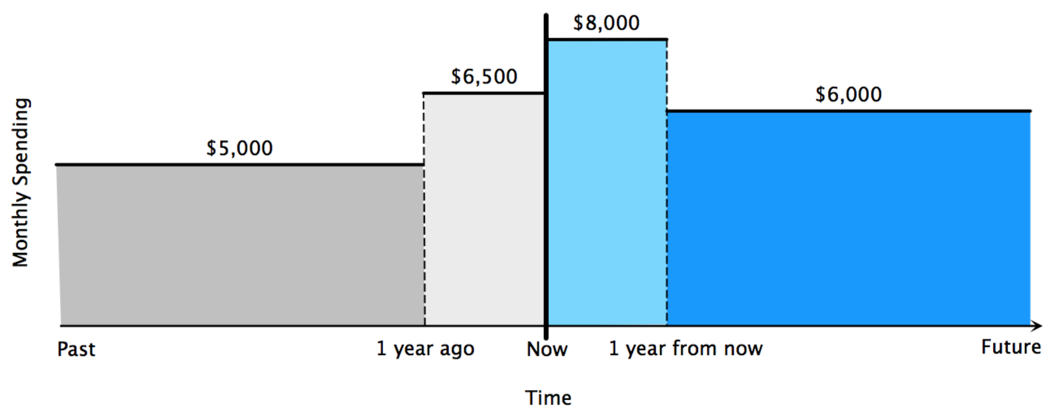


FIGURE A.8: INSTRUCTION—MONTHLY SPENDING GRAPH 5 (REPEATED FROM PAGE 32)

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

- Universe One

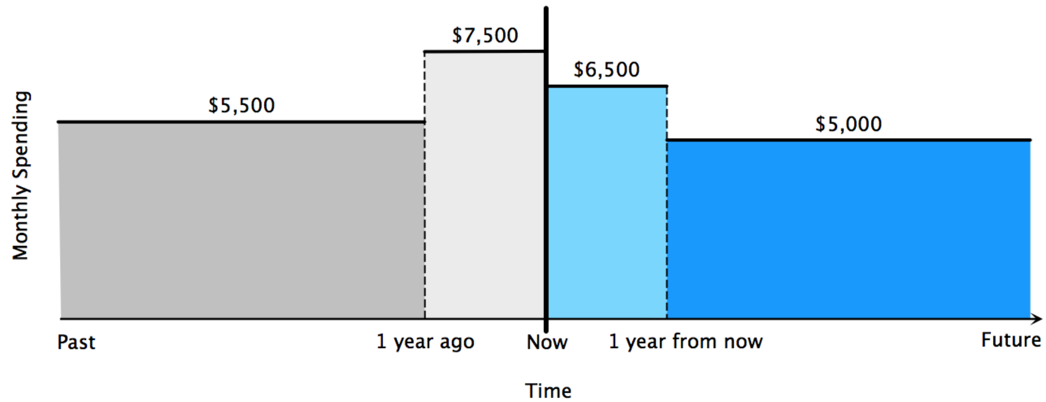


FIGURE A.5: INSTRUCTION—MONTHLY SPENDING GRAPH 2 (REPEATED FROM PAGE 30)

- Universe Two

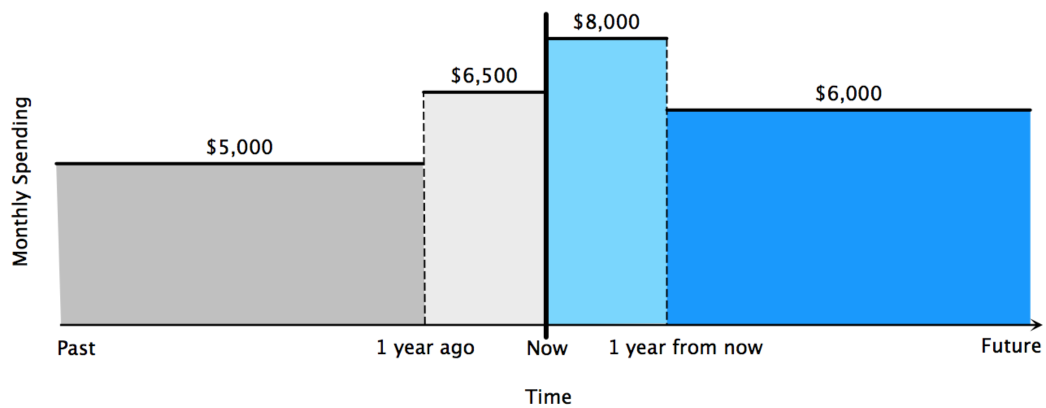


FIGURE A.8: INSTRUCTION—MONTHLY SPENDING GRAPH 5 (REPEATED FROM PAGE 32)

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

- $0
- $500
- $1,000

- $5,500

Final Instruction

A final instruction inoculates the respondents against the seeming repetitiveness of follow-up questions and encourages effort with attention checks and a lottery reward. An opportunity to review previous instructions was also presented.

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

G.2 Option Appearances

To reduce response errors, as mentioned in the paper, the survey darkens slightly the background of an option when respondents' mice hover over it and darkens completely the background of the options they select (Figure A.13).

G.3 Core Modules

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

The flow of each core module is

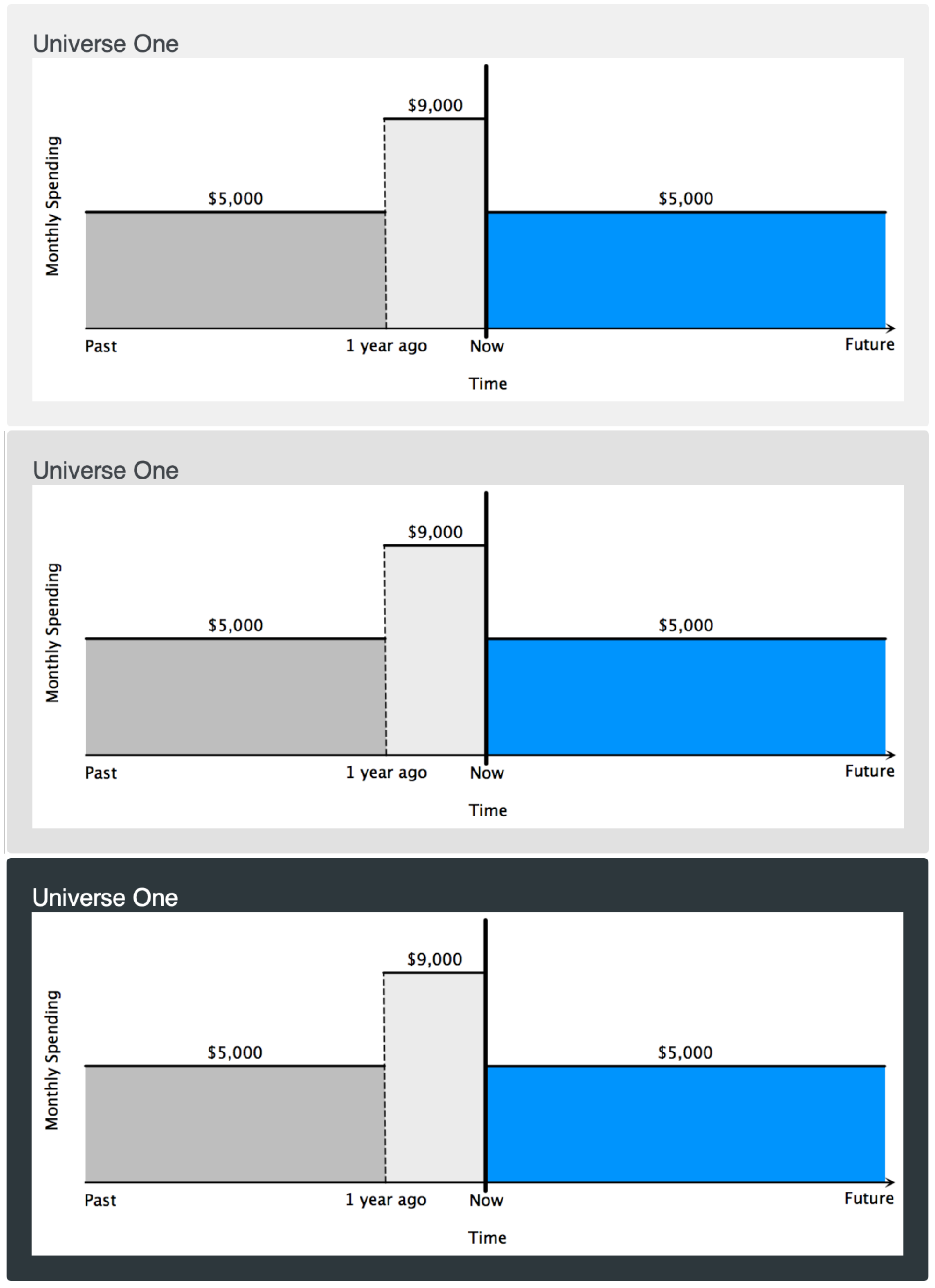


Figure A.13: Option Appearances (Top to Bottom: Initial, Hover, Click)

1. Ask the respondent to choose a universe that brings them a better past experience (e.g., Figure A.14);

2. Ask the respondent to choose a universe that brings them a better future experience (e.g., Figure 2 in the paper);

3. Present the respondent with one to two follow-up questions of monthly spending graphs that vary slightly from the initial question (cf. Figure 5 in the paper).

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

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

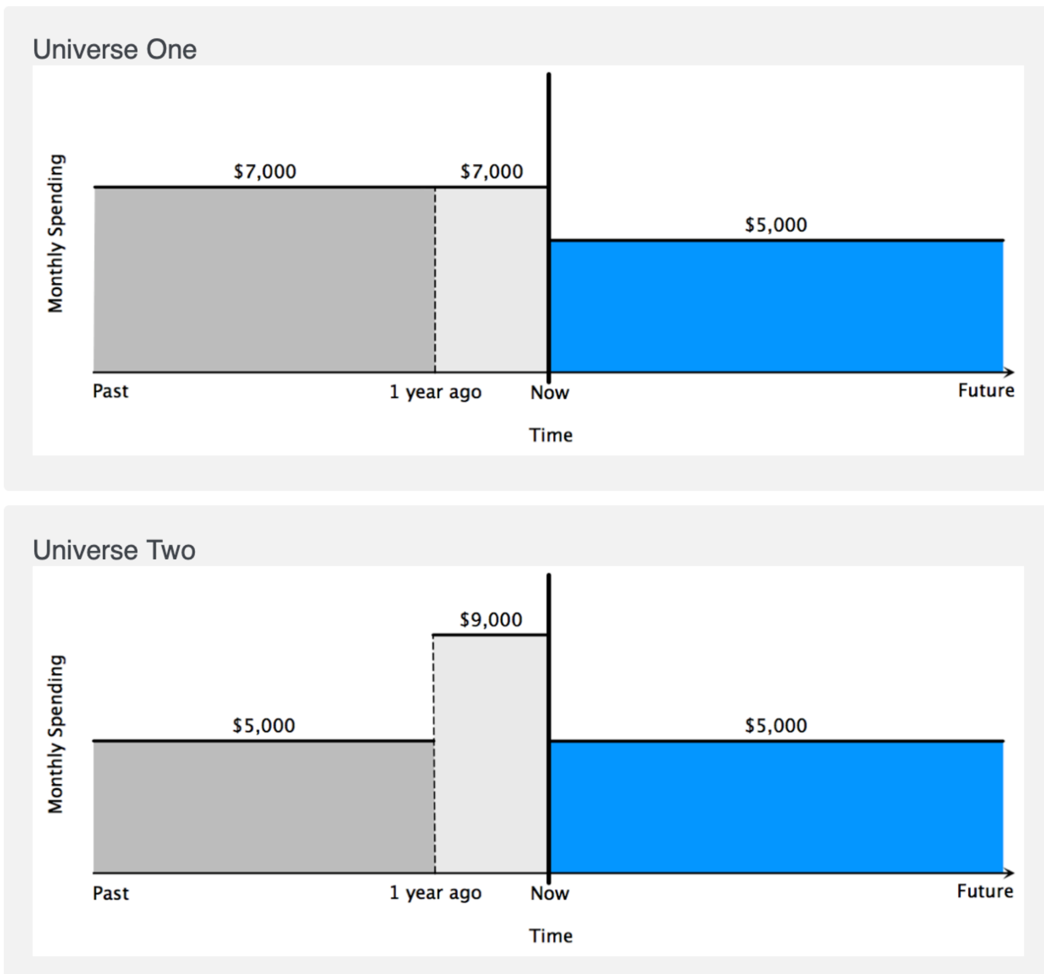


FIGURE A.14: A TYPICAL SURVEY QUESTION (PAST EXPERIENCE)

The main element of the survey questions in each core module is the monthly spending graphs. To save space, I present below only the initial monthly spending graphs from each module. In Table A.13, I summarize the changes between follow-up questions and initial questions in terms of the spending amounts necessary to elicit the preference parameters in Propositions 1, 4 to 9, 11, and 12.

Existence of Internal Habit Formation

See Figure 4 in the paper.

Habit Depreciation Speed

See Figure 2 in the paper.

Time Preference

See Figure A.2.

Slope of Indifference Curve

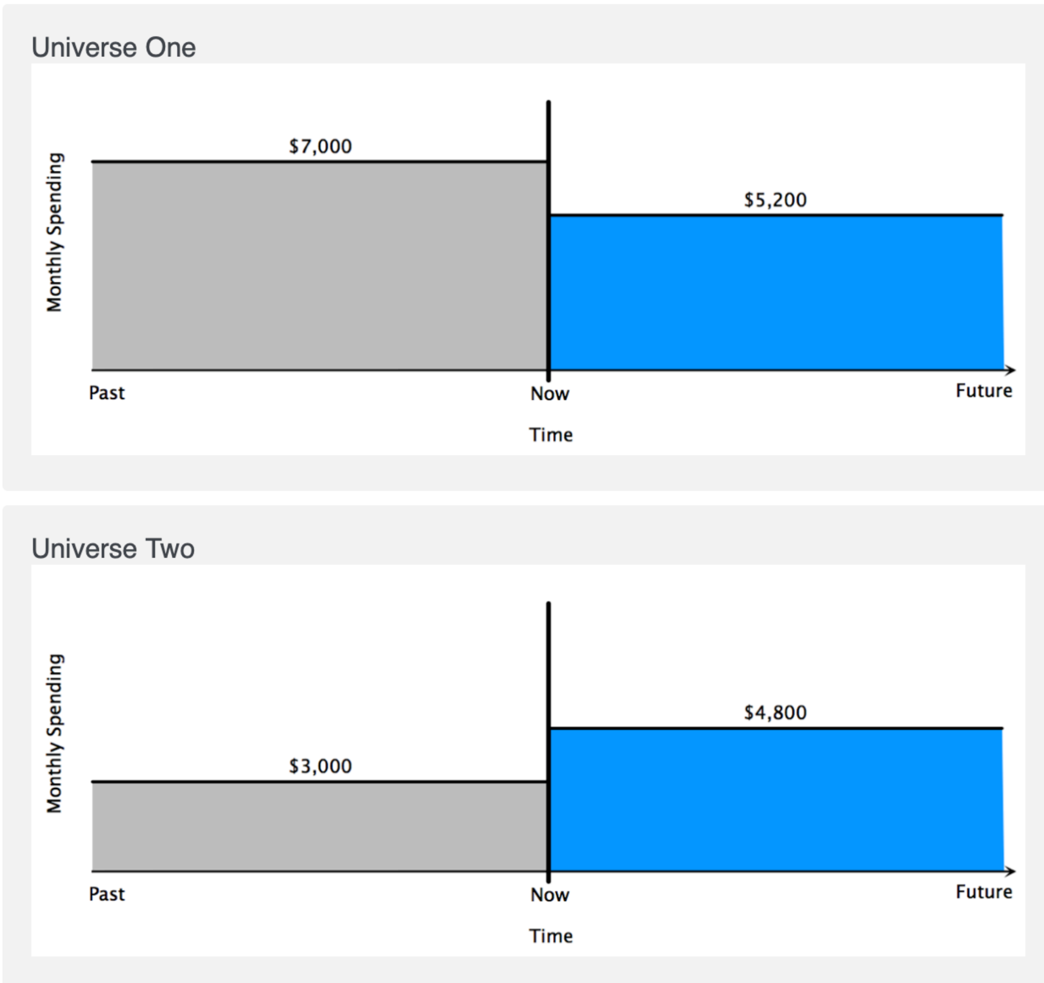


Figure A.15: Monthly Spending Graphs—Slope of Indifference Curve

$\frac{Hu_{HH}}{u_H}$

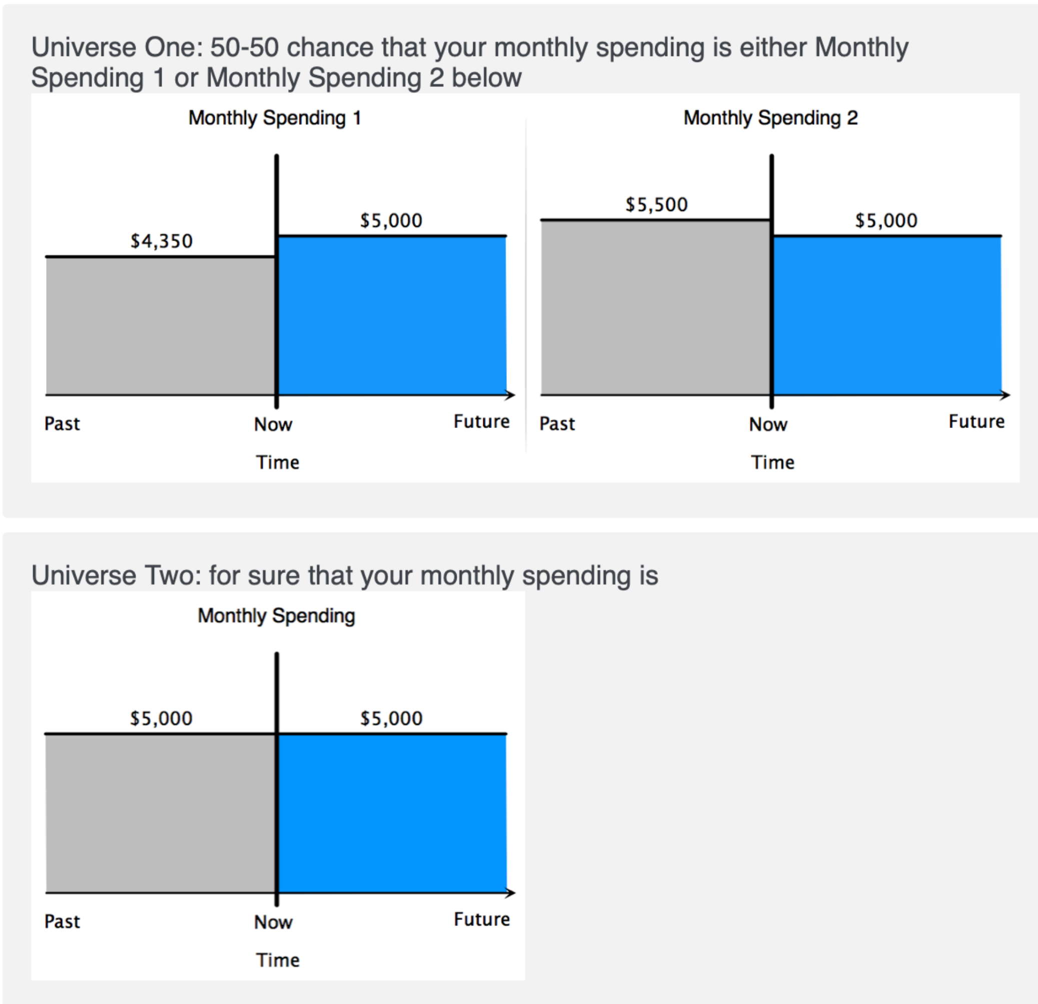


FIGURE A.16: MONTHLY SPENDING GRAPHS—$\frac{Hu_{HH}}{u_H}$

$\frac{u_{CH}}{u_{HH}}$

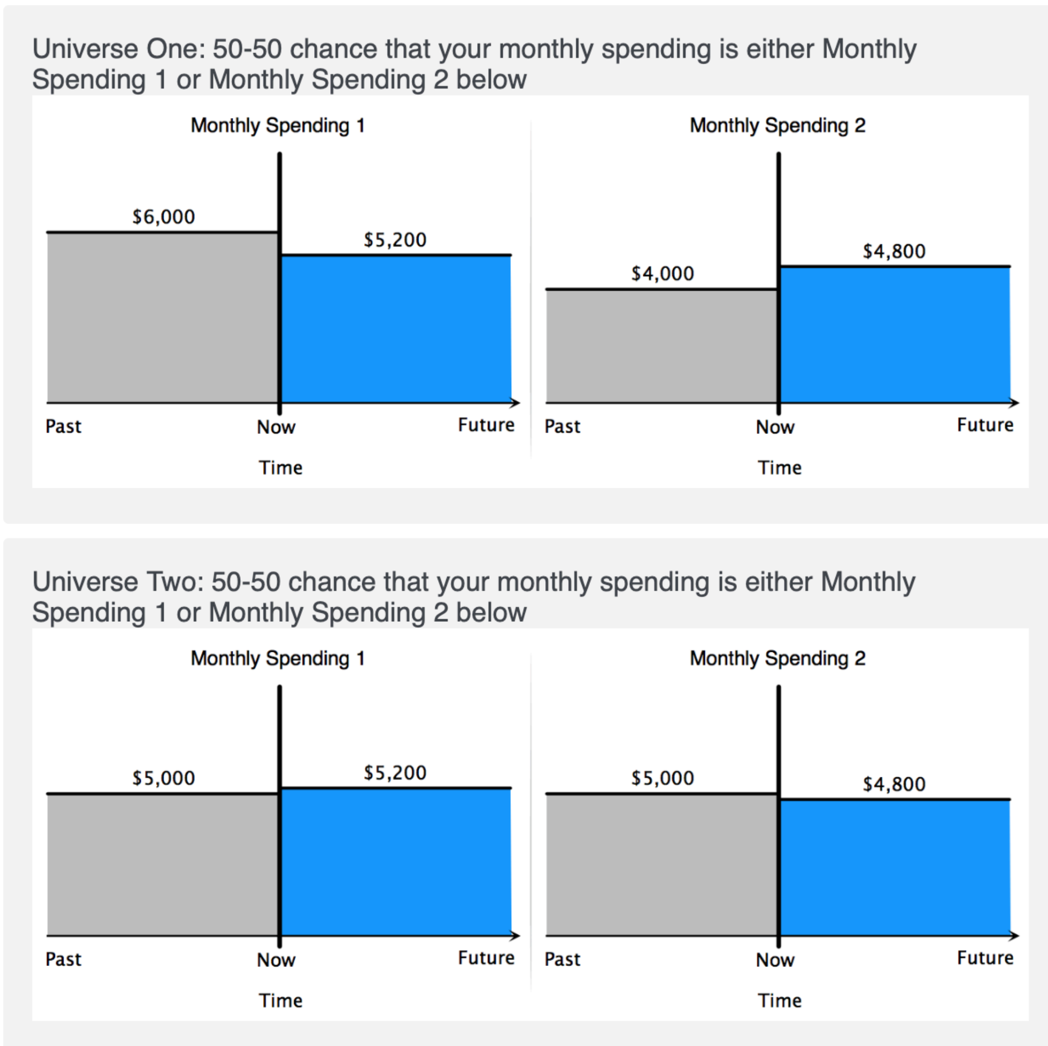


Figure A.17: Monthly Spending Graphs—$\frac{u_{CH}}{u_{HH}}$

$\frac{u_{CC}}{u_{HH}}$

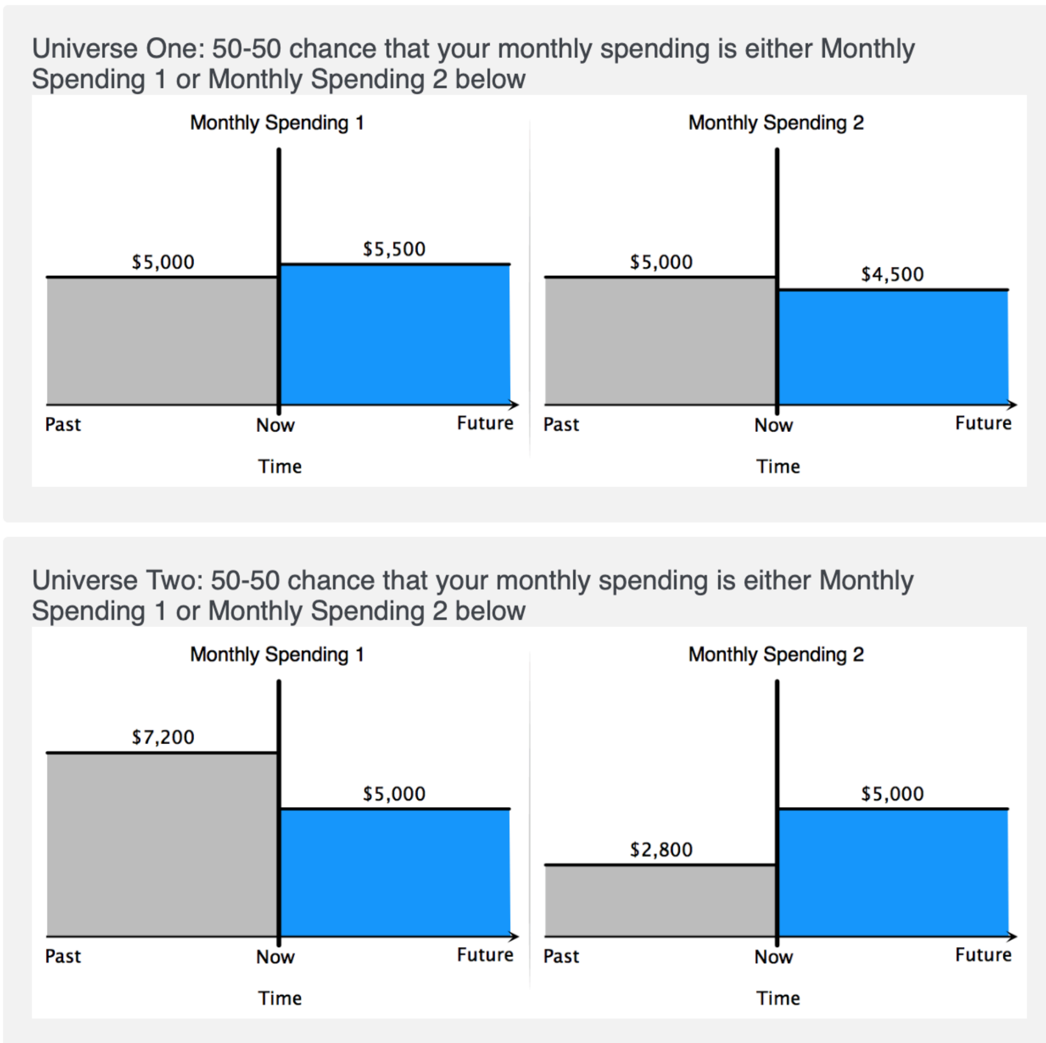


FIGURE A.18: MONTHLY SPENDING GRAPHS—$\frac{u_{CC}}{u_{HH}}$

External Habit Formation and Composition of Habit

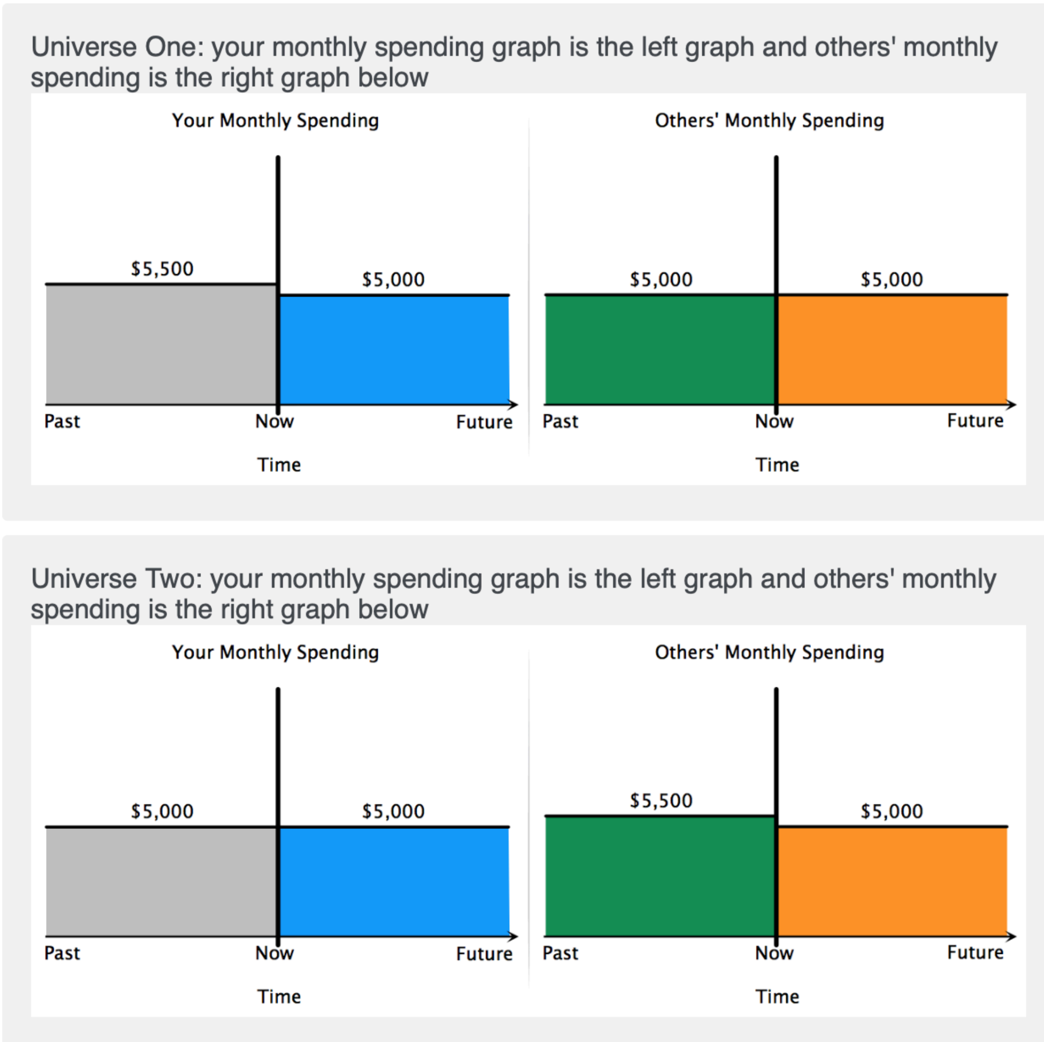


FIGURE A.19: MONTHLY SPENDING GRAPHS—EXTERNAL HABIT FORMATION

Relative Strength of Habit Formation and Peer Effect

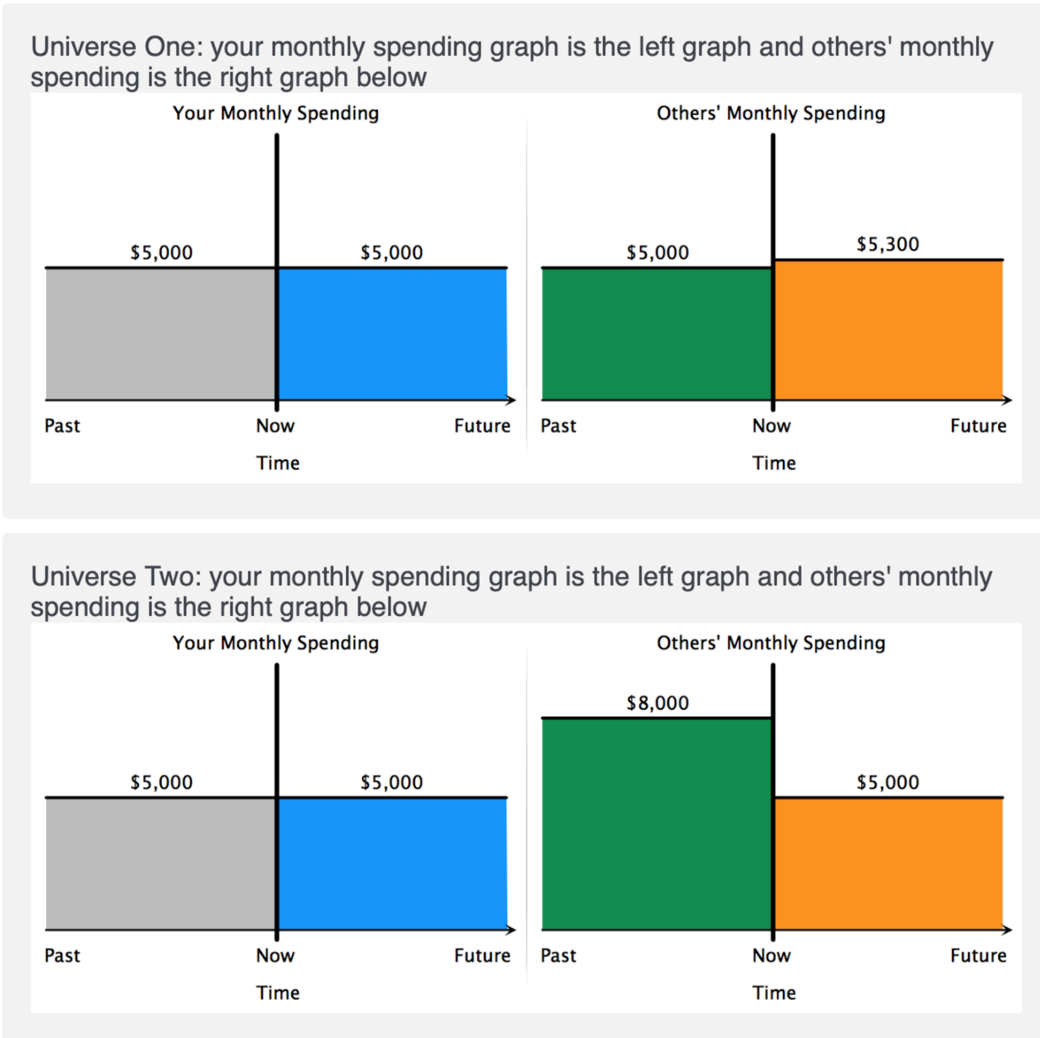


FIGURE A.20: MONTHLY SPENDING GRAPHS—$\frac{u_{C_{\text{OTHERS}}}}{u_H}$

Spending Amounts in Initial and Follow-Up Questions

TABLE A.13: 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
$-u_H/u_C$	Δe	2000	2000	2000	2000	2000
	Δf	20	80	200	400	1000
Hu_{HH}/u_H	Δe	500	500	500	500	500
	Δf	540	600	650	700	800
u_{CH}/u_{HH}	Δe	2000	1600	1000	600	100
	Δf	200	200	200	200	200
u_{CC}/u_{HH}	Δe	500	1500	2200	3000	3500
	Δf	500	500	500	500	500
$u_{C_{\text{others}}}/u_H$	$\Delta C^{U1}_{\text{others}}$	3000	600	300	150	100
	$\Delta C^{U2}_{\text{others}}$	3000	3000	3000	3000	3000
Time discount rate	Δe	2000	2000	2000	2000	2000
	Δf	3300	2500	2200	2100	2040

Notes: U1 and U2 denote Universe One and Universe Two of the monthly spending graphs, respectively. Choosing U1 in the initial question and then U2 in the 1st follow-up question or choosing U2 in the initial question and then U1 in the 1st follow-up question end the module at the end of the 1st follow-up question (cf. Figure 5 in the paper). All amounts are in U.S. dollar.

G.4 End-of-Survey Quiz on the Basic Hypothetical Situation

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

Under the hypothetical situation of this survey, if you can buy 3 bananas with one dollar in the last year, how many bananas can you buy with one dollar in the next year?

- 5
- 3
- 1
- No idea

Under the hypothetical situation of this survey, select any of the following that you own (that is, not rent):

- Residence
- Car
- Furniture
- I do not own any of the above
- No idea

Under the hypothetical situation of this survey, do things you want change over time?

- Yes
- Maybe
- No

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

- Yes
- Maybe
- No

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

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

H Response Frequency

Table A.14 summarizes the response frequency of preference parameters identifiable to scale. Response frequency of parameters identifiable only to sign are reported in Tables 3 and A.1.

TABLE A.14: RESPONSE FREQUENCY (PERCENTAGE) FOR PREFERENCE PARAMETERS IDENTIFIABLE TO SCALE

Question	Wave	Choice					
		1	2	3	4	5	6
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
Time discount rate	1	33	4	7	7	5	43
	2	39	3	5	4	4	45

I Proofs

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

I.1 Lemmas

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

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

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

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

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

$\forall n \in \mathbb{N}^+$.

Now,

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

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

Lemma 3. *For $\Delta e, \Delta f, M, \theta, t\in\mathbb{R}^{+}$, if $M-\Delta e-\Delta f\geq 0$, $u\left(C,H\right)$ 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,*

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

and $K\left(n\right)=k$ 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 $Me^{-\theta t} > 0$,

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

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,

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

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

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

This process of successive applications of Lemmas 1 and 2 can continue until

$\partial^{i_1} u/\partial H^{i_1} > 0$ where $i_1 \geq 2$.

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.[10]

Without loss of generality, suppose, for some $i_k \in I$,

$$\sum_{n=1}^{i_k-1} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) < 0. \tag{A.3}$$

Note that this is satisfied for $u_{HH} > 0$ (or $i_k = 2$).

Lemma 2 implies

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

which allows the application of Lemma 1 for getting

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

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

Now,

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

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 $\partial^{i_1} u/\partial H^{i_1} < \Lambda_1$.

From there, apply Lemma 2 to get

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

and Lemma 1 to get

$$\begin{aligned}
&\sum_{n=1}^{i_k} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) \\
&\quad + \frac{\frac{\sum_{s=0}^{i_k} (\Delta e)^s (\Delta f)^{i_k - s}}{M^{i_k}}}{\frac{\sum_{s=0}^{i_k - 1} (\Delta e)^s (\Delta f)^{i_k - 1 - s}}{M^{i_k - 1}}} \frac{\sum_{s=0}^{i_k - 1} (\Delta e)^s (\Delta f)^{i_k - 1 - s}}{M^{i_k - 1}} \sum_{n=i_k+1}^{\infty} \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(M e^{-\theta t} \right)^n \\
&= \sum_{n=1}^{i_k+1} e^{-n\theta t} M \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(\sum_{s=0}^{n-1} (\Delta e)^s (\Delta f)^{n-1-s} \right) \\
&\quad + \frac{\sum_{s=0}^{i_k} (\Delta e)^s (\Delta f)^{i_k - s}}{M^{i_k}} \sum_{n=i_k+2}^{\infty} \frac{1}{n!} \frac{\partial^n u}{\partial H^n} \left(M e^{-\theta t} \right)^n \\
&< 0.
\end{aligned} \tag{A.4}$$

In summary, the above process can be applied iteratively for all $n \in \mathbb{N}^+$: apply the similar derivation as from inequalities (A.1) to (A.2) when $\partial^n u/\partial H^n \leq 0$, and apply the similar derivation as from inequalities (A.3) to (A.4) when $\partial^n u/\partial H^n > 0$.

Finally, dividing both sides of the resulting inequality by M to get

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

$\square$

I.2 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.[11]

A respondent preferring Universe One for a better future experience (U) implies

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

which, by Lemma I.1, implies

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

or equivalently

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

□

[11] See Table A.13 for all the values of ΔC_{U1} and ΔC_{U2} used in survey.

I.3 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

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

□

I.4 Proof of Proposition 3

Proof.

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

and

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

□

I.5 Proof of Proposition 4

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

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

which, by $u_C > 0$ and $\rho > 0$,[12] implies

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

□

See Table A.13 for all the values of Δe and Δf used in survey.

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

I.6 Proof of Proposition 5

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

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

which, by $u_H < 0$[13] and $H > 0$, implies

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

□

See Table A.13 for all the values of Δe and Δf used in survey.

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

I.7 Proof of Proposition 6

Proof. Under third-order elicitation,[14] a respondent preferring Universe One for a better future experience (U) 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)}] \\
={} & \frac{1}{2}\frac{1}{\rho}\left\{u_C \Delta f + \frac{1}{\rho+\theta} u_H\left(\rho\Delta e + \theta\Delta f\right) + \frac{1}{2}\left[u_{CC}\left(\Delta f\right)^2 + 2\frac{1}{\rho+\theta}u_{CH}\Delta f\left(\rho\Delta e + \theta\Delta f\right)\right.\right. \\
& \left.\left. + \frac{1}{(\rho+\theta)(\rho+2\theta)}u_{HH}\left(\rho(\rho+\theta)(\Delta e)^2 + 2\rho\theta\Delta e\Delta f + 2\theta^2(\Delta f)^2\right)\right]\right\} \\
& + \frac{1}{2}\frac{1}{\rho}\left\{-u_C \Delta f - \frac{1}{\rho+\theta} u_H\left(\rho\Delta e + \theta\Delta f\right) + \frac{1}{2}\left[u_{CC}\left(\Delta f\right)^2 + 2\frac{1}{\rho+\theta}u_{CH}\Delta f\left(\rho\Delta e + \theta\Delta f\right)\right.\right. \\
& \left.\left. + \frac{1}{(\rho+\theta)(\rho+2\theta)}u_{HH}\left(\rho(\rho+\theta)(\Delta e)^2 + 2\rho\theta\Delta e\Delta f + 2\theta^2(\Delta f)^2\right)\right]\right\} \\
& - \frac{1}{2}\frac{1}{\rho}\left\{\left(u_C + \frac{\theta}{\rho+\theta}u_H\right)\Delta f + \frac{1}{2}\left[u_{CC} + 2u_{CH}\frac{\theta}{\rho+\theta} + u_{HH}\frac{2\theta^2}{(\rho+\theta)(\rho+2\theta)}\right](\Delta f)^2\right\} \\
& - \frac{1}{2}\frac{1}{\rho}\left\{\left(u_C + \frac{\theta}{\rho+\theta}u_H\right)(-\Delta f) + \frac{1}{2}\left[u_{CC} + 2u_{CH}\frac{\theta}{\rho+\theta} + u_{HH}\frac{2\theta^2}{(\rho+\theta)(\rho+2\theta)}\right](\Delta f)^2\right\} \\
={} & \frac{1}{2}\frac{1}{\rho}\left[u_{CC}\left(\Delta f\right)^2 + 2\frac{1}{\rho+\theta}u_{CH}\Delta f\left(\rho\Delta e + \theta\Delta f\right)\right. \\
& \left. + \frac{1}{(\rho+\theta)(\rho+2\theta)}u_{HH}\left(\rho(\rho+\theta)(\Delta e)^2 + 2\rho\theta\Delta e\Delta f + 2\theta^2(\Delta f)^2\right)\right] \\
& - \frac{1}{2}\frac{1}{\rho}\left[u_{CC} + 2u_{CH}\frac{\theta}{\rho+\theta} + u_{HH}\frac{2\theta^2}{(\rho+\theta)(\rho+2\theta)}\right](\Delta f)^2 \\
={} & \frac{1}{2}\frac{1}{\rho}\left[2u_{CH}\frac{\rho}{\rho+\theta}\Delta e\Delta f + \frac{1}{(\rho+\theta)(\rho+2\theta)}u_{HH}\left(\rho(\rho+\theta)(\Delta e)^2 + 2\rho\theta\Delta e\Delta f\right)\right] \\
>{} & 0,
\end{aligned}
$$

which, by $u_{HH} < 0$ and $\rho > 0$,[15] implies

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

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

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

□

See Table A.13 for all the values of Δe and Δf used in survey.

I.8 Proof of Proposition 7

Proof. Under third-order elicitation,[16] a respondent preferring Universe One for a better future experience (U) implies

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

which, by $u_{HH} < 0$ and $\rho > 0$,[17] implies

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

□

See Table A.13 for all the values of Δe and Δf used in survey.

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

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

I.9 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.[18]

A respondent preferring Universe One for a better future experience (U) implies

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

which, by Lemma I.1, implies

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

or equivalently

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

□

[18] See Table A.13 for all the values of ΔC and ΔC_{others} used in survey.

I.10 Proof of Proposition 9

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

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

which, by $u_H < 0$ and $\rho > 0$,[19] implies

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

□

See Table A.13 for all the values of $\Delta C^{U1}_{\text{others}}$ and $\Delta C^{U2}_{\text{others}}$ used in survey.

I.11 Proof of Proposition 10

See Section B of this appendix.

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

I.12 Proof of Proposition 11

Proof. Under exact elicitation, a respondent preferring Universe One over Universe Two for a better future experience (U) implies

$$\begin{aligned}
& U\text{ (Universe One)} - U\text{ (Universe Two)} \\
=& \left[U\text{ (Universe One)} - U\text{ (Baseline)}\right] - \left[U\text{ (Universe Two)} - U\text{ (Baseline)}\right] \\
=& \int_0^{\infty} e^{-\rho t} \left\{\left[u\left(C, H\left(\mathbf{C} + \Delta\mathbf{C}, \mathbf{C}_{\text{others}}\right)\right) - u\left(C, H\left(\mathbf{C}, \mathbf{C}_{\text{others}}\right)\right)\right]\right. \\
& \left.- \left[u\left(C, H\left(\mathbf{C}, \mathbf{C}_{\text{others}} + \Delta\mathbf{C}_{\text{others}}\right)\right) - u\left(C, H\left(\mathbf{C}, \mathbf{C}_{\text{others}}\right)\right)\right]\right\} dt \\
>& 0
\end{aligned}$$

where $\Delta\mathbf{C} > 0$ and $\Delta\mathbf{C}_{\text{others}} > 0$.

This implies

$$\begin{aligned}
Q_{EH} &\equiv u\left(C, H\left(\mathbf{C}, \mathbf{C}_{\text{others}} + \Delta\mathbf{C}_{\text{others}}\right)\right) - u\left(C, H\left(\mathbf{C}, \mathbf{C}_{\text{others}}\right)\right) \\
&< u\left(C, H\left(\mathbf{C} + \Delta\mathbf{C}, \mathbf{C}_{\text{others}}\right)\right) - u\left(C, H\left(\mathbf{C}, \mathbf{C}_{\text{others}}\right)\right) \\
&< 0
\end{aligned}$$

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

I.13 Proof of Proposition 12

Proof. Let

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

Under exact elicitation, a respondent preferring Universe One over Universe Two for a better future experience (U) implies

$$\begin{aligned}
&U \text{ (Universe One)} - U \text{ (Universe Two)} \\
=&\left[U \text{ (Universe One)} - U \text{ (Baseline)}\right] - \left[U \text{ (Universe Two)} - U \text{ (Baseline)}\right] \\
=&Q\left(\Delta e\right) - e^{-\rho} Q\left(\Delta f\right) \\
=&\Delta e \frac{Q\left(\Delta e\right)}{\Delta e} - e^{-\rho} \Delta f \frac{Q\left(\Delta f\right)}{\Delta f} \\
>&\left(\Delta e - e^{-\rho} \Delta f\right) \frac{Q\left(\Delta e\right)}{\Delta e} \\
>&0
\end{aligned}$$

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

This implies

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

or equivalently

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

□

See Table A.13 for all the values of Δe and Δf used in survey.

References

This section presents the reference list of studies in Table 1 of the paper.

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

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

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

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

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

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

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

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

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

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

Guariglia, Alessandra, and Mariacristina Rossi. 2002. "Consumption, Habit Formation, and Precautionary Saving: Evidence from the British Household Panel Survey." *Oxford Economic Papers*, 54(1): 1–19.

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

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

Khanal, Aditya R., Ashok 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.

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

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

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

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

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

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

Rhee, Wooheon. 2004. "Habit Formation and Precautionary Saving: Evidence from the Korean

Household Panel Studies." *Journal of Economic Development*, 29(2): 1–19.

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

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

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