A Structural Neural Autopilot Analysis of Social Media Use Around the Pandemic Lockdown*

Yi Xin[1], Lawrence J. Jin[2], Jessica Fong[3], Matthew Shum[1], and Colin Camerer[1]

[1]California Institute of Technology

[2]Cornell University and NBER

[3]University of Michigan†

March 10, 2024

Abstract

This paper describes and estimates a "neural autopilot" model of habit formation. The estimation uses individual-level data on posting behavior from a Chinese social media platform before, during, and after the 2020 pandemic lockdown. The model produces interpretable parameter estimates about autopilot habit formation. It shows that once habit is neuroscientifically formalized, changes in preferences are no longer required to explain observed behavior change. Moreover, the neural autopilot model fits the data better than a traditional model of habit that uses changing preferences to explain choice persistence. We also find that forced experimentation alone does not lead to persistent habitual postings after the lockdown ends. Counterfactual forecasts, which are derived from simulating behavior using the structural model, show that reducing the volatility in posting rewards, in conjunction with forced experimentation, would significantly increase habitual postings. This finding suggests that higher moments of the reward process may play an important role in creating habits.

*We thank Monica Capra, Xiaomin Li, three thoughtful referees, and seminar participants at the 2021 AEA Annual Meeting, the 2021 ISMS Marketing Science Conference, the 2022 SITE Conference, the 2023 Choice Symposium, and the Virtual Quantitative Marketing Seminar for helpful comments. Zhuofang Li provided excellent research assistance. This work is supported by the Sloan Foundation (G-2018-1125) and the National Science Foundation (RAPID SES–1851745). Any opinions, findings, and recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the Sloan Foundation and the National Science Foundation.

†Xin, Shum, and Camerer are affiliated with the Division of the Humanities and Social Sciences, California Institute of Technology, 1200 E California Blvd, MC 228-77, Pasadena, CA 91125. Jin is affiliated with the SC Johnson College of Business, Cornell University, 137 Reservoir Ave, Ithaca, NY 14850 and the National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge, MA 02138. Fong is affiliated with the Ross School of Business, University of Michigan, 701 Tappan Ave, Ann Arbor, MI 48109. The authors' e-mails are: yixin@caltech.edu (Xin), lawrence.jin@cornell.edu (Jin), jyfong@umich.edu (Fong), mshum@caltech.edu (Shum), and camerer@hss.caltech.edu (Camerer), respectively.

1 Introduction

People form habits. Habit formation interests psychologists, doctors, and neuroscientists who seek to understand how habit formation works. Finding ways to create more good habits and fewer bad habits also preoccupies areas of applied psychology and public health. All social sciences are interested in habits of some kind—for example, organizational routines, rituals and norms in sociology and anthropology.

We present one of the first empirical analyses of a model that closely matches how some neuroscientists and cognitive psychologists think about habits. The motivating principle is that people form habits to economize on mental effort: they repeat previous choices when they expect those choices to deliver similar rewards as they have experienced in the past.[1]

Our approach conceptualizes habit as a particular decision process which is intimately related to learning about what is rewarding, rather than about choosing given stable preferences. People repeat a previous choice either because the choice is utility-maximizing, or because they are in a habit mode, whereby the habitual choice can be suboptimal as it neglects consideration of other options. In the latter case, forcing people to experiment with new choices may lead to higher utility.[2] Indeed, Verplanken and Orbell (2022) note that empirically, substantial changes in the choice environment often lead to changes in habits.

What factors create and sustain habits? Can forced experimentation break old habits and generate new habits and persistent behavior change? These remain fundamental questions in social sciences. Our paper addresses these questions by structurally estimating a novel neuroeconomic model of habit formation. The model is a "neural autopilot" model, in which people toggle between previous, reliably rewarded choices, and goal-directed maximization once the habitual choice led to large reward prediction errors (Landry et al., 2021).[3] It is a particular kind of dual-process model, out of many that have been studied in psychology research (Stanovich, 1999; Evans, 2008; Kahneman, 2011; Evans and Stanovich, 2013; Cerigioni, 2021). More precisely, the neural autopilot

[1] Choice repetition can also depend on contextual states, especially in the case of addictive substances for which cues can trigger craving and drug use (Laibson, 2001; Bernheim and Rangel, 2009; Wellsjo, 2021).

[2] Larcom et al. (2017) study forced experimentation in the context of local transit. The authors find that a strike on the London Underground leads to behavior change.

[3] The word "neural" is used because the model's central dual-process mathematical form is closely related to the neuroscientific mechanisms developed in Daw et al. (2005) and subsequent studies. We further discuss this relation in Section 2.1.

model is a "default interventionist" model, in the sense that people are assumed to act on low-cost habits unless large reward prediction errors motivate them to engage in more effortful goal-directed choices.

Most analyses of apparent habits report a "reduced-form" persistence of choice over time. Our structural approach hypothesizes a particular way in which choices are governed by past experience and a set of parameter values. This approach is aimed at explaining when choice persistence arises and when it does not; moreover, the structural model allows us to quantitatively measure the long-term effects of forced experimentation on behavior change. To the best of our knowledge, our paper is among the first studies that use large-scale field data to estimate and test a neuroscientific model of habit formation.

The neural autopilot model works as follows. In every time period, an economic agent chooses from multiple options and is in one of two possible modes of decision making: a "goal-directed" mode and a habit mode. In the goal-directed mode, the agent actively predicts the reward value for each option in her choice set. Then, she chooses an option probabilistically, according to a "softmax" function, where the probability of choosing a given option is an increasing function of the option's predicted reward value. In the habit mode, however, the agent simply repeats her choice from the previous period, without actively predicting each option's reward value.

In this model, the agent's choice always leads to a realized reward—for example, the "likes" received from a social media post. The agent then computes a "reward prediction error," which is the difference between the realized reward and her predicted reward, and uses it to update her reward prediction for the corresponding choice. If an option's realized reward has been close to its reward prediction for many periods—that is, if the reward prediction errors have been small—the reward prediction becomes sufficiently reliable. In this case, the agent is likely in the habit mode.

We apply this autopilot model to a large sample of individual-level activity data on posting behavior from Weibo, one of the biggest social media platforms in China. We focus on the time period around the 2020 lockdown in response to the COVID-19 pandemic.[4] This setting is well-suited to applying our autopilot model, for at least two reasons. First, the lockdown was strictly

[4]Wuhan and some surrounding cities went into lockdown on January 23, 2020. Many other counties of the Hubei province entered lockdown on the following day. More information on the lockdown can be found at `https://apnews.com/article/pandemics-wuhan-china-coronavirus-pandemic-e6147ec0ff88affb99c811149424239d` (accessed March 2024).

enforced, and it caused an unexpected interruption in social activities, especially outdoor activities. The restriction on people's choice set led them to explore new activities such as posting on Weibo—this is evidenced by the large increase in new users observed at the beginning of the lockdown. And such a big change likely disrupts habits, as demonstrated in many other domains by Verplanken and Orbell (2022). Second, our data allow us to directly measure rewards from posting using the number of likes received from each post. By contrast, alternative settings typically used in studying habit formation, such as gym visits, often do not provide a good empirical proxy for rewards, hence preventing a direct test of the neural autopilot model.

We estimate the neural autopilot model using daily posting and reward data for a group of randomly sampled users who had used Weibo prior to the pandemic lockdown. In the data, we observe a decrease in posting probability after the lockdown ends. Our model explains this post-lockdown drop in the posting probability through users' reactions to changes in posting rewards, rather than through changes in their preferences toward non-social media activities during and after the lockdown. Our model suggests that people are likely to be in the goal-directed mode during the lockdown when many of their outside options become unavailable; once the lockdown is lifted, the rewards received from posting decrease, and people form habits for non-social media activities. Compared with a set of alternative models of choice behavior, the neural autopilot model offers higher explanatory power for changes in observed behavior.

Using the estimated model parameters, we investigate the key factors in creating and sustaining habits. We conduct two counterfactual "thought experiments": What does the model say would happen if we increase the average level of posting rewards during the lockdown? And what does the model say would happen if we reduce the volatility of posting rewards—the volatility of likes received from posting—during the lockdown? Overall, we find that forced experimentation does not lead to new habit that persists after the lockdown is lifted.

However, reducing the volatility of posting rewards during the lockdown leads to a temporary but significant increase in habitual postings, particularly for users who can quickly learn about their perceived reliability of predicted rewards; this learning rate is a behavioral parameter in our model and it is likely to be heterogeneous across users. Our finding suggests that higher moments of the reward process may play an important role in creating habits.

Our paper contributes to the literature on habit formation in many ways. Economic and mar-

keting models of habit often assume that the utility of current goods depends on past consumption, without any direct biological motivation for this reduced-form assumption (Pollak, 1970; Becker and Murphy, 1988; Constantinides, 1990; Campbell and Cochrane, 1999; Allcott et al., 2022). By contrast, our model draws directly from neuroscientific evidence (Daw et al., 2005; Dolan and Dayan, 2013; Lee et al., 2014) and makes specific predictions about how and when people form or break a habit. Our paper also relates to the literature on choice persistence and structural state dependence in economics and marketing (Keane, 1997; Seetharaman et al., 1999; Dubé et al., 2010). For example, Dubé et al. (2010) find that structural state dependence can be explained by preference changes due to past consumption, rather than search costs or learning. We contribute to this literature by considering neuroscientific habit formation as an alternative explanation for structural state dependence. We show that our neural autopilot model better explains observed behavior change compared to the traditional approach.

Our paper also contributes to the literature on reward learning and social media. Social media has been called a "Skinner Box for the modern human" (Lindström et al., 2021, page 2). However, it is important to note that distinct challenges exist when inferring behavioral processes—especially, causality—from social media data (Burton et al., 2021).

Several studies have estimated reinforcement learning models with field data, assuming that positive social media feedback serves as a reinforcer. Das and Lavoie (2014) estimate a reinforcement learning model using data from Reddit and demonstrate that this model outperforms alternative models in predicting posting behavior. Lindström et al. (2021) examine both lab and field data, and they document that choice behavior on social media is consistent with reinforcement learning theory. Brady et al. (2021) find that positive social feedback is correlated with an increase in posting on Twitter. Moreover, the effects of reinforcement on subsequent behavior are stronger when people have more followers (Lindström et al., 2021) and longer posting histories (Brady et al., 2021; Anderson and Wood, 2023). Anderson and Wood (2023) show that choice behavior of Facebook users who post more frequently and have stronger self-reported habits is less sensitive to changes in social rewards; this is consistent with the idea that habits create insensitivity to changes in valuation. Our paper contributes to this literature by estimating a structural reinforcement learning model which incorporates a specific transition in and out of habit, one that can be estimated using observable data. None of the previous analyses extend reinforcement learning models to

include explicit habit formation.

More broadly, our paper adds to a growing literature that uses models from psychology and neuroscience to understand human behavior in economic and financial settings (Landry et al., 2021; Khaw et al., 2021; Webb et al., 2022; Frydman and Jin, 2022, 2023; Barberis and Jin, 2023; Wachter and Kahana, 2023).[5] Most related here is Webb et al. (2022), who estimate the neural autopilot model using data from canned tuna purchases; in their setting, a change in the size of the tuna cans shifts consumers from the habit mode to the goal-directed mode. An important difference between the two papers is that their paper focuses on the extent to which habits can explain choice persistence, while our paper focuses on the role of reward schemes in creating habits.

2 The Neural Autopilot Model

Our model builds on the general framework of Landry et al. (2021). We first describe the basic structure in mathematical terms. We then give the intuition behind our modelling choices and provide further evidence and motivation.

We study an economic agent who chooses one action from a set of actions $\mathbf{a} = \{a_1, a_2, ...a_J\}$ in every discrete period. The agent's choice at time t is denoted by c_t. Each action $a \in \mathbf{a}$ is associated with a predicted reward, which is the agent's prediction of the reward she will receive from choosing a. The predicted reward for a at time t is denoted by $r_t(a)$. We assume that the agent learns and updates the reward prediction $r_t(a)$ as follows

$$r_t(a) = \begin{cases} r_{t-1}(a) + \lambda_r \cdot [u_{t-1}(a) - r_{t-1}(a)] & \text{if } a = c_{t-1} \\ r_{t-1}(a) & \text{if } a \neq c_{t-1} \end{cases}. \tag{1}$$

Equation (1) says that at time t, the agent updates only the predicted reward for the previous action c_{t-1} she took at time $t-1$. Specifically, the agent updates $r_t(c_{t-1})$ by the amount of $\lambda_r \cdot [u_{t-1}(c_{t-1}) - r_{t-1}(c_{t-1})]$, where λ_r is a learning rate and the term in square brackets is the reward prediction error (RPE). The RPE is the difference between the realized reward of taking

[5] Allcott et al. (2022) also study habit formation in the context of social media, but their paper does not adopt a neuroscientific approach.

the action c_{t-1}, denoted by $u_{t-1}(c_{t-1})$, and the previously predicted value $r_{t-1}(c_{t-1})$. The updating rule in Equation (1) is based directly on neural evidence accumulated over the past 25 years. For example, studies of human decision making by O'Doherty et al. (2003) and Rangel et al. (2008) find that neural activity in the ventral striatum correlates strongly with the RPE computed in Equation (1). Overall, the neural circuitry of RPE is one of the most well-established regularities in decision neuroscience (see Rangel et al., 2008 and Dolan and Dayan, 2013 for reviews).

The agent also tracks the *reliability* of each action's predicted reward. The reward reliability for action a at time t is denoted by $d_t(a)$ and is often called the "unsigned prediction error." We assume that the agent updates $d_t(a)$ as follows

$$d_t(a) = \begin{cases} (1-\lambda_d)d_{t-1}(a) + \lambda_d \cdot |u_{t-1}(a) - r_{t-1}(a)| & \text{if } a = c_{t-1} \\ (1-\lambda_d)d_{t-1}(a) + \lambda_d & \text{if } a \neq c_{t-1} \end{cases}. \tag{2}$$

Equation (2) says that at time t, the agent updates $d_t(c_{t-1})$, the reward reliability for the previous action she took at time $t-1$, based on $\lambda_d \cdot |u_{t-1}(c_{t-1}) - r_{t-1}(c_{t-1})|$, where $0 < \lambda_d < 1$ is a learning rate for reward reliability d. When the realized reward $u_{t-1}(c_{t-1})$ is very close to the previously predicted reward $r_{t-1}(c_{t-1})$, the agent believes that her reward prediction becomes more reliable; the absolute value of the RPE is close to zero, leading to $d_t(c_{t-1}) < d_{t-1}(c_{t-1})$. Lee et al. (2014) present neural evidence that numerical reliability signals in Equation (2) are encoded in identifiable brain regions.

How does the agent actually choose an action c_t at any time t? If the agent is in the habit mode, she simply repeats her previous action; $c_t = c_{t-1}$. In this case, the agent's choice is not sensitive to rewards; she chooses the same action regardless of its expected reward. This is consistent with the finding of Anderson and Wood (2023) that the posting rates of habitual social media users are less sensitive to changes in social rewards, compared to the posting rates of non-habitual social media users. If the agent is in the goal-directed mode, the agent is then assumed to choose an action probabilistically, where the probability of choosing a given action a is increasing in the action's predicted reward $r_t(a)$:

$$\Pr_t(c_t = a) = \frac{\exp\left[\alpha \cdot r_t(a)\right]}{\sum_j \exp\left[\alpha \cdot r_t(a_j)\right]}. \tag{3}$$

This type of probabilistic choice, known as a "softmax" specification in the reinforcement learning literature, encourages the agent to "explore," in other words, to try an action other than the one that currently has the highest predicted reward (Cesa-Bianchi et al., 2017). The parameter α controls the degree of this directed exploration.[6]

Finally, our model specifies the following rule for the switches between the goal-directed mode and the habit mode:

$$\Pr_t(\text{goal-directed}) = \frac{1}{1 + \exp(-\kappa \cdot (d_t(c_{t-1}) - \phi))}. \tag{4}$$

Equation (4) says that when the agent finds her predicted reward $r_t(c_{t-1})$ sufficiently reliable—that is, when $d_t(c_{t-1})$ is low—the agent will likely enter the habit mode; here, the parameter ϕ can be interpreted as a "threshold" of $d_t(c_{t-1})$ that affects the switches between the goal-directed mode and the habit mode. In Section 3.1, we provide empirical evidence that more reliable rewards are associated with more habitual behavior.

2.1 Relation of Neural Autopilot to Dual-Process and Reinforcement Evidence

We have now described the mathematical details of the neural autopilot model. Next, we turn to three important questions regarding the relation between our model and previous theory and evidence. The first question is concerned about why our model is viewed as a "neural" autopilot model. The second question is about how our model incorporates the cost-benefit tradeoff between fast, low-cost habitual choice and slower, higher-cost model-based choice. And the third and final question is about whether we have modelled reward predictability in a way that is consistent with the large amount of evidence from animal learning (and some human data) under fixed or variable reward "schedules."

We start with the first question: What makes our model "neural"? The updating rule for reward prediction, as expressed in Equation (1), is a form of model-free reinforcement learning and is certainly similar to a long line of associational-strength updating rules originating in Thorndike (1932)'s "Law of Effect" and Rescorla and Wagner (1972) (see Pearce and Bouton, 2001 for history). Neuroscientific data played no role in the early development of these learning rules. However,

[6] Alternatively, probabilistic choice could represent imprecision in encoding or in response; see Daw et al. (2006) and Zajkowski et al. (2017).

starting around the year 2000, neuroscientists began to focus on measuring neural activity and understanding neural implementation of decision modes. In an influential study, Jog et al. (1999) used the term "habit" to describe fast decisions in which, with learning, animal T-maze decisions became faster and better, and neural signals of reward (from recordings of neural spikes) propagated upstream from the time of receipt of the reward to the onset of a goal state which, with learning, predicted reward.

Following Jog et al. (1999), there was progress in understanding habits from many areas such as applied psychology, animal learning, computer science, and neuroscience. The scientific genealogy for our concept of neural autopilot can be traced to multiple-model "mixture of experts." They were suggested in computer science as methods that could perform well in complex domains where simpler single-model reinforcement learning did not (Jacobs et al., 1991). The idea of using reward prediction error to weight different models was suggested in Narendra et al. (1995) and elaborated in Doya et al. (2002) using the term "responsibility signal," a Bayesian posterior similar to that later used by Lee et al. (2014). The next important step was taken by Daw et al. (2005). These authors suggested an "arbitration" between model-free and model-based controllers, drawing on a large amount of imaging and causal evidence of model-free and model-directed systems and their dissociation.[7] They wrote (page 1704):

> "Here we suggest how the brain might estimate this accuracy for the purpose of arbitration by tracking the relative uncertainty of the predictions made by each controller."

Specifically, they hypothesized that the variance of a Bayesian posterior distribution over the learned probability of an action being optimal, could be used to determine whether a tree-directed model-based value or a "cached" model-free value is chosen to drive action. If the posterior variance is high for the model-free recommended choice, a person assigns control to the opposite, model-based system; conversely, if the posterior variance is high for the model-based recommended choice, the person assigns control to the opposite, model-free system. This notion of posterior variance is a precursor of the doubt stock $d_t(a)$ in our autopilot model; the absolute value of the reward prediction error in Equation (2) is akin to the posterior variance in Daw et al. (2005). Both papers study a dual-process model in which the system that currently has lower uncertainty in predicting

[7] Dolan and Dayan (2013) call this "Generation 3" of understanding goals and habits.

future reward is more likely to guide choice. It is important to note that, like our paper, Daw et al. (2005) also did not report new neural data; however, they did make the case that the arbitration process they hypothesized was consistent with various kinds of neural evidence. Following the Daw et al. (2005) emphasis on the reliability of estimating predicted reward, progress has been made in discovering neural circuitry that corresponds to the model-free and model-based choices, their reliability, and the neural "arbitration" that combines both choice tendencies (Lee et al., 2014).

In summary, our model is called "neural" because it came from considerations of a mathematical way of measuring differential predictability of dual-process models and these considerations originated from computer science and theoretical neuroscience.

We now turn to the second question: How does our model incorporate the cost-benefit tradeoff between fast, low-cost habitual choice and slower, higher-cost model-based choice? Attention is scarce. Therefore, it is obviously useful to offload choices to a low-cost mode of decision making when it is safe (reliably rewarding) to do so. However, our neural autopilot model does not explicitly account for mental costs, complexity, and stress. Further improvement of the model should include these factors.

The third and final question is: Have we modelled reward predictability in a way that is consistent with the large amount of evidence from animal and human learning about habitization under fixed or variable reward "schedules"? Many studies have examined how animals and humans learn in an environment where rewards are delivered stochastically or at different time intervals. A clear finding is that when rewards are stochastic and independent in each period (a "variable-ratio" schedule), habits do not form as strongly as when there is a random time interval between rewards (a "variable-interval" schedule). Is this finding consistent with the way we model reward predictability and habit formation in Equations (1) to (4)? The answer is unclear—figuring this out conclusively is not simple and lies beyond the scope and ambition of this paper. However, this question is certainly interesting and important, so we include a more detailed discussion in Appendix A.

3 Details of the Weibo Data and User Behavior

We now use our neural autopilot model to analyze posting behavior on Sina Weibo (Weibo), a Chinese microblogging platform. Launched in 2009, Weibo is one of the biggest social media platforms in China with over 500 million users in 2020. Commonly referred to as the "Twitter of China," Weibo allows its users to post original messages of up to 2,000 Chinese characters (or roughly 1,200 to 1,400 English words), repost messages from other users, and "like" posts.

We focus on users' posting behavior on Weibo around the mass lockdown China imposed between January and April 2020 in response to the COVID-19 pandemic. Note that the lockdown serves as a shock to both the outside option—non-social media activities—and the inside option—social media activities; for example, social media may become more valuable when face-to-face interactions are not possible. Regardless of whether the lockdown impacted the outside or inside option, it generated a shock that impacted behavior.

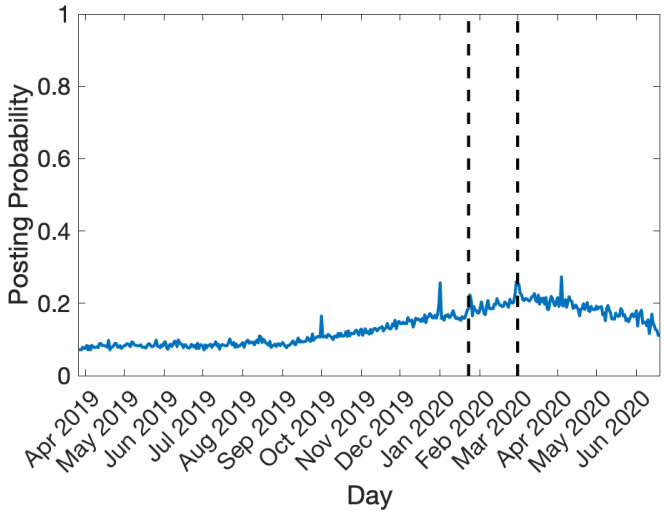
We collect data on users' posting behavior by randomly sampling a set of user accounts; more details on how we sample and clean the data can be found in Appendix B. For each post, we observe its content and timing as well as whether it is an original post or a repost from other users; we also observe the number of likes, the number of reposts, and the comments received by the post. Moreover, we collect a set of user and account characteristics, including the user's gender, age, and location, the number of followers the account has, the number of users the account follows, the date when the account was created, and whether or not the account is verified by Weibo.

In our empirical analysis, we focus on users who created an account before the lockdown; these are users who first posted on Weibo between 2015 and 2019.[8] We exclude users with private accounts and those whose maximum number of likes received in a single day is greater than the 95^{th} percentile cutoff. This leads to a final sample of 1,848 users for whom we observe their entire posting history. On average, we observe each user for about 3 years; the average probability that a user creates at least one post on a given day is 0.094 and the average number of postings per day is 0.321. Conditional on posting, the users receive an average of 0.993 likes per day. Figure 1 plots

[8]The start date, February 16, 2015, is selected to exclude users who joined Weibo at an early stage: early users may behave differently from those who joined the platform when it had become more established. Specifically, this date corresponds to day 2,000 since August 27, 2009, the first observed date in our data set. Appendix Table D.3 shows that our main findings are robust to alternative start dates. We also exclude users who created an account only after January 23, 2020, the beginning of the lockdown; these more recent users may be significantly different from users who joined before the lockdown.

the average probability that a user makes at least one post per day around the lockdown period. One notable pattern is that, after the lockdown, posting probability gradually declines.

Figure 1: Posting Probability



Notes. The two dashed lines mark the beginning and end of the lockdown.

3.1 Empirical Evidence

We now provide empirical evidence for a key implication of the neural autopilot model, namely that more reliable rewards lead to more habitual postings. To measure habitual postings for user i and date t, we use the number of consecutive posts in the next seven days ($\text{NConsecutivePosts}_{i,t}$).[9] Here "consecutive posts" refers to the number of posts made by user i within a continuous sequence of days after posting on date t. For example, if a user posts on dates $t, t+1$ and $t+3$, then $\text{NConsecutivePosts}_{i,t} = 1$.[10] To measure reward reliability for user i and date t, we compute the variance of the number of likes received per day by the user over the last seven days in which the

[9] Our measure of habitual postings focuses on the habit of posting every day, as opposed to posting every other day or every weekend. This is to follow the literature that studies habit in social media; for example, Anderson and Wood (2021) also focus on daily postings when they measure habitual postings.

[10] Note that when users post in the habit mode, by definition they post on consecutive days. Also note that when computing $\text{NConsecutivePosts}_{i,t}$, we use consecutive posts over the next seven days—not over the remainder of the observed time period—to avoid right censoring the data.

user posted ($\text{LikesVar}_{i,t}$). We then estimate the following regression:

$$\begin{aligned}\text{NConsecutivePosts}_{i,t} = \alpha_i + \delta_t + \beta_1 \log(\text{LikesVar}_{i,t}) + \beta_2 \log(\text{LikesMean}_{i,t}) \\ + \beta_3 \log(\text{LikesVar}_{i,t}) \times \log(\text{LikesMean}_{i,t}) + \beta_4 \text{PostToday}_{i,t} + \varepsilon_{i,t}. \quad (5)\end{aligned}$$

We control for the average number of likes ($\text{LikesMean}_{i,t}$) and its interaction with $\text{LikesVar}_{i,t}$ to capture the fact that consistently receiving zero likes may have a different effect on habitual postings than consistently receiving a high number of likes. We also control for whether user i posts on date t ($\text{PostToday}_{i,t}$).

Table 1: Number of Consecutive Posts Regressed on Reward Variables

	(1)	(2)	(3)	(4)
log(LikesVar)	–0.015	–0.048*	–0.035*	–0.065**
	(0.012)	(0.021)	(0.017)	(0.024)
log(LikesMean)	0.143**	0.128**	0.205***	0.185***
	(0.045)	(0.043)	(0.054)	(0.054)
log(LikesVar)× log(LikesMean)		0.029*		0.029+
		(0.013)		(0.015)
PostToday	1.378***	1.378***	1.397***	1.396***
	(0.054)	(0.054)	(0.054)	(0.054)
Last X posting days	7	7	10	10
Observations	1,082,076	1,082,076	969,090	969,090
R-squared	0.522	0.522	0.525	0.525
RMSE	0.86	0.86	0.90	0.90

Notes. This table reports the OLS estimates of Equation (5). For all columns, the dependent variable for user i on date t is the number of consecutive posts after date t until date $t+7$. In Columns 1 and 2, the main regressors are the mean and variance of the number of likes received per day by the user over the last 7 posting days. Columns 3 and 4 report robustness checks, in which the main regressors are the mean and variance of the number of likes received per day by the user over the last 10 posting days. All specifications include individual and date fixed effects. Standard errors are clustered by individuals and reported in parentheses. +, *, **, and *** indicate significance at the 10%, 5%, 1%, and 0.1% level, respectively.

Table 1 reports the estimates of Equation (5). It shows that β_1, the regression coefficient on log(LikesVar), is negative and statistically significant when controlling for the interaction term log(LikesVar) × log(LikesMean); moreover, β_3, the regression coefficient on the interaction term, is positive. This finding suggests that a reduction in the variance of likes increases the number of

consecutive posts in subsequent days, especially when the average number of likes is low. Despite the positive interaction effect, the marginal effect of log(LikesVar) is negative for the vast majority of the support of log(LikesMean) in our data.[11] The observed negative effect of log(LikesVar) on habitual postings (NConsecutivePosts) supports our model's implication that users receiving more reliable rewards are associated with more habitual postings.[12]

4 Structural Estimation

We estimate the neural autopilot model using maximum likelihood. Our estimation uses data on users' daily posting decisions.[13] For each user i and date t, we observe the user's posting decision $c_{i,t}$. If user i chooses to post, then $c_{i,t} = 1$. In this case, we observe the number of likes received, which serves as the empirical measure of the realized reward associated with $c_{i,t}$, and is denoted by $u_{i,t}$.[14] We note that users may derive different amounts of "value" from a like. In our model, we assume homogeneity in the value of a like. That is, one "like" is equally rewarding for all users. If user i chooses not to post, then $c_{i,t} = 0$. In this case, user i receives a reward from non-social media activities, denoted by u_0, but we, the econometricians, do not observe this reward. The interpretations of $u_{i,t}$ and u_0 are worth some discussions. We think of $u_{i,t}$ as a stimulus; it is transitory, and it triggers brain activities from user i. We think of u_0 as capturing users' intrinsic *preferences* toward non-social media activities, so it is interpreted as a utility level.

We form the likelihood of users' daily posting decisions according to the model described in Section 2. At each time t, the probability that user i posts depends on (1) the probability that the user is in the habit or goal-directed mode, and (2) the choice rules in these two modes. The model-implied probabilities are functions of the parameters specified in the neural autopilot model; moreover, they depend on $\boldsymbol{r}_{i,t} = (r_{i,t}(1), r_{i,t}(0))$ and $\boldsymbol{d}_{i,t} = (d_{i,t}(1), d_{i,t}(0))$, which represent the predicted reward and reward reliability for action $a \in \{0, 1\}$, respectively. Note that we do not directly observe $(\boldsymbol{r}_{i,t}, \boldsymbol{d}_{i,t})$ in the data. Instead, we compute them using choice and rewards data

[11] Appendix Figure D.1 plots the marginal effect of log(LikesVar) on habitual postings (NConsecutivePosts) from the 10th to 90th percentile value of log(LikesMean).

[12] The patterns in our data are also consistent with the findings of Anderson and Wood (2023) and Perez and Dickinson (2020) that choice behavior of habitual users is less sensitive to changes in rewards, compared to that of non-habitual users. We present our evidence in Appendix Table D.1.

[13] We use the term "choice" and "decision" interchangeably.

[14] Our model estimates are robust to using the average likes per post per day as the measure for the realized reward, as opposed to the total likes per day. We report these estimates in Appendix Table D.2.

$(c_{i,t}, u_{i,t})$ as well as the updating rules specified in Equations (1) and (2). Our estimation strategy is to search for the model parameters such that the observed posting behavior is most probable.

The set of model parameters we estimate includes: λ_r, the learning rate for reward prediction r; λ_d, the learning rate for reward reliability d; κ, a parameter in Equation (4) that captures the stochasticity in the switches between the habit mode and the goal-directed mode; ϕ, the threshold of reward reliability as shown in Equation (4); and finally, u_0, which represents the realized reward for the outside option—any activities other than posting on Weibo.[15] Given our focus on behavior change around the pandemic lockdown, we estimate u_0^{before}, u_0^{during}, and u_0^{after} separately for the reward value of the outside option before, during, and after the lockdown. Moreover, for each user i, we allow the reward reliability threshold ϕ to depend on user characteristics $\boldsymbol{X}_i$ (for example, the proportion of original posts); we assume $\phi_i = \phi_0 + \boldsymbol{\beta}'\boldsymbol{X}_i$ and estimate ϕ_0 and $\boldsymbol{\beta}$.[16] We use $\boldsymbol{\theta} \equiv (\lambda_r, \lambda_d, \kappa, \phi_0, \boldsymbol{\beta}, u_0^{before}, u_0^{during}, u_0^{after})$ to denote the vector of model parameters that we estimate. The estimation procedure is detailed in Appendix C.

4.1 Estimation Results

Table 2 presents the parameter estimates and their standard errors for our neural autopilot model. We find that λ_d, the learning rate for reward reliability, is significantly higher than λ_r, the learning rate for the predicted reward; that is, users learn reward reliability more quickly than they learn the predicted reward. We also find that ϕ, the threshold of reward reliability that drives the switches between the habit mode and the goal-directed mode, is higher for users with a higher proportion of original posts. This implies that users with a higher proportion of original posts are more likely to be in habit mode. More important, u_0^{during} is approximately equal to u_0^{after}. This suggests that our model explains the drop in posting probability post-lockdown through changes in posting rewards, not through changes in users' preferences toward non-social media activities. In other words, in a model where habit is neuroscientifically formalized, observed behavior change may be simply due to changes in reward prediction and reward reliability; changes in preferences are not required. In other words, our model can explain behavior change even when the estimated utility levels u_0^{during}

[15] When estimating the model, we fix the exploration parameter α in Equation (3) at one. This parameter cannot be jointly identified with u_0, the reward value for the outside option.

[16] For simplicity, $\boldsymbol{X}_i$ only includes the proportion of original posts in the main text. In Appendix Table D.4, $\boldsymbol{X}_i$ includes additional user characteristics, such as gender, whether user i is in a developed city, whether the user has a high credit score, and whether he or she has a verified account.

and u_0^{after} are basically the same.

Table 2: Parameter Estimates: Neural Autopilot Model

	Autopilot	
Parameters	Est.	Std. Err.
λ_r	0.009	(0.000)
λ_d	0.121	(0.001)
κ	7.291	(0.033)
ϕ_0	0.327	(0.001)
β	0.070	(0.002)
u_0^{before}	1.216	(0.001)
u_0^{during}	1.072	(0.005)
u_0^{after}	1.054	(0.004)
Log-likelihood	–389,048	
Number of users	1,848	
AIC	778,111	
BIC	778,211	

Our main estimates from Table 2 do not vary significantly when we allow additional individual heterogeneities to drive ϕ, the threshold parameter in Equation (4). In particular, Appendix Table D.4 presents the estimation results in which ϕ for each user is influenced not only by the user's proportion of original posts, but also by additional attributes that include the user's gender, whether the user has a verified account, is in a developed city, and has high Sesame Credit.[17] By comparing Table 2 and Table D.4, we find that the estimates for λ_r, λ_d, κ, u_0^{before}, u_0^{during}, and u_0^{after} are quite similar. Moreover, Table D.4 shows that for male, verified users who have low credit scores and post more original content, the threshold parameter ϕ tends to be higher, indicating that these users, compared to the other users, are more likely to be in the habit mode.

4.2 Comparison with Alternative Models

We compare the neural autopilot model with four alternative models of choice behavior. The first alternative model is an "autopilot + lagged choice" model. Although lagged choice is already part of the baseline neural autopilot model as it affects the updating rule for reward reliability and

[17]Sesame Credit is a private credit scoring system developed by Ant Group, an affiliate company of the Chinese conglomerate Alibaba Group.

reward prediction, this alternative model generalizes the neural autopilot model by allowing users' choice probabilities in the goal-directed mode to depend directly on their lagged choice. Specifically, when user i is in the goal-directed mode, the choice rule is now

$$\Pr(c_{i,t} = 1 | c_{i,t-1}, \boldsymbol{r}_{i,t}) = \frac{\exp(r_{i,t}(1) + \gamma \cdot c_{i,t-1})}{\exp(r_{i,t}(1) + \gamma \cdot c_{i,t-1}) + \exp(r_{i,t}(0))}, \tag{6}$$

where the coefficient γ on the lagged choice $c_{i,t-1}$ is an additional parameter we estimate.

Why is it sensible to include the lagged choice on the right hand side of Equation (6)? A low-level interpretation is that animals often exhibit "perseveration"—repetition of previous choices—that is seemingly independent of values and learning; see, for example, Miller et al. (2019) for evidence and a recent analysis. A higher-level interpretation is that dependence of current choice probabilities on lagged choice allows us to perform reduced-form estimations of adjacent complementarity between past choices and current utility of the same choices.

The second alternative model is a simple Q-learning model from the reinforcement learning literature (see Sutton and Barto, 2019 for a review). In this model, $r_t(a)$, the predicted reward for action a at time t, again evolves according to Equation (1). Moreover, action selection is governed by Equation (3); that is, users are *always* in the goal-directed mode. This Q-learning model can be viewed as a special case of our baseline neural autopilot model, with the threshold parameter ϕ in Equation (4) set to $-\infty$.

The third alternative model is a standard "epsilon-greedy" multi-armed bandits framework. As before, $r_t(a)$ evolves according to Equation (1). With probability $1 - \varepsilon$, action selection is given by

$$a_t = \arg\max_a \{r_t(a)\}. \tag{7}$$

With the remaining probability ε, action a_t is randomly selected across all possible choices. Here, the parameter ε controls the degree of exploration, and we estimate it using data.

The final alternative model we examine is a traditional "lagged choice" logit model. This model does not have a neuroscientific foundation. It captures choice persistence purely from the reduced-form association between lagged and current choices. In this model, the choice of user i at time t

is given by

$$c_{i,t} = \mathbb{1}_{\delta_0 c_{i,t-1} + \boldsymbol{\delta}_1' \boldsymbol{Z}_{i,t} + \varepsilon_{i,t} \geq 0}, \tag{8}$$

where $\boldsymbol{Z}_{i,t}$ is a vector of control variables that include the lagged posting reward from time $t-1$, user characteristics, and a dummy variable that indicates whether time t is before or during the lockdown. The coefficient δ_0 in Equation (8) captures the direct impact of the lagged choice $c_{i,t-1}$ on the utility from posting. The latent utility of the outside option after the lockdown is normalized to zero. This model is similar to those commonly used in the structural state dependence literature, in which the last period choice directly impacts the current period utility; recent examples include Dubé et al. (2010) and Allcott et al. (2022).

Table 3 presents the parameter estimates and their standard errors for the four alternative models described above. Columns 2 and 3 present the estimation results for the "autopilot + lagged choice" model. Compared to the baseline estimation results reported in Table 2, further including lagged choice in the model does not significantly change the estimates of many parameters. This suggests that the neural autopilot model captures habitization above and beyond structural state dependence. At the same time, there are some notable differences. First, adding lagged choice to the model shrinks λ_r toward zero. Note that λ_r captures the persistence of predicted reward across periods and hence contributes to choice dependency. As such, its estimation could be confounded by γ, the coefficient on the lagged choice $c_{i,t-1}$ from Equation (6). Second, adding lagged choice to the model leads to a larger κ and hence makes users more likely to switch between the habit mode and the goal-directed mode upon a given change in reward reliability. This in part offsets the choice persistence introduced by the lagged choice term $\gamma \cdot c_{i,t-1}$ in Equation (6). Finally, adding lagged choice to the model leads to a smaller ϕ_0, indicating that for identical levels of reward reliability, users in the "autopilot + lagged choice" model are more likely to be in the goal-directed mode, compared to users in the baseline neural autopilot model. In the baseline model, choosing repeated actions that yield lower expected rewards is primarily explained by the user being in the habit mode. However, in the "autopilot + lagged choice" model, this behavior is also in part driven by the user deriving utility directly from repeating the last period's choice. As such, a smaller proportion of repeated choices are attributed to the habit mode in the "autopilot + lagged choice"

model, leading to a lower ϕ_0.

Table 3: Parameter Estimates: Alternative Models

	Autopilot+Lagged Choice		Q-learning		Epsilon-Greedy		Logit	
Parameters	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
λ_r	0.001	(0.000)	0.004	(0.000)	0.009	(0.000)		
λ_d	0.093	(0.001)						
κ	23.627	(0.041)						
ϕ_0	0.127	(0.002)						
β	0.012	(0.003)						
u_0^{before}	1.527	(0.002)	2.864	(0.073)	2.232	(0.058)		
u_0^{during}	1.492	(0.003)	–3.631	(0.028)	0.110	(0.999)		
u_0^{after}	1.577	(0.008)	4.862	(0.100)	0.070	(1.000)		
ε					0.194	(0.333)		
c_{t-1}	1.334	(0.007)					2.822	(0.011)
Likes_{t-1}							0.012	(0.021)
Prop orig posts							–0.868	(0.009)
Before lockdown							–0.666	(0.008)
During lockdown							0.043	(0.002)
Constant							–1.957	(0.009)
Log-likelihood	–376,201		–547,961		–593,489		–448,485	
Number of users	1,848		1,848		1,848		1,848	
AIC	752,420		1,095,931		1,186,988		896,981	
BIC	752,532		1,095,981		1,187,050		897,056	

Columns 4 to 7 present the estimation results for the Q-learning model and the "epsilon-greedy" model. Compared to the baseline neural autopilot model, these two alternative models' fits to the data are significantly worse; their AIC and BIC are about 40% higher. These model comparisons highlight the importance of reward reliability and the switches between the habit mode and the goal-directed mode in driving users' decision making.[18]

Columns 8 and 9 present the estimation results for the "logged choice" logit model. This model generates a coefficient on the lagged choice $c_{i,t-1}$ that is positive and significant, indicating that current choice depends strongly on lagged choice. Moreover, the estimated differences between the

[18]For the "epsilon-greedy" model, the parameter estimates imply that with 19.4% probability, users fully randomize between posting and not posting. With the remaining 80.6% probability, users follow Equation (7) to choose the action that is associated with the highest reward prediction; in this case, users most likely choose not to post. Taken together, the model generates a posting probability of about 10%, matching the 9.4% posting probability observed in the data.

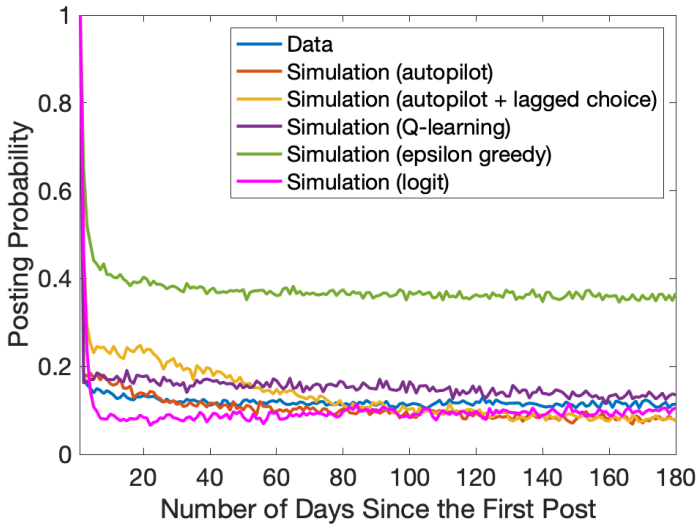
utility levels during and after the lockdown are sizeable and significant. This suggests that the logit model relies heavily on changes in preferences to explain the drop in the posting probability after the lockdown. In contrast, our neural autopilot model shows that once habit is neuroscientifically formalized, changes in preferences are no longer required to explain the observed behavior change.

4.3 Comparison between Model-Implied Choice Patterns and Data

With the parameter estimates at hand, we now study the model's implications for users' posting behavior in the time series. We compare the choice probabilities implied by each model—the neural autopilot model, the "autopilot + lagged choice" model, the Q-learning model, the "epsilon-greedy" model, and the logit model—with the observed data. For each user and each day, we use the simulated rewards and the respective model's parameter estimates to predict whether the user posts on that day. We then aggregate, for each event day, the model-implied posting behavior across all available users. For example, for event day 2 (i.e., one day after the day of the user's first post), we compute the model-implied posting probability as the total number of users who post on the next day following their first post, divided by the total number of such users. Figure 2 plots, for each event day since a user's first post, the posting probability implied by each of the five models; as a comparison, it also plots the actual posting probability observed in the data.

The actual posting probability observed in the data (blue line) exhibits two notable features. First, the posting probability drops significantly almost immediately after the first day; only 17% of users post on the second day. Second, over subsequent periods, the posting probability gradually decreases and eventually converges to 10%. The posting probability implied by the neural autopilot model (red line) captures these features well: the posting probability on the second day is about 18%, very close to the empirically observed level; and over subsequent periods, the posting probability decreases gradually. In contrast, the posting probability implied by the "autopilot + lagged choice" model (yellow line) does not fit well with the data; on the second day, it is significantly higher than the empirically observed posting probability. It also takes more days to converge to the steady state, compared to the autopilot simulation and the observed data. This occurs because the "autopilot + lagged choice" model imposes an additional layer of "state-dependence" through the lagged choice term $\gamma \cdot c_{i,t-1}$ in Equation (6). As such, the model-implied posting probability on the second day depends heavily on the users' behavior on the first day, where by construction

Figure 2: Users' Posting Behavior in the Time Series: Data versus Model Simulations



Notes. This figure plots, for each event day since a user's first post, the posting probability implied by the neural autopilot model, the "autopilot + lagged choice" model, the Q-learning model, the "epsilon-greedy" model, and the logit model, respectively. As a comparison, it also plots the actual posting probability observed in the data.

every user posts.

Moreover, the posting probability implied by the logit model (magenta line) also does not fit well with the data: on the second day, it is significantly lower than the empirically observed posting probability; and over subsequent periods, it quickly converges to the steady state level of about 10%. The logit model does not allow the dynamic evolutions of predicted reward and reward reliability to directly affect users' behavior. Finally, the posting probabilities implied by the Q-learning model (purple line) and the "epsilon-greedy" model (green line) fail to capture the key features observed from the actual posting probability. The Q-learning model tends to overpredict the posting probability because it assumes that users are always in a goal-directed mode. The "epsilon-greedy" model generates posting probabilities that are too high and too persistent. From the first day of posting, users might receive likes and hence update positively their predicted reward from posting; the action selection rule in Equation (7) then implies that these users have a high probability of posting again on the second day.

In summary, Figure 2 suggests that the neural autopilot model offers significant explanatory power of changes in observed behavior above and beyond traditional models of state dependence. Moreover, the autopilot model also outperforms widely-used reinforcement learning models such as

Q-learning and the "epsilon-greedy" model.

5 Counterfactual Analysis

Our empirical analysis adopts a structural estimation approach, rather than a "reduced-form regression" approach. Our structural model hypothesizes decision rules that are followed by purposive agents who face a choice set, a set of environmental states, and some *policy-invariant* parameters that allow for natural interpretations.[19] Note that many cognitive science studies, such as the Lindström et al. (2021) paper on reinforcement learning from social media, use the term "generative" model, which is synonymous with our use of "structural" model.[20]

By fitting the structural model with our data, we estimate parameters that govern the neuroscientific process of habit formation. Then, with the parameter estimates, we conduct thought experiments—in other words, counterfactual analysis—to investigate what factors create and sustain habits. This structural estimation approach is particularly useful for our case in which the decision making process is complex involving multi-layered stochasticity and individual heterogeneity and conducting lab or field experiments is costly.

In this section, we evaluate two counterfactual policies relating to the reward schedules: increasing the average level of posting rewards during the lockdown, and reducing the volatility of posting rewards during the lockdown. These two counterfactual policies allow us to assess the importance of reward levels and reward reliability in creating and sustaining habitual postings. We also evaluate counterfactual policies relating to users' learning rates. More specifically, we explore the role of the learning rates in promoting habits.

Evaluating these counterfactual policies requires a baseline scenario. We simulate posting decisions for 5,000 users over a total of 2,000 time periods. For all users, we assume that the lockdown starts at day 500 and ends at day 600. Conditional on posting, we assume that the number of likes is drawn from a lognormal distribution, denoted by Lognormal(μ, σ^2), and is i.i.d. over time. We set $\mu = 1$ and $\sigma^2 = 8$ such that the mean and variance of the simulated likes are similar to those

[19] Policy-invariant parameters refer to parameters that remain unchanged across different policy regimes or interventions. These parameters are an essential component of the structural model and are unaffected by variations in external conditions. For example, learning rates are assumed to remain constant, regardless of the reward schedule.

[20] However, to our knowledge, the extension of structural models to forecast behavior in counterfactual scenarios is not standard in cognitive science generative modelling. For example, this is not done in Lindström et al. (2021).

of actual likes observed in the data. Finally, conditional on not posting, we set u_0, the utility from the outside option, to the estimated levels reported in Table 2.

We compare each counterfactual policy with this baseline scenario.[21] In the first counterfactual policy, we increase the average level of simulated rewards during the lockdown by changing μ from 1 to 2; outside the lockdown, μ remains at 1. In the second counterfactual policy, we reduce the volatility of simulated rewards during the lockdown by changing σ^2 to zero; outside the lockdown, σ^2 remains at eight.

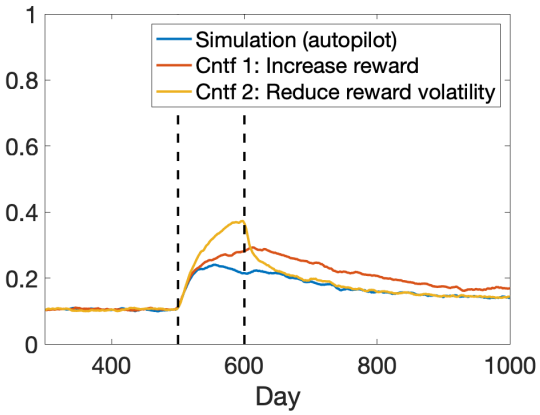
Figures 3 plots the average posting probability, the average probability that users are in the goal-directed mode, the average probability that users are in the habit mode of posting, and the average probability that users are in the habit mode of not posting. The baseline scenario and the counterfactual policies confirm the rationale for using the lockdown as a shock that interrupts habitual behavior: when the utility level of the outside option changes, users are likely to switch from the habit mode to the goal-directed mode.

The first counterfactual policy shows that an increase in the level of rewards leads to a significant increase in the posting probability, and this effect lasts for many periods after the lockdown ends. Moreover, the increase in rewards leads to an increase in the probability that users are in the goal-directed mode. Through the lens of our model, these findings suggest that behavior change associated with an increase in rewards is unlikely due to habit formation; rather, users adjust the predicted rewards and actively choose the option that has the highest predicted reward.

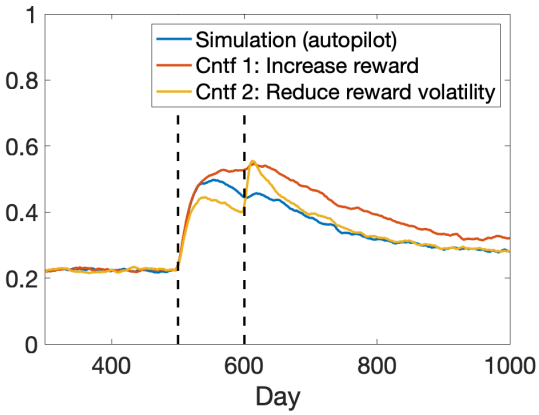
The second counterfactual policy shows that a reduction in reward volatility also leads to a significant increase in the posting probability, although the effect disappears soon after the lockdown ends. Importantly, this increase in the posting probability is due to habit: Figure 3c shows a notable uptick in the probability that users are in the habit mode of posting. In our neural autopilot model, lower reward volatility makes the reward prediction more reliable. As such, the agent is more likely to engage in habitual postings. The comparison between the two counterfactual policies suggests that the second moment of the reward process can be more important than the first moment in creating habits.

[21] When simulating posting decisions for the baseline scenario and the counterfactual policies, we calibrate $r_{i,1}(1)$, the initial value of the predicted reward from posting, to the average number of likes we observe from a post. Moreover, we calibrate $r_{i,1}(0)$, the initial value of the predicted reward from not posting, to u_0^{before}, the estimated utility level of the outside option during the pre-lockdown period.

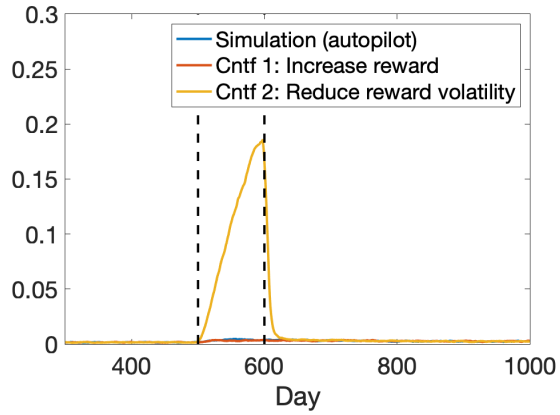
Figure 3: Counterfactual Policies



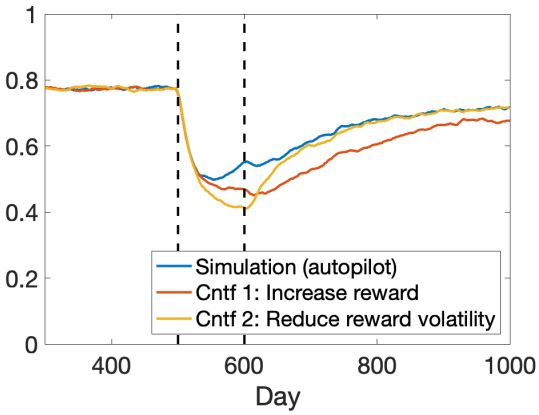
(a) Posting Probability



(b) Pr(goal-directed)



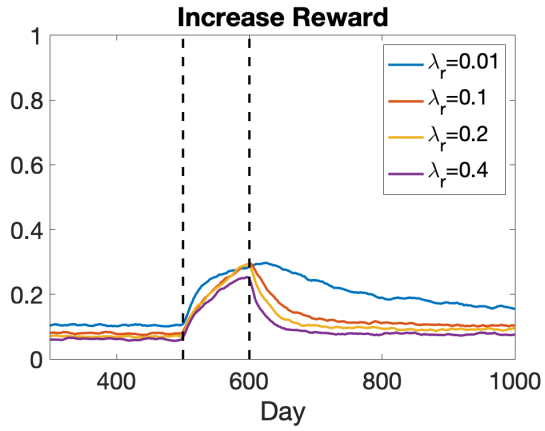
(c) Pr(habit mode of posting)



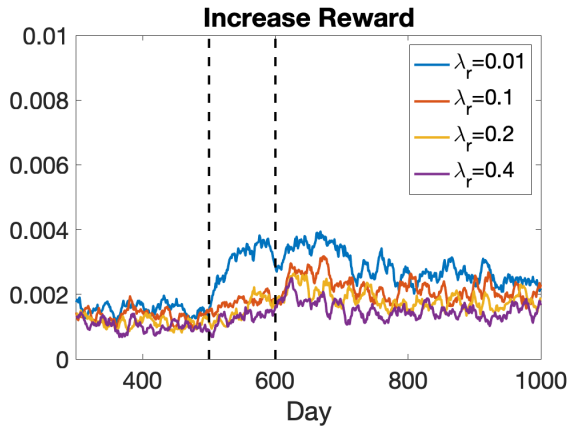
(d) Pr(habit mode of not posting)

Given the importance of learning rates in driving users' posting behavior, we further evaluate counterfactual policies for users with different values of λ_r and λ_d. Figure 4 plots, for users with different values of λ_r, the average posting probability and the average probability that users post in the habit mode. We find that users with a lower λ_r have a more persistent response to the increase in posting rewards: when λ_r is low, the predicted reward for posting decreases slowly after the drop in rewards that happens at the end of the lockdown period; as such, the posting probability remains elevated for many periods after the lockdown (see Figure 4a). Interestingly, users with a lower λ_r also have a stronger response to the reduction in reward volatility: for these users, the probability of posting in the habit mode increases more rapidly during the lockdown.

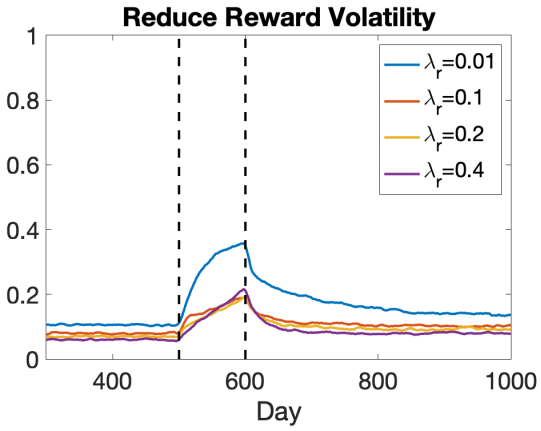
Figure 4: Counterfactuals: Users with Different λ_r



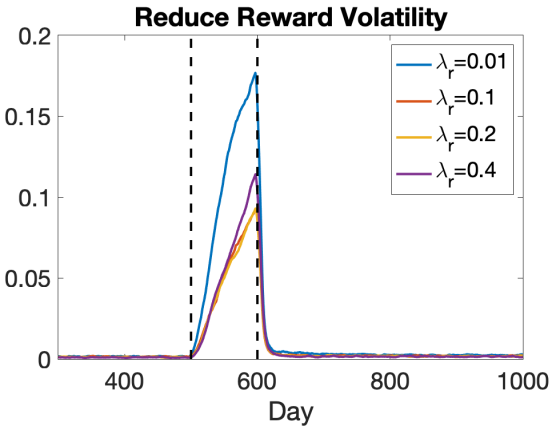
(a) Cntf 1: Posting Probability



(b) Cntf 1: Pr(habit mode of posting)

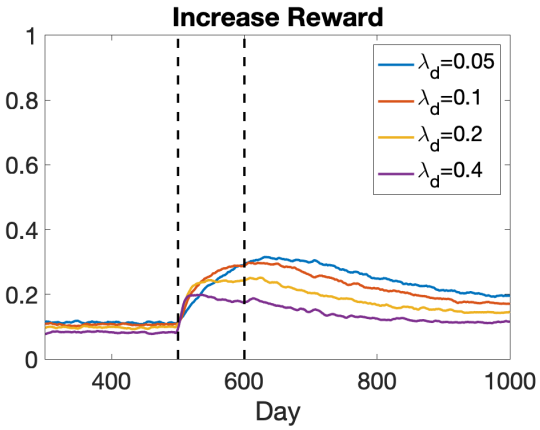


(c) Cntf 2: Posting Probability

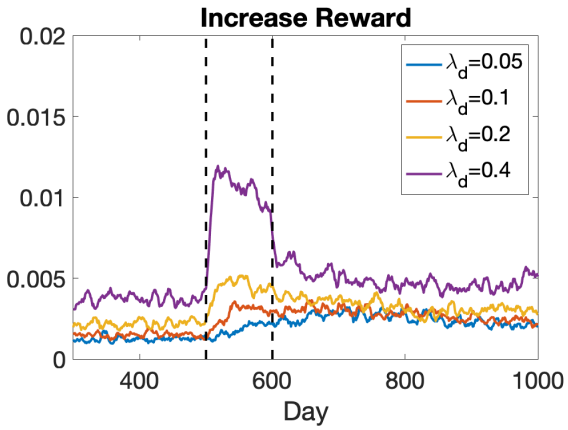


(d) Cntf 2: Pr(habit mode of posting)

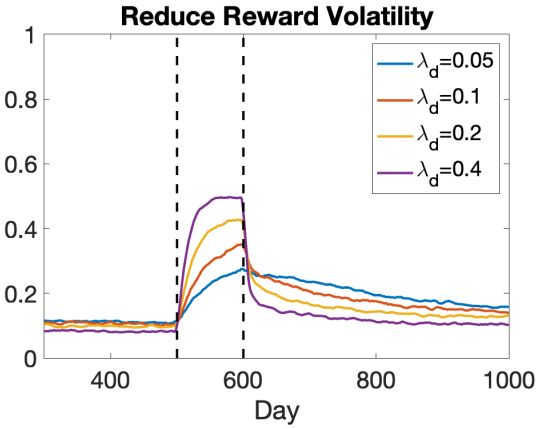
Figure 5: Counterfactuals: Users with Different λ_d



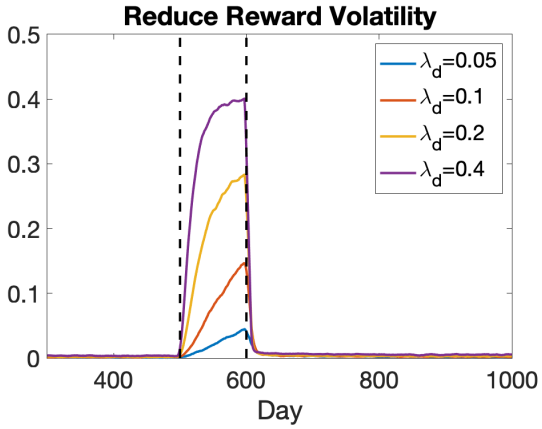
(a) Cntf 1: Posting Probability



(b) Cntf 1: Pr(habit mode of posting)



(c) Cntf 2: Posting Probability



(d) Cntf 2: Pr(habit mode of posting)

Figure 5 plots, for users with different values of λ_d, the average posting probability and the average probability that users post in the habit mode. We find that users with a higher λ_d have a stronger response to the reduction in reward volatility. For instance, reducing reward volatility during the lockdown—changing σ^2 from 8 to 0— causes a 40% increase in the probability of posting in the habit mode for users with a λ_d of 0.4, as compared to a 20% increase for those with a λ_d of 0.1. Moreover, when reward volatility returns to the baseline level after the lockdown, users with a high λ_d will soon find the predicted reward for posting unreliable. Therefore, their posting probability drops precipitously (see Figures 5c and 5d).

In summary, our counterfactual analysis shows that reducing reward volatility is more effective than increasing the level of rewards in creating habits; reducing reward volatility is particularly effective in promoting habit formation for users with a high λ_d, the learning rate for reward reliability. Moreover, the counterfactual policies we consider have direct implications for how social media platforms should design recommendation algorithms (e.g, how Facebook should select and display posts via Feed). If a platform's objective is to increase social media engagement, our findings suggest that maintaining a *steady* level of likes can be important for creating and sustaining habitual postings.

6 Conclusion

In the neural autopilot model, agents learn the value of choices through reinforcement, updating the predicted reward using the reward prediction error (RPE). The absolute value of the RPE is also used to compute the reliability of the predicted reward. If a choice has consistently yielded low absolute RPEs, then it is safe to repeat the same choice—this is what constitutes a habit.

We use large-scale, individual-level field data to estimate and test the neural autopilot model. Note that these types of neuroeconomics models were created and honed on much simpler, stylized paradigms, typically with animals that make hundreds of regular simple choices such as tapping a lever for food pellets. In this paper, we test whether the same basic model can fit more complicated human choices.

The structural estimation yields interpretable parameter estimates that are consistent with intuitions about habit and with prior research. For example, the learning rates we estimate from

field data are in the ballpark of values derived in human decision neuroscience for abstract reward learning. Moreover, our counterfactual simulation shows that reducing the volatility in posting rewards leads to a significant increase in habitual postings; this is due to the importance of reward *reliability*—not just rewards—in creating habit.

We compare our model with alternative models of choice behavior. If the apparent habitization captured by the neural autopilot model was just a complicated way of approximating choice persistence, then including lagged choice in the model should erase the influence of parameters associated neural autopilot. This does not happen: Table 3 shows that these parameters remain significant. At the same time, the added term on lagged choice significantly predicts current choice, and it leads to an incremental improvement in the model's fit. This means that there is some residual effect of previous choice on current choice, on top of all the learning and autopilot structure. Taken together, our results suggest that the neural autopilot model offers significant explanatory power of actual choices above and beyond traditional models of state dependence. We also compare the neural autopilot model with widely-used reinforcement learning models such as Q-learning and the "epsilon-greedy" model. We find that the neural autopilot model significantly outperforms these alternative reinforcement learning models.

There are many opportunities for future research about the neural autopilot model. The amount of information processing required in the habit mode is significantly lower than in the goal-directed mode. In the habit mode, agents only need to recall their previous choice c_{t-1}, the reward reliability of that choice $d_t(c_{t-1})$, and then compare $d_t(c_{t-1})$ to a threshold ϕ, which is a simple computation.[22] In the goal-directed mode, however, agents need to evaluate all choices, and this is much more effortful. In principle then, attention measures, response times, and other mental effort measures could be used along with choice data to identify the habit and goal-directed modes. Moreover, theorists studying rational inattention and resource rationality may be interested in figuring out whether the neural autopilot model is an optimized response to an environment with nonstationarity and costs of mental effort or attention. Lastly, some important features of habit, such as context dependence, are not incorporated by our neural autopilot model. Extending the model to include such features can make it more broadly useful.

[22] A habitized person does not even need to recall the predicted reward value; the fact that a choice was made previously is like a proxy for the choice having high predicted reward.

References

Allcott, H., M. Gentzkow, and L. Song (2022): "Digital Addiction," American Economic Review, 112, 2424–63.

Anderson, I. A. and W. Wood (2021): "Habits and the Electronic Herd: The Psychology Behind Social Media's Successes and Failures," Consumer Psychology Review, 4, 83–99.

——— (2023): "Social Motivations' Limited Influence on Habitual Behavior: Tests from Social Media Engagement," Motivation Science, 9, 107–119.

Arad, A., U. Gneezy, and E. Mograbi (2023): "Intermittent Incentives to Encourage Exercising in the Long Run," Journal of Economic Behavior & Organization, 205, 560–573.

Barberis, N. and L. J. Jin (2023): "Model-free and Model-Based Learning as Joint Drivers of Investor Behavior," Working paper.

Baum, W. M. (1973): "The Correlation-Based Law of Effect," Journal of the Experimental Analysis of Behavior, 20, 137–153.

Becker, G. S. and K. M. Murphy (1988): "A Theory of Rational Addiction," Journal of Political Economy, 96, 675–700.

Bernheim, B. D. and A. Rangel (2009): "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics," Quarterly Journal of Economics, 124, 51–104.

Brady, W. J., K. McLoughlin, T. N. Doan, and M. J. Crockett (2021): "How Social Learning Amplifies Moral Outrage Expression in Online Social Networks," Science Advances, 7, eabe5641.

Burton, J. W., N. Cruz, and U. Hahn (2021): "Reconsidering Evidence of Moral Contagion in Online Social Networks," Nature Human Behaviour, 5, 1629–1635.

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

Cerigioni, F. (2021): "Dual Decision Processes: Retrieving Preferences When Some Choices Are Automatic," Journal of Political Economy, 129, 1667–1704.

Cesa-Bianchi, N., C. Gentile, G. Lugosi, and G. Neu (2017): "Boltzmann Exploration Done Right," in Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, USA: Curran Associates Inc., NIPS'17, 6287–6296.

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

Das, S. and A. Lavoie (2014): "The Effects of Feedback on Human Behavior in Social Media: An Inverse Reinforcement Learning Model," in Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems, Citeseer, 653–660.

Daw, N., Y. Niv, and P. Dayan (2005): "Uncertainty-Based Competition between Prefrontal and Dorsolateral Striatal Systems for Behavioral Control," Nature Neuroscience, 8, 1704–1711.

Daw, N. D., J. P. O'Doherty, P. Dayan, B. Seymour, and R. J. Dolan (2006): "Cortical Substrates for Exploratory Cecisions in Humans," Nature, 441, 876–879.

Dolan, R. J. and P. Dayan (2013): "Goals and Habits in the Brain," Neuron, 80, 312–325.

Doya, K., K. Samejima, K.-i. Katagiri, and M. Kawato (2002): "Multiple Model-Based Reinforcement Learning," Neural Computation, 14, 1347–1369.

Dubé, J.-P., G. J. Hitsch, and P. E. Rossi (2010): "State Dependence and Alternative Explanations for Consumer Inertia," The RAND Journal of Economics, 41, 417–445.

Evans, J. S. B. T. (2008): "Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition," Annual Review of Psychology, 59, 255–278.

Evans, J. S. B. T. and K. E. Stanovich (2013): "Dual-Process Theories of Higher Cognition: Advancing the Debate," Perspectives on Psychological Science, 8, 223–241.

Frydman, C. and L. J. Jin (2022): "Efficient Coding and Risky Choice," Quarterly Journal of Economics, 137, 161–213.

——— (2023): "On the Source and Instability of Probability Weighting," Working paper.

Jacobs, R. A., M. I. Jordan, S. J. Nowlan, and G. E. Hinton (1991): "Adaptive Mixtures of Local Experts," Neural Computation, 3, 79–87.

Jog, M. S., Y. Kubota, C. I. Connolly, V. Hillegaart, and A. M. Graybiel (1999): "Building Neural Representations of Habits," Science, 286, 1745–1749.

Kahneman, D. (2011): Thinking, Fast and Slow, Farrar, Straus and Giroux.

Keane, M. P. (1997): "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior," Journal of Business & Economic Statistics, 15, 310–327.

Khaw, M. W., Z. Li, and M. Woodford (2021): "Cognitive Imprecision and Small-States Risk Aversion," Review of Economic Studies, 88, 1979–2013.

Laibson, D. (2001): "A Cue-Theory of Consumption," Quarterly Journal of Economics, 116, 81–119.

Landry, P., R. Webb, and C. Camerer (2021): "A Neural Autopilot Theory of Habit," Working paper.

Larcom, S., F. Rauch, and T. Willems (2017): "The Benefits of Forced Experimentation: Striking Evidence from the London Underground Network," Quarterly Journal of Economics, 132, 2019–2055.

Lee, S. W., S. Shimojo, and J. P. O'Doherty (2014): "Neural Computations Underlying Arbitration between Model-Based and Model-free Learning," Neuron, 81, 687–699.

Lindström, B., M. Bellander, D. T. Schultner, A. Chang, P. N. Tobler, and D. M. Amodio (2021): "A Computational Reward Learning Account of Social Media Engagement," Nature Communications, 12, 1311.

Maki, A., R. J. Burns, L. Ha, and A. J. Rothman (2016): "Paying People to Protect the Environment: A Meta-Analysis of Financial Incentive Interventions to Promote Proenvironmental Behaviors," Journal of Environmental Psychology, 47, 242–255.

Miller, K. J., A. Shenhav, and E. A. Ludvig (2019): "Habits Without Values," Psychological Review, 126, 292–311.

Mowrer, O. H. and H. Jones (1945): "Habit Strength as a Function of the Pattern of Reinforcement," Journal of Experimental Psychology, 35, 293–311.

Narendra, K., J. Balakrishnan, and M. Ciliz (1995): "Adaptation and Learning Using Multiple Models, Switching, and Tuning," IEEE Control Systems Magazine, 15, 37–51.

O'Doherty, J. P., P. Dayan, K. Friston, H. Critchley, and R. J. Dolan (2003): "Temporal Difference Models and Reward-Related Learning in the Human Brain," Neuron, 38, 329–337.

Pearce, J. M. and M. E. Bouton (2001): "Theories of Associative Learning in Animals," Annual Review of Psychology, 52, 111–139.

Perez, O. D. and A. Dickinson (2020): "A Theory of Actions and Habits: The Interaction of Rate Correlation and Contiguity Systems in Free-Operant Behavior," Psychological Review, 127, 945–971.

Pollak, R. A. (1970): "Habit Formation and Dynamic Demand Functions," Journal of Political Economy, 78, 745–763.

Rangel, A., C. Camerer, and P. R. Montague (2008): "A Framework for Studying the Neurobiology of Value-Based Decision Making," Nature Reviews Neuroscience, 9, 545–556.

Rescorla, R. A. and A. R. Wagner (1972): "A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement," in Classical Conditioning II: Current Theory and Research, Appleton-Century-Crofts, New York.

Seetharaman, P., A. Ainslie, and P. K. Chintagunta (1999): "Investigating Household State Dependence Effects Across Categories," Journal of Marketing Research, 36, 488–500.

Stanovich, K. E. (1999): Who is Rational? Studies of Individual Differences in Reasoning, Mahwah, NJ: Elrbaum.

Sutton, R. S. and A. G. Barto (2019): Reinforcement Learning: An Introduction, MIT Press.

Thorndike, E. L. (1932): The Fundamentals of Learning, New York, NY, US: Teachers College Bureau of Publications.

Verplanken, B. and S. Orbell (2022): "Attitudes, Habits, and Behavior Change," Annual Review of Psychology, 73, 327–352.

Wachter, J. A. and M. J. Kahana (2023): "A Retrieved-Context Theory of Financial Decisions," Quarterly Journal of Economics, forthcoming.

Webb, R., P. Landry, M. Osborne, C. Zhao, and C. Camerer (2022): "A Neural-Autopilot Theory of Habit: Evidence from Canned Tuna," Working paper.

Wellsjo, A. S. (2021): "Simple Actions, Complex Habits: Lessons from Hospital Hand Hygiene," Working paper.

Wood, W. and D. T. Neal (2009): "The Habitual Consumer," Journal of Consumer Psychology, 19, 579–592.

Zajkowski, W. K., M. Kossut, and R. C. Wilson (2017): "A Causal Role for Right Frontopolar Cortex in Directed, but not Random, Exploration," eLife, 6, e27430.

A Reinforcement Schedules and the Autopilot Model

There is a lot of evidence from animal and human learning about how motivation and habit are affected by various reinforcement "schedules." The stylized facts from this long line of research may present a challenge to the idea, originating in Daw et al. (2005) and at the core of the role of $d_t(a)$, that more predictable reward creates habit. The goal of this section is to briefly describe the stylized facts and speculate about the relation between these facts and the hypothesized role of reward predictability in neural autopilot.

To fix terminology, a reinforcement schedule is a specification of when rewarding reinforcers are delivered. Schedules can be based on real clock time, previous behavioral history, or randomness. In lab experiments of animal and human learning, participants are usually *not* explicitly instructed about the reinforcement schedule; they have to learn the schedule from trial and error. However, in most field experiments on behavior change, including the two experiments described below, human participants are instructed about the reinforcement schedule.

There are two distinct types of reinforcement schedules: continuous reinforcement and partial reinforcement. Continuous reinforcement refers to a reinforcement schedule in which every response is followed by the rewarding reinforcer. By contrast, partial reinforcement refers to a reinforcement schedule in which response is reinforced only a fraction of the time. Partial reinforcement schedules can be classified into two main categories: "ratio" schedules and "interval" schedules. And within each of these two categories, there are two types of reinforcement patterns: fixed patterns and variable patterns.

A "fixed-ratio" schedule delivers the rewarding reinforcer after a fixed number of responses; one example is a coupon for a free coffee offered after every tenth purchase. A "variable-ratio" schedule delivers the rewarding reinforcer after a random number of responses, but with a fixed *average* ratio of rewards to responses; one example is that after each response, a reward is offered with a $p = 10\%$ probability, independent of the reward history (this is also called an R10 schedule, because on average, a reward is delivered every $r = 1/p = 10$ trials). Note that in both the fixed-ratio and variable-ratio examples, the long-run average is to receive reward on 10% of the trials; however, the timing of the reward is different across the two examples.

In contrast to ratio schedules, interval schedules deliver rewards after elapsed time intervals. In

a "fixed-interval" schedule, the time interval is kept at a constant. In a "variable-interval" schedule, the amount of time or behavior that triggers reward is random; for example, the rewarding device might become "baited" after some time interval that is between 3 and 6 seconds (this is to mimic ecological processes such as consuming fruit that ripen after some random period of time).

In domains such as animal training, where the sequence of behavior is controlled by a trainer, continuous reinforcement clearly works best to create habits (e.g., Mowrer and Jones, 1945). This is consistent with the mathematics of the autopilot model, in which habitization arises from a reduction in reward reliability $d_t(a)$. A fixed amount of continuous reinforcement reduces reward prediction errors (RPE) as quickly as possible and hence reduces $d_t(a)$ toward zero as quickly as possible. In this kind of animal training, there is often interest in eventually "thinning" the reinforcement schedule—transitioning to partial reinforcement—slowly enough to maintain an animal's habits.

One well-accepted stylized fact is that variable-ratio schedules are less effective in creating habits, compared to variable-interval schedules. An influential theory, due to Perez and Dickinson (2020) and precursors (Baum, 1973), posits that a crucial feature for forming habits is the correlation between the *rates* of reinforcement and behavior, as aggregated in temporal memory windows; this is called "rate correlation" theory. Consider a variable-ratio schedule in which the rewarding reinforcer is delivered with a $p = 1/r$ probability on each trial. Suppose the memory windows are of T-period long. If r is not too large relative to T, then the percentage of rewards and the percentage of responses observed in these memory windows tend to be correlated. In these variable-ratio schedules, Perez and Dickinson (2020) show, using animal data, that rate-correlated behavior does not lead to habitization. However, in variable-interval schedules with low empirical rate correlation, habits are more strongly formed.

Moreover, studies in consumer behavior present evidence consistent with the notion that low rate correlation sustains habits. As Wood and Neal (2009) wrote (page 586):

> "Early in habit learning, rewards promote repetition. Rewards also facilitate the transition from outcome-oriented to context-cued responding when they are presented in ways that minimize the experience of the contingency between the behavior and the rewarding outcome (i.e, low rate correlation)."

Two questions arise regarding how the animal learning data from continuous and partial reinforcement schedules are linked to the crucial role of the reward reliability variable $d_t(a)$ in the neural autopilot model. The first question is concerned about whether learning with variable-interval schedules implies higher predictability—that is, lower $d_t(a)$—and stronger habit formation, given the evidence that variable-interval schedules are better at creating habits, compared to variable-ratio schedules. The second question is about whether consistency exists between data and the autopilot model's implication that continuous reinforcement, rather than partial reinforcement, leads to the fastest reduction in $d_t(a)$ and hence the quickest formation of habits. Figuring out clear answers to these two questions is beyond the scope of the paper; it likely requires the introduction of a new model of habit formation, and it likely requires structurally estimating and testing this new model using field data. Nonetheless, we provide some preliminary thoughts below.

We start with the first question: Does learning with variable-interval schedules imply lower $d_t(a)$ and stronger habit formation, given the evidence that variable-interval schedules are better at creating habits, compared to variable-ratio schedules? We first examine the evolution of $d_t(a)$ under a variable-ratio schedule. Recall that Equation (2), the updating rule for $d_t(a)$ in the neural autopilot model, does not contain the memory-smoothing windows that Perez and Dickinson (2020) study. As such, a variable-ratio schedule will likely generate a large positive prediction error once the reward is dispensed. Then, over the next few periods, in absence of another reward, the reward prediction error will tend to be substantial and negative; and gradually, it will become less negative, until the occurrence of the next reward. Overall, variable-ratio schedules might generate high time-series variability in RPEs in the autopilot model, such that $d_t(a)$ never becomes sufficiently low to induce habit formation. It is also conceivable that with a higher reward probability $p = 1/r$ or with a lower learning rate λ_d, variable-ratio schedules will induce higher variability in RPEs, making it less likely for the autopilot model to create habits. These results—which are conjectured rather than proven—are consistent with the prediction of rate correlation theory that variable-ratio schedules are not conducive to habit formation.

We now examine the evolution of $d_t(a)$ under a variable-interval schedule. Recall the stylized fact that variable-interval schedules are more likely to create habits, compared to variable-ratio schedules. Empirically, animal learning under interval reward schedules typically shows a pattern of "scalloping": there is a low rate of response until the earliest time at which the next reward

is likely to arrive. And this is especially true for animals who can more quickly learn when the interval time has passed and respond accordingly. Consider a fixed-interval schedule with a reward of either one or zero. In this case, a scalloping animal will wait until the interval length has passed, respond with a predicted reward of $r_t(a) = 1$, earn the reward, and experience a reward prediction error of zero. Intuitively, a scalloping animal in free operant conditioning will essentially transform the fixed-interval schedule into a sequence of continuous reinforcements. According to the neural autopilot model, this animal will have a low $d_t(a)$ and hence form strong habits.[23]

Let us now consider a variable-interval schedule. Suppose the random interval ranges from 3 to 5 seconds. In this case, a semi-scalloping animal will likely wait three seconds and then begin to respond.[24] If the range of the random intervals is narrow, the reward prediction errors will likely be small in magnitude; as such, $d_t(a)$ will be close to zero, and the autopilot model implies that the animal is likely to form habits. Conversely, if the range of the random intervals is wide, then soon after the random interval onset begins, the reward prediction errors will likely be high in magnitude; as such, $d_t(a)$ will likely be large, and according to the model, habits are less likely to form.

Together, our conjectures can be summarized as follows: (1) with a continuous reinforcement schedule, the reward reliability variable $d_t(a)$ in the neural autopilot model will likely be low, hence predicting strong habitization, (2) with a variable-ratio schedule, $d_t(a)$ will likely be high, hence predicting a lack of habitization, and (3) with a variable-interval schedule, the magnitude of $d_t(a)$ will likely depend on the range of the random intervals.

We now turn to the second question: Are the data consistent with the autopilot model's implication that continuous reinforcement, rather than partial reinforcement, leads to the fastest reduction in $d_t(a)$ and hence the quickest formation of habits? In our understanding, there is not much direct, replicated evidence in *human field data* that partial reinforcement produces stronger and more persistent habit-like choice than continuous reinforcement. However, a prominent laboratory finding,

[23]The only complication here is the added assumption that the animal decides which period is optimal for it to respond. The neural autopilot model, in its current form, does not have this feature; however, the model can be easily extended to allow for it.

[24]How animals actually behave under random intervals will depend on whether reward predictions are simply backward-looking—as modelled in Equation (1)—or use information about the expected arrival of future reward. For example, when the animals are prepared to respond after 3 seconds, they may predict a reward of 1/3. This idea of injecting an element of model-based thinking into the autopilot structure can be a sensible direction for future adaptations of the model.

the "partial reinforcement extinction effect" (PREE), is worth noting. The PREE effect refers to the fact that during extinction (trials with no reinforcement), animals tend to lever-press more frequently and more persistently if they had been partially reinforced rather than continuously reinforced. This effect is essentially the same as insensitivity to outcome devaluation, which is thought to be a hallmark of habit.[25]

While the PREE effect is well-established in lab experiments, it is not clearly established in human field data. Below, we briefly discuss two field studies that we view as representative. The first field study compares the causal effect on gym attendance from three different types of reward schedules—a continuous per-visit payment, an intermittent incentive schedule of monetary rewards at increasing intervals, and an intermittent schedule of monetary rewards with unpredictable timing (Arad et al., 2023). During the two-month treatment period, all three reward schedules led to more gym visits, compared to a no-incentive control group. Moreover, the continuous payment schedule actually led to slightly more gym visits, and more participants visiting at least once, than the intermittent schedule of monetary rewards with unpredictable timing.

The second field study is a meta-analysis of 30 effect sizes from environmentally-friendly incentives (Maki et al., 2016). The effects in changing behaviors were generally positive and substantial; the effect size d-value ranges from 0.30 to 0.45. When incentives were applied, variable reinforcement was more effective than fixed reinforcement ($d = 0.45$ vs. $d = 0.30$). However, after incentives were removed, the effects were similar and not significantly different ($d = 0.35$ vs. $d = 0.44$); the comparison of post-incentive effects is hampered by the small sample size of reported post-incentive behaviors.

The discussion of this section leads us to draw two conclusions. First, any autopilot-type model based on reward predictability needs to explain animal and human evidence from continuous and partial reinforcement; the model needs to be modified if it cannot explain well-accepted evidence. Second, there is a big leap from cued and free operant conditioning of animals in lab environments to human choice in field settings. As we have discussed in the main text, it does seem evident that humans are reinforced by social media rewards in ways that often conform to simple reinforcement learning models; so there is promise for unified principles that govern animal and human choice.

[25] Also note that the PREE effect can be rationalized as Bayesian learning in which animals take more trials to learn that reward has ended after partial reinforcement than after continuous reinforcement. Bayesian learning is an input to goal-directed or model-based choice, rather than to habitual choice.

At the same time, it is also possible that the ideal mathematical descriptions applied to humans in field settings are not quite the same as the mathematical descriptions applied to animals in lab environments. Figuring out the differences and similarities of these mathematical descriptions should be a priority for future research.

B Data Collection and Cleaning

We randomly sampled a set of Weibo users using their 10-digit account number; this number is unique to each account and allows us to access the user's profile and activity pages. For example, "https://m.weibo.cn/u/1669879400" is linked to the profile and activity pages of user "1669879400." For each user, we obtain a complete timeline of posts and platform activities that are publicly available online.

Weibo allows users to set their accounts to a private mode. When this happens, only posts within the last six months are publicly available. Because we cannot capture the entire timeline of posting behavior for users whose accounts are in private mode at the time of data collection, we exclude these users by dropping all users whose first observed post is within six months of the time when we access their profile.

C Estimation Procedure

We estimate the neural autopilot model using maximum likelihood. Users are indexed by $i = 1, 2, \ldots, N$. For user i, we observe her actions on $t = 1, 2, \ldots, T_i$, where $t = 1$ is the date when she first posts on Weibo, and T_i is the last date that her posting decision is observed in our data. $\boldsymbol{X}_i$ denotes a vector of user characteristics. For each user i and date t, we observe the user's posting decision $c_{i,t}$, which is treated as a binary choice: $c_{i,t} = 1$ refers to the case where user i chooses to post at least once on date t; and $c_{i,t} = 0$ refers to the case where user i chooses not to post on date t. Moreover, we observe the realized reward associated with $c_{i,t}$, which we denote by $u_{i,t}$. We measure $u_{i,t}$ by the number of likes that user i receives from her posts on date t. Taken together, the data set we use for model estimation is $\{(\boldsymbol{X}_i, c_{i,t}, u_{i,t}) \text{ for } t = 1, 2, \ldots, T_i\}_{i=1}^{N}$. Also recall from the main text that the set of model parameters that we estimate is $\boldsymbol{\theta} = (\lambda_r, \lambda_d, \kappa, \phi_0, \boldsymbol{\beta}, u_0^{before}, u_0^{during}, u_0^{after})$.

We maximize the following log-likelihood function

$$LL(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \left[\prod_{t=2}^{T_i} \left(\prod_{a \in \{0,1\}} \Pr(c_{i,t} = a | c_{i,t-1}, \boldsymbol{r}_{i,t}, \boldsymbol{d}_{i,t}, \boldsymbol{X}_i; \boldsymbol{\theta})^{\mathbb{1}\{c_{i,t}=a\}} \right) \right] \tag{C.1}$$

over $\boldsymbol{\theta}$, where $\boldsymbol{r}_{i,t} = (r_{i,t}(1), r_{i,t}(0))$ represents the predicted rewards for $a \in \{0,1\}$ for user i on date t, and similarly, $\boldsymbol{d}_{i,t} = (d_{i,t}(1), d_{i,t}(0))$ represents the reward predictability for $a \in \{0,1\}$. Note that we do not directly observe $\{(\boldsymbol{r}_{i,t}, \boldsymbol{d}_{i,t})$ for $t = 1, 2, \ldots, T_i\}_{i=1}^{N}$. Instead, we compute them using choice and rewards data $\{(c_{i,t}, u_{i,t})$ for $t = 1, 2, \ldots, T_i\}_{i=1}^{N}$, the estimated learning rates λ_r and λ_d, the estimated utility levels u_0^{before}, u_0^{during}, and u_0^{after}, and the update rules in Equations (1) and (2) of the main text.[26] When computing $\boldsymbol{r}_{i,t}$ and $\boldsymbol{d}_{i,t}$, we set the following initial values: for each user i, we set $r_{i,1}(1) = 0$; we calibrate $r_{i,1}(0)$ based on the observed second-date ($t = 2$) posting probability averaged across all users; and we set $d_{i,1}(1) = d_{i,1}(0) = 1$.[27]

At each date t, user i is either in the habit mode or in the goal-directed mode; over time, the user can switch between these two modes. We define a latent state variable $h_{i,t}$: $h_{i,t} = 1$ means user i is in the habit mode on date t; and $h_{i,t} = 0$ means user i is in the goal-directed mode on date t.[28] Then, according to the neural autopilot model described in Section 2 of the main text, we write the model-implied probability that user i posts on date t—namely, $\Pr(c_{i,t} = 1 | c_{i,t-1}, \boldsymbol{r}_{i,t}, \boldsymbol{d}_{i,t}, \boldsymbol{X}_i; \boldsymbol{\theta})$ in Equation (C.1)—as

$$\begin{aligned} &\Pr(c_{i,t} = 1 | c_{i,t-1}, \boldsymbol{r}_{i,t}, \boldsymbol{d}_{i,t}, \boldsymbol{X}_i; \boldsymbol{\theta}) \\ &= \begin{cases} \Pr(h_{i,t} = 1 | d_{i,t}(c_{i,t-1}), \boldsymbol{X_i}) & \\ + \Pr(c_{i,t} = 1 | \boldsymbol{r}_{i,t}, h_{i,t} = 0) \cdot \Pr(h_{i,t} = 0 | d_{i,t}(c_{i,t-1}), \boldsymbol{X}_i) & \text{if } c_{i,t-1} = 1 \\ \Pr(c_{i,t} = 1 | \boldsymbol{r}_{i,t}, h_{i,t} = 0) \cdot \Pr(h_{i,t} = 0 | d_{i,t}(c_{i,t-1}), \boldsymbol{X}_i) & \text{if } c_{i,t-1} = 0 \end{cases}. \end{aligned} \tag{C.2}$$

[26] One potential concern about our data set is that the number of likes $u_{i,t}$ observed by researchers may not be the same as the actual number of likes received by user i on date t. This happens if likes for a post published on date t trickle in slowly over the next few days. Although we are unable to rule out such a possibility, we note that on Twitter, 50% of all retweets happen within the first 10 minutes after the tweet has been published (see `https://tinyurl.com/2r4f8j22`). This suggests that most social media activities take place soon after the post is published.

[27] The users are likely to be in the goal-directed mode when they first start to post on Weibo. We make this assumption because when a user first posts on Weibo on date t, they are making a different choice than the choice on $t-1$, and therefore by definition is not in habit mode. As such, Equation (3) of the main text suggests that the model-implied second-date ($t = 2$) posting probability can be approximated by $\exp(r_{i,1}(1))/(\exp(r_{i,1}(1)) + \exp(r_{i,1}(0)))$. To solve for $r_{i,1}(0)$, we equate this model-implied probability to the observed second-date posting probability averaged across all users.

[28] Note that the econometricians do not directly observe $h_{i,t}$.

In Equation (C.2), the probability that user i is in the goal-directed mode on date t, $\Pr(h_{i,t} = 0|d_{i,t}(c_{i,t-1}), \boldsymbol{X_i})$, is given by Equation (4) of the main text

$$\Pr(h_{i,t} = 0|d_{i,t}(c_{i,t-1}), \boldsymbol{X}_i) = \frac{1}{1 + \exp(-\kappa \cdot (d_{i,t}(c_{i,t-1}) - (\phi_0 + \boldsymbol{\beta}' \boldsymbol{X}_i)))}. \tag{C.3}$$

The probability that the user is in the habit mode on date t is: $\Pr(h_{i,t} = 1|d_{i,t}(c_{i,t-1}), \boldsymbol{X_i}) = 1 - \Pr(h_{i,t} = 0|d_{i,t}(c_{i,t-1}), \boldsymbol{X}_i)$. Lastly, the probability that i chooses to post on date t, conditional on being in the goal-directed mode, is given by Equation (3) of the main text

$$\Pr(c_{i,t} = 1|\boldsymbol{r}_{i,t}, h_{i,t} = 0) = \frac{\exp(r_{i,t}(1))}{\exp(r_{i,t}(1)) + \exp(r_{i,t}(0))}, \tag{C.4}$$

with α set to 1.[29] In summary, we estimate

$$\boldsymbol{\theta} = (\lambda_r, \lambda_d, \kappa, \phi_0, \boldsymbol{\beta}, u_0^{before}, u_0^{during}, u_0^{after})$$

for users observed in the sample using the maximum likelihood approach described in Equations (C.1) to (C.4). To ensure that the parameter values are not sensitive to the estimation procedure, we have also tried the Metropolis-Hastings algorithm with uninformative priors as well as the maximum likelihood estimation with multiple initial values. All three procedures converge to similar parameter values.

In Section 4.2 of the main text, we compare our baseline model with an alternative "autopilot + lagged choice" model. This model generalizes the neural autopilot model by allowing the users' choice probabilities in the goal-directed mode to depend directly on their lagged choice. Specifically, we replace Equation (C.4) with

$$\Pr(c_{i,t} = 1|\boldsymbol{r}_{i,t}, c_{i,t-1}, h_{i,t} = 0) = \frac{\exp(r_{i,t}(1) + \gamma \cdot c_{i,t-1})}{\exp(r_{i,t}(1) + \gamma \cdot c_{i,t-1}) + \exp(r_{i,t}(0))}, \tag{C.5}$$

where the coefficient γ on the lagged choice $c_{i,t-1}$ is an additional parameter we estimate.

[29] One microfoundation of Equation (C.4) is as follows. Suppose a user actively compares the reward predictions of two alternative options, $(r_{i,t}(1) + \varepsilon_{i,t}(1), r_{i,t}(0) + \varepsilon_{i,t}(0))$, and chooses the larger one. Further suppose $(\varepsilon_{i,t}(1), \varepsilon_{i,t}(0))$ are idiosyncratic shocks that follow the type I extreme value distribution. Then, the probability that the user chooses $a = 1$ ("posting online") is given by Equation (C.4).

D Additional Tables and Figures

Table D.1: Posting Behavior Regressed on Reward Variables and Previous Posting History

	(1)	(2)
log(TotalLikes)	0.367***	
	(0.024)	
log(TotalLikes) × log(NPrevConsecutivePosts)	–0.132***	
	(0.012)	
log(TotalLikesPerDayPosting)		0.351***
		(0.099)
log(TotalLikesPerDayPosting) × log(NPrevConsecutivePosts)		–0.112*
		(0.046)
log(NPrevConsecutivePosts)	1.289***	0.893***
	(0.044)	(0.031)
PostToday	0.984***	1.060***
	(0.025)	(0.026)
Observations	1,802,377	1,802,377
R-squared	0.368	0.362
RMSE	0.24	0.24

Notes. Column 1 reports the logit estimates of the following regression:

$$\begin{aligned}\text{PostTomorrow}_{i,t} = \alpha_i + \delta_t + \beta_1 \log(\text{TotalLikes}_{i,t}) + \beta_2 \log(\text{NPrevConsecutivePosts}_{i,t}) \\ + \beta_3 \log(\text{TotalLikes}_{i,t}) \times \log(\text{NPrevConsecutivePosts}_{i,t}) + \beta_4 \text{PostToday}_{i,t} + \varepsilon_{i,t}.\end{aligned}$$

The dependent variable $\text{PostTomorrow}_{i,t}$ is an indicator that equals one if user i posts on date $t+1$ and equals zero otherwise. The independent variable $\text{TotalLikes}_{i,t}$ is the cumulative number of likes user i has received up until and including date t; the independent variable $\text{NPrevConsecutivePosts}_{i,t}$ is the number of days that user i has posted consecutively before date t. For example, if user i has posted on dates $t-2$, $t-1$, and t (but not on date $t-3$), then $\text{NPrevConsecutivePosts}_{i,t} = 2$. Column 2 reports the logit estimates of a regression that is similar to the one in Column 1; instead of $\text{TotalLikes}_{i,t}$, this regression uses $\text{TotalLikesPerDayPosting}_{i,t}$, which is the cumulative number of likes divided by the total number of days when the user has created at least one post, up until and including date t. Standard errors are clustered by individuals and reported in parentheses. *, **, and *** indicate significance at the 5%, 1%, and 0.1% level, respectively.

Table D.2: Autopilot Estimates: Comparing Average Likes and Total Likes Per Day

	Average Likes		Total Likes	
Parameters	Est.	Std. Err.	Est.	Std. Err.
λ_r	0.008	(0.000)	0.009	(0.000)
λ_d	0.126	(0.001)	0.121	(0.001)
κ	7.599	(0.034)	7.291	(0.033)
ϕ_0	0.315	(0.002)	0.327	(0.001)
β	0.071	(0.004)	0.070	(0.002)
u_0^{before}	1.242	(0.011)	1.072	(0.005)
u_0^{during}	1.095	(0.013)	1.216	(0.001)
u_0^{after}	1.078	(0.011)	1.054	(0.004)
Log-likelihood	–380,473		–389,048	
Number of users	1,848		1,848	
AIC	760,962		778,111	
BIC	761,061		778,211	

Notes: This table reports the parameter estimates for the autopilot model. Columns 1 and 2 use the average likes per post per day as the measure for the realized reward; Columns 3 and 4 follow the main text by using the total likes per day as the measure for the realized reward.

Table D.3: Autopilot Estimates: Users with Different Start Dates

	Cohort 1		Cohort 2		Cohort 3	
Parameters	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
λ_r	0.007	(0.000)	0.008	(0.000)	0.010	(0.000)
λ_d	0.113	(0.001)	0.120	(0.001)	0.126	(0.001)
κ	7.371	(0.029)	7.247	(0.027)	6.902	(0.037)
ϕ_0	0.318	(0.034)	0.329	(0.001)	0.351	(0.002)
β	0.073	(0.023)	0.069	(0.001)	0.071	(0.003)
u_0^{before}	1.215	(0.003)	1.222	(0.001)	1.176	(0.001)
u_0^{during}	1.087	(0.008)	1.078	(0.000)	1.010	(0.004)
u_0^{after}	1.082	(0.018)	1.062	(0.001)	0.983	(0.005)
Log-likelihood	–594,110		–413,222		–257,967	
Number of users	2,293		1,906		1,462	
AIC	1,188,236		826,460		515,950	
BIC	1,188,339		826,560		516,045	
Start date	January 1, 2014		January 1, 2015		January 1, 2016	

Notes. Our analysis in the main text uses a sample of users who created a Weibo account prior to January 23, 2020 and first posted on Weibo between February 16, 2015 and January 1, 2020. We selected the start date of February 16, 2015 as it corresponds to day 2,000 since August 27, 2009, the first observed date in our data set. This table reports the parameter estimates with alternative start dates: Cohorts 1, 2, and 3 include users who first posted on Weibo after January 1, 2014, January 1, 2015, and January 1, 2016, respectively.

Table D.4: Autopilot Estimates: Allowing for Additional Individual Heterogeneities

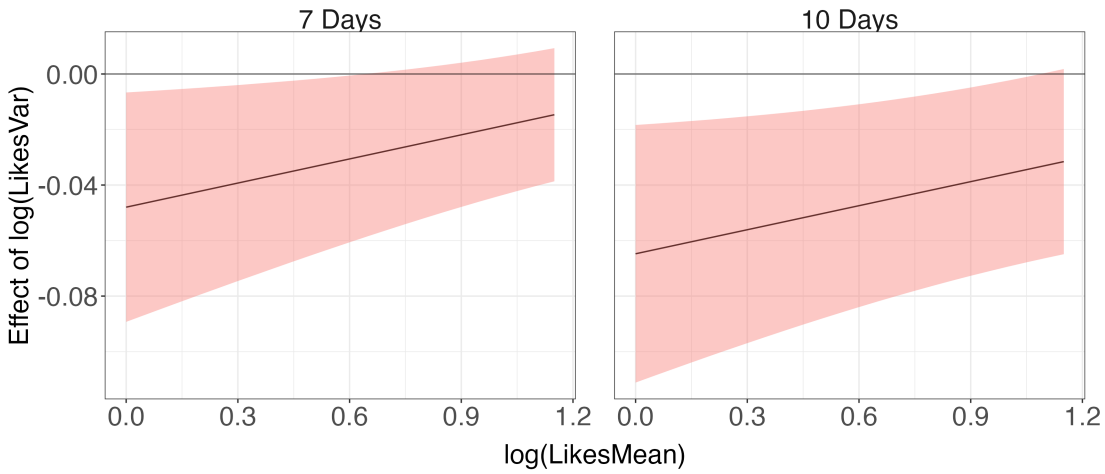
	Autopilot	
Parameters	Est.	Std. Err.
λ_r	0.008	(0.000)
λ_d	0.120	(0.001)
κ	7.213	(0.040)
ϕ_0	0.482	(0.006)
β^{male}	0.030	(0.003)
$\beta^{verified}$	0.020	(0.007)
$\beta^{developed}$	0.001	(0.006)
β^{highCS}	–0.168	(0.008)
β^{orig}	0.071	(0.010)
u_0^{before}	1.210	(0.032)
u_0^{during}	1.060	(0.029)
u_0^{after}	1.044	(0.009)
Log-likelihood	–387,656	
Number of users	1,848	
AIC	775,337	
BIC	775,486	

Note: This table reports the parameter estimates for the neural autopilot model when we allow additional individual heterogeneities to drive ϕ, the threshold parameter in Equation (4) of the main text. Specifically, ϕ for each user is affected not only by the user's proportion of original posts, but also by additional attributes that include the user's gender, whether the user has a verified account, is in a developed city, and has high Sesame Credit. We assume

$$\begin{aligned}\phi_i = \phi_0 \quad &+ \quad \beta^{male}\mathbb{1}_{male_i=1} + \beta^{verified}\mathbb{1}_{verified_i=1} + \beta^{developed}\mathbb{1}_{developed_i=1} \\ &+ \quad \beta^{highCS}\mathbb{1}_{highCS_i=1} + \beta^{orig}\text{OrigPostRatio}_i,\end{aligned}$$

where OrigPostRatio_i is the proportion of original posts for user i.

Figure D.1: Marginal Effect of log(LikesVar) on Habitual Postings



Notes. This figure plots the marginal effect of log(LikesVar) on habitual postings (NConsecutivePosts) as a function of log(LikesMean). The left panel computes "LikesMean" and "LikesVar" as the mean and variance of the number of likes received per day by the user over the last 7 posting days. The right panel computes "LikesMean" and "LikesVar" as the mean and variance of the number of likes received per day by the user over the last 10 posting days. The shaded regions represent the 95% confidence intervals. The range of log(LikesMean) is from its 10th percentile value to its 90th percentile value.