Economic agents as imperfect problem solvers*

Cosmin Ilut
Duke University & NBER

Rosen Valchev
Boston College

February 2021

Abstract

We develop a tractable model of limited cognitive perception of the optimal *policy function*, with agents using costly reasoning effort to update beliefs about this optimal mapping of economic states into actions. A key result is that agents reason less (more) when observing usual (unusual) states, producing state- and history-dependent behavior. Our application is a standard incomplete markets model with ex-ante identical agents that hold no a-priori behavioral biases. The resulting ergodic distribution of actions and beliefs is characterized by "learning traps", where locally stable dynamics of wealth generate "familiar" regions of the state space within which behavior appears to follow past-experience-based heuristics. We show qualitatively and quantitatively how these traps have empirically desirable properties: the marginal propensity to consume is higher, hand-to-mouth status is more frequent and persistent, and there is more wealth inequality than in the standard model.

JEL Codes: D83, D91, E21, E71, C11

*Email addresses: Ilut cosmin.ilut@duke.edu, Valchev valchev@bc.edu. We would like to thank our discussants George-Marios Angeletos, Paolo Bonomolo, Tarek Hassan and Luigi Iovino, as well as Ian Dew-Becker, Ryan Chahrour, Guido Lorenzoni, Filip Matějka, Kristoffer Nimark, Philipp Sadowski, Todd Sarver, Andrei Shleifer and Mirko Wiederholt, and seminar and conference participants at Bank of Finland, CERGE, CEU, EEA Congress, Green Line Macro Meetings, NBER EFBEM and EFCE Groups, Northwestern Macro Conference, Society for Economic Dynamics and Computing in Economics and Finance for helpful discussions and comments. This project is supported by NSF Award SES-1824367.

1 Introduction

Standard models assume that, given beliefs about the state of the world, decision-makers face no cognitive limitations in figuring out their optimal course of action. In other words, agents are always fully aware of the answer to a critical question: 'What is the action that yields the highest value, given my current circumstances?'. Since in most economic environments individual circumstances vary over time, this view of human cognition essentially assumes that people can costlessly solve for their optimal policy *functions* – the state-contingent mapping of beliefs about states, economic structure and etc., into the best course of action.

In reality, however, coming up with a "good" plan of action is the outcome of conscious reasoning and deliberation, which requires cognitive effort. Motivated by a long-standing interest in modeling cognitive resources as scarce (e.g. Simon (1955, 1956)), in this paper we develop a model where figuring out the optimal action requires costly reasoning effort.[1]

We focus on two major questions. The first is how to model costly reasoning about the unknown optimal policy function with a general, yet tractable framework that is applicable across different economic environments, while capturing key empirical insights on how human reasoning works. Specifically, we aim to model the reasoning process as (i) noisy, (ii) resource-rational and (iii) subject to 'episodic memory', or the differential recall of past information based on its similarity with the current situation, all of which are well-documented features by the neuroscience and psychology literature. We achieve these goals by casting the agents' reasoning as an 'as if' Bayesian non-parametric estimation of the unknown optimal policy function, subject to a trade-off between accuracy and mental cost. The second question is to establish whether and how the resulting bounded-rationality mechanism matters for micro and macro behavior. To do so, we use a standard Aiyagari (1994) incomplete markets model as a laboratory to show that, even though reasoning errors are i.i.d. and agents have no a-priori behavioral biases, costly reasoning *endogenously* alters ergodic behavior in a systematic and quantitatively promising way.

To fix ideas and to preview our application, consider a consumption-savings problem where agents choose their level of consumption and all payoff relevant information is contained in the current value of the sufficient economic state – the available cash-on-hand y. Our agents are imperfect problem solvers in the sense that they do not have immediate access to the optimal solution to this decision-making problem – i.e. the optimal consumption policy function $c^*(y)$. Therefore, agents face uncertainty about the optimal consumption level in any given period t, $c_t \equiv c^*(y_t)$, even though their current circumstances (encoded in y_t) are

[1] See Klaes et al. (2005) for a conceptual history of bounded rationality and Conlisk (1996) for an early review, among many others (eg. Rubinstein (1998), Todd and Gigerenzer (2003), or DellaVigna (2009)), on the evidence and challenges in incorporating human cognition as a scarce resource in economics.

perfectly known.[2] This uncertainty is subjective and due to the fact that solving for the true optimal rule c^* is cognitively costly, hence agents might not have a perfect grasp of it.

Intuitively, at each point in time, an agent sees her current financial resources y_t (and its history) and asks herself the question: 'What is my best course of action given current circumstances?' Imperfect reasoning leads her to face uncertainty over that answer and act according to an imprecise guess over the optimal level of consumption. In each period, the agent can also engage in costly internal deliberation, through which she gets a new (noisy) idea, or mental information, about the unknown level of the optimal action. She is not sure that this is the right answer to the question she posed herself, but understands that the more costly reasoning effort she invests the more precise is her updated guess.[3] Moreover, she does not simply act on the new information she just got, but reflects on it in combination with any knowledge she has accumulated about optimal consumption behavior through reasoning about the same question in the past, albeit in the context of different past states of the world y_{t-k}. In particular, she figures that her past reasoning ideas are likely to be more relevant to her now if they occurred at past circumstances y_{t-k} that are more similar to her current y_t.[4]

Methodological contribution. Our key methodological contribution is to model this reasoning process in a general, yet tractable way. To do so, we represent the unknown optimal policy function as a projection on a set of kernel basis functions, and assume that the basis functions and the history of state values are known to the agent. Hence uncertainty over the optimal action is stemming from uncertainty over the projection coefficients that effectively define the optimal policy c^*. We then follow insights from machine learning and Bayesian statistics and assume Gaussian priors over the projection coefficients, allowing us to conveniently integrate them out and work directly with a closed-form expression for the implied distribution over the space of functions c^*. Thus, effectively we model the reasoning process as an abstract 'as if' representation in terms of a Bayesian non-parametric functional

[2]Our interest is thus complementary, but distinct, to a literature inspired by Sims (1998, 2003) that models imperfect perceptions of the relevant state of the world arising from agents' limited attention capacity. A common feature within this approach is that, *conditional* on beliefs about the state, the optimal policy *function* mapping those beliefs into the best perceived action is derived under no additional cognitive cost.

[3]First, the noisy perception implied by these reasoning signals means that our agents exhibit stochastic choice, i.e. even conditioning on the same observed circumstance their actions may differ. See Tversky (1969) and more recently Ballinger and Wilcox (1997) and Hey (2001) for experimental evidence of stochastic choice. Second, the notion that agents are "resource-rational" and exhibit behavior that is the outcome of limited but appropriate deliberation, has been advocated as early as Simon (1976).

[4]This episodic nature of memory has been advocated as a critical psychological feature as early as Tulving (1972), with the accumulation of evidence reviewed in the surveys by Tulving (2002) and Gershman and Daw (2017). For example, Plonsky et al. (2015) and Bornstein et al. (2017) document through experiments and neuroimaging how the current decision is influenced by a past choice whose weight depends not simply on how recent it is (i.e. a standard temporal difference) but, critically, how this weight is larger if that past choice occurred in a similar context or circumstance to the current one (i.e. a 'circumstance-based' difference).

estimation, which tractably captures the three key features of reasoning we identified earlier.

Specifically, in each period t, reasoning generates a noisy and unbiased signal, η_t, about the optimal action today $c^*(y_t)$. This signal updates beliefs about the whole function c^*, and the resulting conditional expectation of $c^*(y_t)$, in particular, is the agent's time-t action, as it is her best guess of the optimal action given today's state y_t. The agent is 'resource-rational' in her reasoning, as she optimally chooses the signal precision each period, by trading-off the reduction in uncertainty over the optimal action achieved by the new signal, and a cognitive cost proportional to the amount of information carried in that signal, as measured by entropy.

Finally, due to the non-parametric nature of the estimation, while a signal η_t updates beliefs about the function c^* over the whole state space, in their time-t update agents optimally put more weight on past knowledge that was derived in similar situations (i.e. at state values y_{t-k} close to the current state y_t), as opposed to simply 'recent' memories (i.e. y_{t-k} for small k). Thus, our agents display 'episodic' or 'associative' memory, with past information being weighted by a similarity notion, which is both an intuitive feature of memory and a well-documented empirical phenomenon (e.g. Plonsky et al. (2015)).

The key emerging property of our framework is the state- and history-dependence of subjective uncertainty. In particular, for realizations of the state y_t close to the past y_{t-k} where the agent has already reasoned, the beginning-of-period uncertainty over $c^*(y_t)$ is relatively low, since the agent has accumulated some useful information already. In such "familiar" parts of the state space, the agent optimally chooses to invest little additional cognitive effort, and thus, the resulting action is close to her previous-period beliefs about $c^*(y_t)$. Through this interaction of episodic memory and resource-rationality, the model endogenously generates a dual-type of reasoning (a well-documented psychological phenomenon, eg. Stanovich and West (2000)): *habitual* or *heuristic* behavior in familiar circumstances (i.e a system 1), with a change to a more deliberative approach in unfamiliar circumstances, in the form of significant new reasoning effort and thus revision in beliefs and actions (i.e. a system 2).

Applied contribution. To showcase how the proposed costly reasoning friction matters for observable behavior both at the micro and macro level, we analyze our mechanism in the setting of the Aiyagari (1994) incomplete markets model of consumption-savings decisions.

The resulting typical behavior is underpinned by a fundamental feedback between the state-dependent, local reduction in uncertainty and the endogenous dynamics of the state (i.e. wealth). Namely, when wealth drifts into new and uncertain parts of the state space, an agent's conditional beliefs about the optimal consumption function c^* are likely to change, as she increases reasoning efforts in response to the increase in uncertainty. On the contrary, if the accumulated past reasoning signals lead to a policy function estimate that establishes stable wealth dynamics, and thus a high likelihood for wealth to remain within a particular neighborhood, the evolution of beliefs slows down significantly. As wealth fluctuates in such a familiar, low-uncertainty region, the agent has little incentive to reason much further. However, without acquiring new precise signals her conditional beliefs largely stop evolving and remain the same, which in turn perpetuates the stable dynamics in wealth. There are two such patterns of mistakes that can be self-perpetuating, one that generally characterizes agents near the borrowing constraint and one that characterizes unconstrained agents.

We call such situations "learning traps", and they underscore a powerful *selection effect* of *when* agents choose to reason intensely, which leads to a selection in what kind of errors in the policy function estimates tend to persist and are thus over-represented at the stochastic steady-state of the model. Due to such selection, the resulting stochastic steady-state behavior is systematically different from that implied by the full-information (FI) c^*, in ways that are both empirically promising and important for aggregate dynamics.

First, our costly reasoning model rationalizes the related facts that (i) a surprisingly high fraction (23% in the data) of households have near-zero net wealth and that (ii) those households also remain in this effective "Hand-to-Mouth" (HtM) situation *persistently* (Aguiar et al. (2020)). This feature of the data is challenging for standard models where the precautionary motives of agents makes HtM status both very rare and temporary. Our mechanism is consistent with both features of the data because some of our agents settle in "learning traps" characterized by a noisy policy estimate that points to a high average *level* of consumption near the borrowing constraint, which generates a negative drift in assets and thus endogenously keeps wealth near the constraint persistently. Without variation in their circumstances, these agents perceive no further need to reason a lot, perpetuating their high consumption behavior. Such agents are naturally over-represented in the left-tail of the ergodic wealth distribution, hence the HtM agents in our model display a *habitually* high level of consumption and remain HtM persistently. Furthermore, we show that this behavior can also rationalize the puzzling lack of adequate savings of agents with temporarily available liquidity, who are otherwise near the constraint (Ganong and Noel (2019)).

Second, in line with the data, but in contrast to standard frameworks, our model also produces high MPCs even for rich, unconstrained agents.[5] While this consumption property is usually viewed as a separate theoretical challenge, in our model it arises from the same general mechanism of endogenous selection of beliefs. In particular, away from the liquidity constraint, wealth dynamics are endogenously mean-reverting precisely when the agent's consumption policy estimate is relatively steep. Intuitively, in such a case the agent tends to consume "too much" out of high realizations of income shocks, and "too little" otherwise,

[5] Parker (2017), Kueng (2018), Lewis et al. (2019), Olafsson and Pagel (2018), Fagereng et al. (2020) and McDowall (2020) document that rich people, even with high liquid wealth, have puzzlingly high MPCs.

keeping her resulting wealth stable around its average. Thus, the resulting "learning traps" make high MPC behavior the norm, even for the richest agents. For example, in our model the average MPC of the richest quintile of agents is equal to 0.15, which is in-line with the data, and four times higher than in the FI benchmark, where rich agents are essentially permanent-income-hypothesis consumers.

Finally, we discuss several dimensions through which our mechanism matters in the aggregate. First, both the large fraction of HtM agents and the high MPC for the wealthy, unconstrained agents, that our model generats contribute to an empirically relevant, high aggregate MPC. In particular, in our model the average MPC is 0.29, well in line with empirical estimates (Carroll et al. (2017)), while the mean MPC for the FI version is counterfactually low at 0.05. Second, while the beliefs of all agents tend towards a steep policy estimate and stable wealth dynamics, the differences in the specific history of reasoning errors across agents means that the eventual steady-state wealth levels around which beliefs and wealth stabilize is heterogeneous. This leads to significantly larger wealth heterogeneity – for example, our model displays a Gini coefficient of 0.58, 50% higher than the FI version. Third, we illustrate the policy importance of modeling bounded, but "resource-rational" agents, by considering a fiscal stimulus at a time when additional information arrives that lowers agents' confidence in their previous reasoning about c^*. We find that in this case the average MPC falls substantially on impact, as agents abandon their "business-as-usual" consumption patterns, with implications for policymakers that hope to leverage the usual high MPCs.

2 Related literature

We share the broad interest of modeling mistakes in decision-making with a long tradition of studying imperfect perception of the relevant state of the world. There are numerous such approaches, ranging from the Rational Inattention literature inspired by Sims (1998, 2003) where the attention choice is optimal, to work inspired by neuroscience evidence on the imperfect perception of stimuli (eg. Girshick et al. (2011), Wei and Stocker (2015) and Woodford (2019)), where the precision is exogenously given. Our paper is closest in spirit to the optimal inattention approach, as our reasoning choice is similarly resource-rational. However, the uncertain object agents learn about is different, with that approach maintaining the usual assumption that agents know the mapping of their (otherwise imperfect) beliefs about the unobserved state into their optimal actions (see Wiederholt (2010), Gabaix (2019) and Mackowiak et al. (2020) for surveys and Woodford (2003), Reis (2006), Maćkowiak and Wiederholt (2009), Gabaix (2014), Matějka and McKay (2014) or Stevens (2020) for specific modeling examples). In our approach we are specifically interested in relaxing that usual assumption of knowing c^* and modeling uncertainty over the optimal policy function.

In our modeling choices, we first build on a large literature in cognitive sciences (eg. Gershman et al. (2015), Griffiths et al. (2015) and Shenhav et al. (2017), surveyed by Lieder and Griffiths (2020)), that promotes the view of human reasoning and cognition as "resource-rational", trading off accuracy and cognitive cost in reaching the best perceived action for a given circumstance. Second, consistent with a large experimental evidence on stochastic (as opposed to deterministic) choices (see for example Mosteller and Nogee (1951), Tversky (1969), Ballinger and Wilcox (1997), Hey (2001)), we are interested in modeling noisy perceptions of the optimal action, and as such we also related to approaches producing stochastic choice based on bounded rationality (eg. Ratcliff and McKoon (2008), Manzini and Mariotti (2014) and Woodford (2014), surveyed in Johnson and Ratcliff (2014)). Third, we connect to extensive evidence in the experimental, psychology and neuroscience literature (see Tulving (2002) and Gershman and Daw (2017) for surveys) documenting the episodic, or similarity-based, properties of memory and decision-making. This evidence ranges from early real-world situations (like chess players in Chase and Simon (1973), or fire-fighters in Klein et al. (1986)) to recent controlled experiments and neuroimaging (eg. Plonsky et al. (2015), Bornstein et al. (2017) and Bornstein and Norman (2017)).

In this context, our paper connects to existing modeling approaches in economics that have been built around this similarity-based property of decision-making, including the early model of case-based decision theory (Gilboa and Schmeidler (1995, 2001)) and the recent memory-based anchoring model (Bordalo et al. (2020)). In this view, shared with the classic computational literature on reinforcement learning (eg. Kaelbling et al. (1996)), the information that guides agents' current choices typically takes the form of the experienced outcomes of past actions and circumstances, making the precision of each signal naturally fixed, since agents learn only from actual experiences. Instead, in our setup new information arrives as the outcome of internal deliberation and reasoning, a view shared with a theory literature exemplified by Aragones et al. (2005) or Alaoui and Penta (2016). The precision of this type of reasoning information is then naturally under the agent's control, and can thus vary, depending on its cost-benefit tradeoff, conditional on a given circumstance. Conceptually, this allows us to study the *interaction* between similarity-based learning and the constrained-optimal choice of the precision of new information over the optimal action.

The joint study of episodic memory and resource-rationality also connects our model to a large literature and body of evidence in psychology that emphasizes reasoning as a dual-process (see Stanovich and West (2000) and Evans (2003) for reviews of the evidence and related theories, and Sloman (1996) and Kahneman (2011) for specific interpretations). That view describes reasoning as the interplay between a 'System 1', which is an intuitive, associative, habitual mode of decision-making, and a 'System 2', which activates deliberative and cognitively-demanding reasoning. Our framework's key qualitative result is consistent with this view in that associative/habitual actions are taken in familiar circumstances, where little (if at all) additional costly reasoning is engaged (thus acting like relying on System 1), but agents actively deliberate (like in System 2) when faced with un-familiar circumstances.

Our incomplete markets application relates to a large literature aimed at addressing two well-documented and fundamental challenges for canonical models. First is the puzzling lack of saving among poor households, which is essentially a statement about the *level* of consumption. The typical approach in the literature is to appeal to permanent heterogeneity in preferences (usually in discount rates) to generate a large mass of 'spender' agents with high average consumption, and thus low steady-state wealth and frequent HtM status (e.g. Krueger et al. (2016), Carroll et al. (2017), and Aguiar et al. (2020)). The second is to understand why rich and wealthy households have surprisingly high MPCs and do not smooth out transitory income shocks, as suggested by the permanent-income-hypothesis (a statement about the *slope* of the consumption function).[6] Here, liquidity frictions (as in Kaplan and Violante (2014)) are typically used to account for the high MPC of rich and liquidity-constrained agents, but typically fall short in explaining the evidence for the liquid wealthy agents (as argued for example by Kueng (2018), Olafsson and Pagel (2018), Fagereng et al. (2020) and McDowall (2020)). This shortcoming has spurred recent interest in behavioral models that can lead to high MPCs even for such rich and liquid agents (e.g. Lian (2020)).

Our paper proposes a complementary mechanism that can *jointly* and parsimoniously speak to these challenges by relaxing the common assumption that agents costlessly optimize over their actions, which results in persistent differences in the perception of the optimal policy function and hence behavior consistent with the puzzles. In the process, the same mechanism also delivers significantly larger wealth heterogeneity, in a way that is consistent with the empirical evidence that emphasizes the role of apparent unobserved heterogeneity in behavior driving wealth dispersion (e.g. Bernheim et al. (2001), Ameriks et al. (2003), or Hendricks (2007), and see De Nardi and Fella (2017) for a broader survey). Overall, understanding these puzzles is central for the macro analysis of incomplete market models, as they drive the large average MPC that underpins the key propagation mechanism.

Section 3 describes the general framework of costly reasoning. Section 4 introduces the reasoning friction into the Aiyagari (1994) environment and analyzes its qualitative insights. A numerical analysis of the model is discussed in Section 5.

[6] This challenge is further exacerbated in the class of models that aim to explain HtM behavior through discount-rate heterogeneity (eg. Carroll et al. (2017), Aguiar et al. (2020)), since there the 'saver' (high β) types become the typical wealthy agents, but their stronger incentive to smooth consumption also leads, everything else constant, to lower MPCs in a non-HtM circumstance, i.e. rich and highly liquid.

3 General framework

We consider a generic economic situation where an agent chooses an action, denoted as c_t, so as to maximize her expected payoff. The agent is perfectly aware of all payoff-relevant details of the economic environment, such as her utility function $u(c)$, the set of constraints that restrict the feasible action c_t, and the process and current realization y_t of the state variable relevant for that decision. In other words, the agent knows the full mathematical description of the maximization problem she faces. Naturally, the solution – the policy function $c^*(y)$ which gives the optimal action for any realization of the state variable y – follows logically from the details of the agent's maximization problem, hence, it might seem tempting to conclude that the agent must also know $c^*(y)$.

In reality, however, people are subject to cognitive limitations that prevent this. Knowing a fact does not mean that all of its logical implications are also immediately obvious to us – deducing such further implications takes time and reasoning effort.[7] As economists we encounter this every time we write down a new model – even though we know all basic facts about the structure of the problem we ourselves created, the solution and full set of implications takes time and effort to figure out. Beyond introspection, there is a large and growing experimental and field evidence that people indeed need time and effort to "solve" tasks, and that the quality of decision making is negatively affected by the complexity of the decision problem (see for example Caplin et al. (2011), Kalaycı and Serra-Garcia (2016) and Carvalho and Silverman (2019) and Stanovich and West (2000) and Deck and Jahedi (2015) for a recent survey).

To represent the notion that agents perfectly know the objective description of their environment, but do not immediately know the exact solution to the maximization problem, we build on general insights from prior decision-theory work (Lipman (1991, 1999)) and model agents 'as if' facing subjective uncertainty about the optimal policy function $c^*(y)$. Furthermore, we ground the particular modeling details that operationalize this broad idea of reasoning as the reduction in subjective uncertainty by drawing on empirical insights on how human reasoning works from psychology and neuroscience, as detailed below. Overall, we thus view our proposed framework as an 'as if' mathematical representation of deliberation that provides a tractable and disciplined way of formalizing key empirical features of reasoning.

To simplify the exposition, we restrict attention to an economic environment described by a scalar state $y \in \mathbb{R}$ and where the action space is also the real line, hence $c^*(y) : \mathbb{R} \to \mathbb{R}$.

[7] In the sense of Hintikka (1975), people thus lack 'logical omniscience'. For example, knowing all axioms of a given subfield of mathematics does not immediately imply that one also knows all theorems that can be proved with those axioms. This is despite the fact that the theorems follow logically from the axioms – i.e. there is no "new" or additional information in the theorems themselves.

The framework readily generalizes to multivariate settings as discussed in the Appendix.

3.1 Uncertainty about the optimal action

To formalize the notion of uncertainty over the optimal policy function $c^*(y)$, we start by the general representation of c^* as the sum of a (complete) set of basis functions,[8]

$$c^*(y) = \lim_{N\to\infty} \sum_{j=1}^{N} \theta_j \phi_j(y), \tag{1}$$

where the set of basis functions $\{\phi_j\}$ is known, but the projection coefficients θ_j are *unknown*. Hence, when we state that the agent is uncertain about the optimal action at time t, i.e. $c^*(y_t) = \sum_j \theta_j \phi_j(y_t)$, that uncertainty is entirely due to the fact that the coefficients θ_j's are unknown, since both the set of ϕ_j and the state realization y_t are known.

We assume that the agent's priors over θ_j are Gaussian and independent of one another, so $\theta_j \sim N(\mu_j, \sigma_j^2)$. This helps make the analysis particularly tractable, as it implies that the resulting distribution over the space of functions c^* is a Gaussian Process (GP) distribution. Thanks to this result, we can directly work with a closed-form expression for beliefs over c^* itself, rather than having to keep track of beliefs about an infinite collection of θ_j coefficients.[9]

Lemma 1. *If θ_j follow independent Gaussian distributions $N(\mu_j, \sigma_j^2)$, then c^* has a Gaussian Process distribution, denoted as $c^* \sim \mathcal{GP}(\widehat{c}_0, \widehat{\sigma}_0)$, and meaning that for any pair of state values y, y' the joint distribution of the resulting function values is given by:*

$$\begin{bmatrix} c^*(y) \\ c^*(y') \end{bmatrix} \sim N\left(\begin{bmatrix} \widehat{c}_0(y) \\ \widehat{c}_0(y') \end{bmatrix}, \begin{bmatrix} \widehat{\sigma}_0(y,y) & \widehat{\sigma}_0(y,y') \\ \widehat{\sigma}_0(y,y') & \widehat{\sigma}(y',y') \end{bmatrix} \right)$$

where

$$\widehat{c}_0(y) = \sum_{j=1}^{\infty} \mu_j \phi_j(y); \qquad \widehat{\sigma}_0(y,y') = \sum_{j=1}^{\infty} \sigma_j^2 \phi_j(y)\phi_j(y')$$

Proof. Details are in Appendix A. □

The defining feature of the Gaussian Process distribution is that the joint distribution of the values of the unknown function at any two inputs y and y' is Normal, with mean and

[8] A set of basis function is complete when they span the space of functions which we seek to represent, and hence the infinite sum of basis functions exactly reproduces $c^*(y)$.

[9] A GP distribution is the limit of the multivariate Gaussian distribution for infinite vectors, and it is often used as a prior for Bayesian inference on functions (Liu et al. (2011)). In economics, it has recently been used for non-parametric learning about the distribution of the state (but still assuming knowledge of mapping to optimal actions) – e.g. Dew-Becker and Nathanson (2019), Ilut et al. (2020) and Kozlowski et al. (2020).

variance fully characterized by the two functions $\widehat{c}_0(y) : \mathbb{R} \to \mathbb{R}$ and $\widehat{\sigma}_0 : \mathbb{R}^2 \to \mathbb{R}$, which are known in the Bayesian statistics literature as the "mean" and "covariance" functions. These two functions depend on both the choice of the parameters of the priors for θ_j (i.e. μ_j and σ_j^2) and the choice of basis functions ϕ_j. The latter is important, as it encodes the agent's prior beliefs about the likely functional shapes of c^* – e.g. if we assume that ϕ_j are a set of linear functions, then we impose the ex-ante assumption that the agent believes c^* to be linear.[10]

Naturally, the "mean function" $\widehat{c}_0(y)$ specifies the prior mean of $c^*(y)$ for any y. Note that any difference between the unknown, to the agent, optimal policy function c^* and the prior mean $\widehat{c}_0$ represents an ex-ante bias in beliefs. Our motivating properties of reasoning do not impose structure on this bias and so it represents a degree of freedom for us as analysts.

To eliminate the role of this degree of freedom we center the prior beliefs of the agents over the true unknown function c^*:

$$\widehat{c}_0(y) = \mathbb{E}(c^*(y)) = c^*(y).$$

Essentially, this means that the agent's prior beliefs about the θ_j's are centered around the true projection coefficients, and as a result the agent holds *no ex-ante bias* in her beliefs.[11]

Nevertheless, the agent of course still faces *uncertainty* around her mean prior belief $\widehat{c}_0(y)$, and that is encoded by the covariance function $\widehat{\sigma}_0(y, y')$, which specifies the covariance between the values of the function c^* at any pair of inputs y and y':

$$\widehat{\sigma}_0(y, y') = \mathbb{E}\left((c^*(y) - \widehat{c}_0(y))(c^*(y') - \widehat{c}_0(y'))\right).$$

As seen directly from Lemma 1, the level of uncertainty, in the sense of the prior variance over the value of the function evaluated at any given point y, i.e.

$$\widehat{\sigma}_0^2(y) \equiv \mathrm{Var}(c^*(y)) = \sigma_0(y, y) = \sum_{j=1}^{\infty} \sigma_j^2 \phi_j(y)^2,$$

is increasing in the variance of the θ_j priors – the series $\{\sigma_j^2\}$. Naturally, the bigger is the variance over the unknown coefficients θ_j, the larger is the ball of uncertainty around the mean prior belief $\widehat{c}_0(y)$, and hence, those parameters govern the overall level of prior uncertainty.

In addition, the covariance function also specifies how the agent believes that the values of c^* evaluated at two different points y and y' are likely to correlate with each other.

[10] Note that some care needs to be exercised when picking $\{\mu_j, \sigma_j^2\}$ and $\{\phi_j\}$ so that the resulting GP distribution has well-defined, finite moments. We describe our particular choices further below.

[11] While our framework allows for general prior shapes of $\widehat{c}_0(y)$, the no ex-ante bias modeling choice squarely identifies the role of costly reasoning to *endogenously* drive a different ergodic behavior than implied by $c^*(y)$.

Intuitively, this speaks to the prior belief about the types of functional shapes that c^* is likely to take (e.g. linear, polynomial, etc.) and this is controlled by the choice of basis functions ϕ_j. For example, if the set $\{\phi_j\}$ includes only linear functions, then the beliefs about $c^*(y)$ and $c^*(y')$ are perfectly correlated across the state space because a shift in beliefs about the value of the function at some point y directly implies a shift in beliefs about $c^*(y')$ at all other y'. Formally, in this case a change in the estimate of one θ_j implies a change in the estimate of the slope of c^* everywhere, and hence changes the estimated level of $c^*(y)$ for *all* y.

Thus, the choice of $\{\phi_j\}$ effectively governs the way the agent extrapolates information about the value of the function c^* at a point y to its value at another point y'. To discipline this choice, we are guided by a growing body of evidence based on both experiments and neuroimaging, which shows that at a basic neurological level the brain learns to solve problems and make decisions by drawing on specific individual, or 'episodic', memories that are relevant to the current task (see Plonsky et al. (2015), Bornstein et al. (2017), Bornstein and Norman (2017) and the survey in Gershman and Daw (2017)). A similar conclusion of primarily drawing on "similar" memories during the decision-making process also emerges from studies focused on traditional field evidence (like chess players in Chase and Simon (1973), fire-fighters in Klein et al. (1986); see Tulving (2002) for a survey on this psychology literature).

To operationalize the idea that such 'episodic' or 'associative' memory plays a crucial role in the learning process, we assume that the basis functions the agent uses are a collection of Gaussian kernels centered uniformly over the whole real line, i.e. we set

$$\phi_j = \exp(-\psi(y - r_j)^2),$$

where r_j is the "node" around which the j-th basis function is centered. Substituting these kernels into equation (1), the learning problem is essentially equivalent to the statistical problem of a Bayesian, non-parametric kernel regression. The non-parametric nature of the estimation is crucial to capturing the notion of episodic memory, as inference in non-parametric methods is naturally localized and the value of each piece of information is specific to its location in the state space. We provide formal derivations in Section 3.3, after we define the nature of the "data" or reasoning information the agent has access to.

Importantly, this choice of basis functions also makes our learning framework both *general* and *tractable*.[12] The non-parametric nature of the estimation does not impose any ad-hoc functional form assumptions on the underlying c^* (e.g. linearity), and moreover, the set of Gaussian kernels is very flexible and is able to approximate arbitrarily well any

[12] We thus relate to increased interest in the cognitive and neuroscience literature (eg. Lucas et al. (2015) and Gershman and Daw (2017)), in using Bayesian non-parametric, kernel regressions as a general way to model learning over potentially complex functions that is also characterized by similarity-based inference.

continuous function (i.e. it satisfies the Universal Approximation Theorem). Despite this generality, with the additional assumption that σ_j^2 are equal for all j, the setup is quite tractable, as the infinite sum defining the covariance function reduces to a simple expression.

Corollary 1. *Assuming that for every N, $\{\phi_j\}_{k=1}^N$ is a set of Gaussian kernels, with the same precision ψ and with means r_j that are uniformly distributed in the interval $[-N, N]$, and that θ_j follow the independent Gaussian distributions $N(\mu_j, \frac{\sigma_c^2}{N})$, then in the limit $N \to \infty$*

$$\widehat{\sigma}_0(y, y') = \sigma_c^2 \exp(-\psi(y - y')^2).$$

Essentially, the combination of independent Gaussian priors over θ_j that have the same variance, but potentially different means, and Gaussian kernels as the basis functions $\{\phi_j\}$, is a convenient pair of "conjugate priors", which lead to a closed form, and tractable distribution over c^* itself.[13] Specifically, this distribution is characterized by two parameters. First, σ_c^2 controls the prior uncertainty about the value of $c^*(y)$ at any given point y, capturing the "quantity" of uncertainty agents face around $\widehat{c}_0(y)$. Second, ψ controls the extent to which information about the value of the function at a point y is informative about its value at a different point y'. Formally, ψ controls the bandwidth of the underlying kernels, with a higher ψ resulting in a lower bandwidth and thus information being more localized. Intuitively, a higher ψ means that agents are unwilling to extrapolate far, hence information about the value of the function at a particular point y, i.e. $c^*(y)$, is less informative about the value of the function at other $y' \neq y$ as the distance between y' and y increases.

Overall, the proposed framework puts only very weak restrictions on the agents' prior beliefs over the unknown policy function, and it is thus portable and applicable to a wide range of economic problems. By avoiding the need to select specific functional forms for the prior, which can easily differ from the actual optimal policy function in a specific economic application, learning is thus not a-priori misspecified.[14] In that way, our agents are not systematically "fooled" ex-ante, and their priors are consistent with the underlying true c^*.

A second sense in which our proposed framework is general is more subtle, but important in applications. Because all payoff relevant variables and constraints are observed, the model allows us to easily study applications with meaningful constraints on the set of relevant

[13] Our setup is closely related to a long literature in Bayesian statistics on non-parametric functional estimation, known as Gaussian Process Regressions or kriging, which uses the same priors as us, because of their attractive combination of flexibility, tractability and degrees of freedom (see Rasmussen and Williams (2006)). The kernel representation of the state space also has many computational benefits, which allow one to efficiently handle approximations of policy functions over large and complex state space and also learn with sparse and little data, both of which are real life features (e.g. Powell (2007), Gershman et al. (2015)).

[14] Thus, in contrast to a literature on learning under misspecification (eg. Gagnon-Bartsch et al. (2020)), our later insights on ergodic "learning traps" do not arise from updating an a-priori misspecified distribution.

actions, such as budget and borrowing constraints. For example, in the consumption-savings application we discuss below, agents observe their income y and understand how a budget constraint imposes a deterministic restriction between any given consumption policy function $c(y)$ and an implied savings rule $a(y) = y - c(y)$. As a result, given the primitives on the prior beliefs (e.g. σ_c^2 and ψ), there is no change in the structure of uncertainty facing the agent whether reasoning occurs over over $c^*(y)$ or over the optimal savings rule $a^*(y)$, since the latter case is simply a deterministic translation of the Gaussian Process distribution over $c^*(y)$ to one over $a^*(y) = y - c^*(y)$. Thus, the model delivers equivalent behavioral implications either way. In contrast, if the state y is imprecisely perceived, as it is often assumed in the canonical literature on imperfect information, such equivalence may not hold.[15]

3.2 Noisy reasoning signals and updating beliefs

The agent does not simply act on her prior beliefs, but in any given period t can deliberate on the optimal course of action, which produces additional information about the unknown optimal action $c^*(y_t)$. In this way, through time, the agent gradually learns more about the unknown policy function. We model the outcome of the reasoning process as an unbiased, but noisy, signal about the actual optimal action at the current state y_t

$$\eta_t = c^*(y_t) + \varepsilon_t,$$

where $\varepsilon_t \overset{iid}{\sim} N(0, \sigma_{\eta,t}^2)$. The variance of the signal noise, $\sigma_{\eta,t}^2$, is endogenous, and chosen by the agent subject to an information cost that we detail later. Formally, the reasoning signals reduce perceived uncertainty by updating the agent's beliefs about the unknown function c^*, with the *conditional* distribution of beliefs following a tractable Kalman-filter like recursion.

Lemma 2. *Given the time-0 prior belief $c^* \sim \mathcal{GP}(\widehat{c}_0, \widehat{\sigma}_0)$, the time-t conditional beliefs $c^* \big| \{\eta^t, y^t\}$ are distributed as $\sim \mathcal{GP}(\widehat{c}_t, \widehat{\sigma}_t)$ with moments given by the recursive expressions*

$$\widehat{c}_t(y) = \widehat{c}_{t-1}(y) + \alpha_t(y)(\eta_t - \widehat{c}_{t-1}(y_t)), \tag{2}$$

$$\widehat{\sigma}_t(y, y') = \widehat{\sigma}_{t-1}(y, y') - \alpha_t(y)\widehat{\sigma}_{t-1}(y', y_t) \tag{3}$$

where $\widehat{c}_t(y) \equiv E_t(c^*(y)|\eta^t)$ and $\widehat{\sigma}_t(y, y') \equiv \mathrm{Cov}(c^*(y), c^*(y')|\eta^t)$ are the posterior mean and covariance functions, $\widehat{\sigma}_t^2(y) \equiv \widehat{\sigma}_t(y, y)$ denotes the posterior variance at a given y and

[15] The typical assumption in the consumption-savings literature on imperfect perception of wealth is to let savings be the residual action that simply clears the otherwise imperfectly perceived budget constraint (see for example Sims (2003) and Maćkowiak and Wiederholt (2015)). However, in a model of inattention, Reis (2006) shows that behavior is different when the residual action is consumption as opposed to savings.

$$\alpha_t(y) \equiv \frac{\widehat{\sigma}_{t-1}(y, y_t)}{\widehat{\sigma}^2_{t-1}(y_t) + \sigma^2_{\eta,t}}$$

is the effective signal-to-noise ratio of the time-t signal η_t.

Proof. Details are in Appendix A. □

We thus model reasoning as a form of '*fact-free*' learning – our agent already knows all objective payoff relevant information, as encoded in the sufficient state variable y_t. Her deliberation efforts instead help her deduce specific implications of these objective facts for her optimal course of action, and thus reduce the subjective uncertainty she faces.[16] A more intense reasoning effort helps the agent obtain a better sense of the 'right answer' of the question what to do right now, i.e. a more precise estimate of the unknown optimal action.

In addition, to the extent to which the outcome of reasoning is noisy and imperfect, this process of internal deliberation implies that ultimately agents exhibit *stochastic choice* as they act on their eventual updated beliefs (as we detail below). Thus, consistent with a large experimental evidence on such probabilistic (as opposed to deterministic) choices, even conditioning on the same observed state (and history) agents' actions may differ.[17]

3.3 The local nature of information and learning

A fundamental feature of our framework is that information is 'local', in the sense that information about the value of the optimal policy at a point y, i.e. $c^*(y)$, is less informative about the value of the function at other $y' \neq y$ as the distance between y' and y increases. This is due to the underlying non-parametric nature of the estimation, and can be seen most directly in the fact that the prior covariance function is decreasing in the distance $||y - y'||$.

As a result, in period t, when agents update beliefs about the optimal action at their current circumstances ($c^*(y_t)$), those updates put a higher weight on reasoning signals η_{t-k} for which y_{t-k} is closer to the current situation, i.e. a lower $||y_t - y_{t-k}||$. Thus, agents rely more on past information derived in situations that are *similar* to the current one, as opposed to not weighing or simply putting more weight on 'recent' memories (i.e. η_{t-k} for small k).

[16] Reasoning is thus simply internal reflection that helps the agent get closer to the optimal decision even without what an outside observed would register as new objective information. We share this view with a theory literature such as Aragones et al. (2005) and Alaoui and Penta (2016). Our specific implementation of this concept allows us to further represent internal deliberation as a tractable Bayesian non-parametric regression, subject to episodic memory and resource rationality, as we detail further below.

[17] See Mosteller and Nogee (1951), Tversky (1969), Ballinger and Wilcox (1997), Hey (2001) among others for evidence and Ratcliff and McKoon (2008), Manzini and Mariotti (2014), Woodford (2014) for recent modeling approaches based on bounded rationality, surveyed more generally in Johnson and Ratcliff (2014).

To get some intuition, notice that by iterating on equation (2) the updated belief at time-t can be expressed as a weighted sum of signals η_{t-k} and the unconditional prior $\widehat{c}_0(y)$:

$$\widehat{c}_t(y) = (1 - \sum_{k=0}^{t} \omega_{t-k}(y))\widehat{c}_0(y) + \sum_{k=0}^{t} \omega_{t-k}(y)\eta_{t-k} \tag{4}$$

where the time-t weight on each of the η_{t-k} signals is given by

$$\omega_{t-k}(y) \equiv \alpha_{t-k}(y) \prod_{s=t-k+1}^{t} (1 - \alpha_s(y)).$$

This weight depends on the value of y for which beliefs are updated, and in particular any given signal η_{t-k} has the strongest effect on beliefs $\widehat{c}_t(y)$ for y values close to the particular state realization y_{t-k} where the reasoning signal η_{t-k} was obtained (thus maximizing $\alpha_{t-k}(y)$).

To illustrate, in Lemma 3 below we consider the two straightforward cases of updating beliefs at times $t = 1$ and $t = 2$, when the agent updates based on one and two reasoning signals respectively. For simplicity of the exposition, in this Lemma we assume that the signal-noise variances of the two signals is the same: $\sigma^2_{\eta,1} = \sigma^2_{\eta,2} = \sigma^2_{\eta}$, but this is not essential. In the first case, at $t = 1$, the weight on η_1 is a particularly clean and revealing expression, which directly showcases the fact that the updating weight on a signal is the highest in the neighborhood of the state realization at which the signal was obtained. In the second case, at $t = 2$, the updating weights on the two signals η_1 and η_2 are more complicated and less intuitive visually (we leave the formulas to the Appendix), but they still imply the basic result that when agents update beliefs about the optimal action at some state value y, the update puts a higher weight on whichever of the two signals is closer. Similar results can be proved for the case of arbitrarily many signals and arbitrary signal-noise variances $\sigma^2_{\eta,t}$.

Lemma 3. *At time $t = 1$, the single signal η_1 gets a weight $\omega_1(y)$:*

$$\omega_1(y) = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_\eta^2} \exp(-\psi(y - y_1)^2).$$

The posterior variance is hence the lowest at $y = y_1$, and increases with $||y - y_1||$:

$$\widehat{\sigma}_1^2(y) = \sigma_c^2(1 - \omega_1(y)).$$

At $t = 2$, the update incorporates two signals, and puts a higher weight on the closer signal:

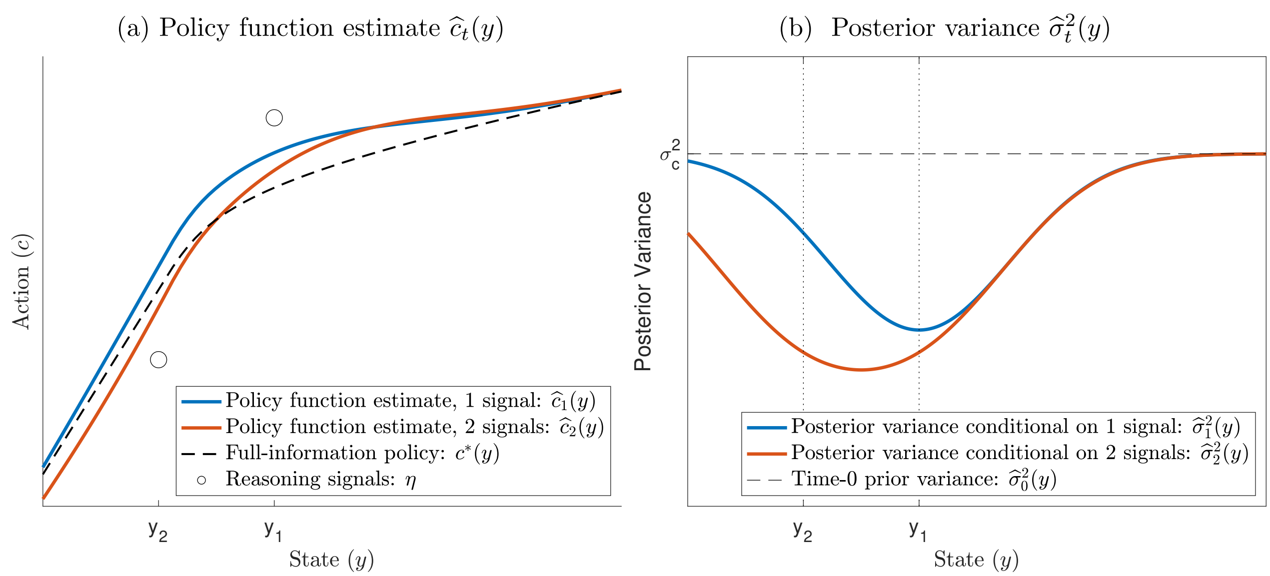
$$\omega_2(y) > \omega_1(y) \textit{ if and only if } ||y - y_2|| < ||y - y_1||$$

Proof. Details are in Appendix A. □

To visualize the basic result, Figure 1 plots the resulting posterior mean and variance functions at $t = 1$ and $t = 2$, when η_1 contains a positive surprise and η_2 a negative one (the two circles) – intuitively, the "surprise" in the signal is akin to the agent getting an idea or insight about the problem that he is trying to solve that differs from his prior. For example, thinking about a consumption-savings application, a positive "surprise" would mean that the agent deduces his prior belief about consumption is sub-optimally low, and that he should be consuming at a higher level.

Starting with the update based on one signal, the blue line in panel (a) shows that $\widehat{c}_1(y)$ is indeed most affected by the positive surprise in η_1 in the neighborhood of where that signal is centered – i.e. for y close to y_1. Moreover, conditional on both signals, e.g. $\widehat{c}_2(y)$ (red line), the updated belief is high in the neighborhood of signal η_1, but low in the neighborhood of η_2, showcasing how in different parts of the state space the agent largely relies on different signals, thus displaying episodic memory. The resulting state- and history-dependent uncertainty reduction can be seen most directly in panel (b), where we plot the posterior variance, which is the lowest in the part of the state space where the signals are centered, and rises back to the unconditional variance in the far parts of the state space.

Figure 1: Conditional beliefs and episodic memory



This local effect of information is both interesting and empirically relevant, given the evidence on the importance of 'episodic memory' reviewed earlier (e.g. Tulving (2002)). That is, both in the data and in our model, agents appear to put higher weight on specific memories (signals in our framework) that are similar to the current circumstances (i.e. current state).

Another feature of the local nature of information is that if two past signals are observed in proximity to one another, the posterior uncertainty can be the lowest in the neighborhood *between* the two signals (as illustrated by the red line in panel (b)).

Corollary 2 (Spillovers). *There is a threshold $\tau > 0$ s.t. if $||y_1 - y_2|| < \tau$ then*

$$\widehat{\sigma}_2^2(y) < \widehat{\sigma}_2^2(y_1) = \widehat{\sigma}_2^2(y_2), \forall y \in (y_1, y_2), \text{ and moreover } \frac{\partial \tau}{\partial \psi} < 0$$

Intuitively, the agent can be more confident about the optimal action at in-between states y, because she can effectively rely on both prior instances of reasoning at y_1 and y_2, as those two noisy signals effectively "bracket" the implied optimal action for $y \in (y_1, y_2)$. The extent to which that such "bracketing" is effective depends on the agent's prior belief of how granular and situation-specific past deductions are. When experiences and knowledge are highly specific (high ψ), then separate signals are not very informative for the in-between ground. On the other hand, if the agent believes the true optimal to be a relatively smooth function (low ψ), then having two prior anchor points is very useful for updating the middle.

3.4 The cost-benefit tradeoff of reasoning

Having described the uncertainty over the unknown function $c^*(y)$ and the outcome of the reasoning process, we close our model by introducing an intuitive cost-benefit tradeoff in choosing the reasoning noise variance $\sigma_{\eta,t}^2$. This choice reflects the agent's intensity of deliberation – the more time and effort spent on thinking about the optimal behavior, the more precise is the resulting signal, and thus the more accurate are the resulting posterior beliefs. In this view, agents are "resource-rational" and choose the optimal amount of cognitive effort, given its cost and expected benefit. This approach is not only intuitive, but also supported by a wealth of empirical evidence documented in the cognitive sciences literature – for example, Levy and Baxter (2002) show that the basic physiological processes of the brain itself trade-off metabolic cost with resulting accuracy (see Lieder and Griffiths (2020) for a broader review of the evidence, and Gershman et al. (2015), Griffiths et al. (2015), and Shenhav et al. (2017) for specific examples).

To operationalize this idea, we model the costly cognitive effort as an information cost. In particular, any given signal η_t carries an utility cost that is proportional to the amount of information about the optimal action $c^*(y_t)$ revealed by that signal, where we measure information flow as the reduction in entropy, i.e. Shannon Mutual Information,

$$H(c^*(y_t)|\eta^{t-1}) - H(c^*(y_t)|\eta_t, \eta^{t-1}) = \frac{1}{2} \ln \left(\frac{\widehat{\sigma}_{t-1}^2(y_t)}{\widehat{\sigma}_t^2(y_t)} \right), \tag{5}$$

where $H(X)$ denotes the entropy of a random variable X, and is the standard measure of uncertainty in information theory (eg. Sims (2003)). Thus, equation (5) measures the reduction in uncertainty about the unknown optimal action $c^*(y_t)$ from seeing the new signal η_t, given the history of past deliberation signals η^{t-1}. The implicit assumption here is that cognitive effort is directly proportional to the resulting information about the optimal action.

To model the "benefit" of the costly reasoning process, we draw on a long tradition in costly information models and assume that the agent is facing a standard tracking problem where, given the known state y_t, the agent wants to choose her action, c_t, to be as close as possible to the action implied by the unknown optimal policy function, $c^*(y_t)$. Thus, in choosing her action at in any given period t she minimizes the expected squared deviations

$$\min_{c_t,\sigma^2_{\eta,t}} \mathbb{E}_t(c_t - c^*(y_t))^2, \tag{6}$$

where $\mathbb{E}_t$ denotes the conditional expectation over the distribution of c^* with moments recursively determined by equations (2) and (3). The solution to the tracking problem is to act according to the best estimate of the unknown optimal action, and thus set $c_t = \widehat{c}_t(y_t)$. Given this choice of time-t action, the expected quadratic deviation from the optimal in any given period is simply the posterior variance under the agents beliefs: $\mathbb{E}_t(c_t - c^*(y_t))^2 = \widehat{\sigma}^2_t(y)$.

Thus, the cost-benefit tradeoff of reasoning can be cast as the information problem

$$\min_{\widehat{\sigma}^2_t(y_t)} \widehat{\sigma}^2_t(y_t) + \kappa \ln\left(\frac{\widehat{\sigma}^2_{t-1}(y_t)}{\widehat{\sigma}^2_t(y_t)}\right), \tag{7}$$

$$\text{s.t.} \qquad \widehat{\sigma}^2_t(y_t) \leq \widehat{\sigma}^2_{t-1}(y_t).$$

The first component is the benefit of reasoning, in the form of a lower dispersion of the action c_t around the unknown optimal action $c^*(y_t)$ – i.e. $\widehat{\sigma}^2_t(y_t)$. The second represents the cost of reasoning, which is proportional to the information content of the new time-t signal, with the parameter κ controlling the implied constant marginal cost of a unit of information. For example, κ will be higher for individuals with a higher cost of deliberation – either because they have a higher opportunity cost of time or because their particular deliberation process takes longer to achieve a given improvement in precision. In addition, κ would also be higher if the economic environment facing the agent is more complex, and thus the optimal action is objectively harder to figure out – for example solving a difficult math problem.

Lastly, the minimization in (7) is subject to the "no forgetting constraint" $\widehat{\sigma}^2_t(y_t) \leq \widehat{\sigma}^2_{t-1}(y_t)$ which ensures that the chosen variance of the noise in the signal, $\sigma^2_{\eta,t}$, is non-negative. Otherwise, the agent can gain utility by "forgetting" some of her past information.

The optimal deliberation choice that solves equation (7) is given by

$$\widehat{\sigma}_t^{*2}(y_t) = \min\left[\kappa, \widehat{\sigma}_{t-1}^2(y_t)\right],$$

meaning that the optimal level of posterior variance which equates the marginal benefit and cost of reasoning is simply κ. Intuitively, the desired precision in actions is larger when the deliberation cost κ is lower. The *min* function enforces the no-forgetting constraint – if the agent's beginning-of-period conditional variance over the optimal action at y_t is lower than the optimal target κ, then she does not acquire any further information and the posterior variance at time t remains equal to $\widehat{\sigma}_{t-1}^2(y_t)$. This leads to the following optimal behavior.

Proposition 1. *The optimal signal noise variance is given by*

$$\sigma_{\eta,t}^{*2} = \begin{cases} \frac{\kappa\widehat{\sigma}_{t-1}^2(y_t)}{\widehat{\sigma}_{t-1}^2(y_t)-\kappa} & , \textit{ if } \widehat{\sigma}_{t-1}^2(y_t) \geq \kappa \\ \infty & , \textit{ if } \widehat{\sigma}_{t-1}^2(y_t) < \kappa \end{cases} \tag{8}$$

and this in turn implies the time-t action

$$c_t = \widehat{c}_t(y_t) = \widehat{c}_{t-1}(y_t) + \alpha_t^*(y_t)(c^*(y_t) + \varepsilon_t - \widehat{c}_{t-1}(y_t)), \tag{9}$$

where the optimal weight put on the new reasoning signal, $\alpha_t^*(y_t)$ depends on the current state y_t and the history $\{y^{t-1}, \sigma_\eta^{t-1}\}$ of past signals' location and precision:

$$\alpha_t^*(y_t) \equiv \frac{\widehat{\sigma}_{t-1}^2(y_t)}{\widehat{\sigma}_{t-1}^2(y_t) + \sigma_{\eta,t}^{*2}} = \max\left[1 - \frac{\kappa}{\widehat{\sigma}_{t-1}^2(y_t)}, 0\right]. \tag{10}$$

Proof. Details are in Appendix A. □

Thus, since posterior uncertainty $\widehat{\sigma}_t^2(y)$ is state and history dependent, both the optimal reasoning choice, in the form of signal-noise variance $\sigma_{\eta,t}^2$, and the effective action c_t are also state- and history-dependent.

The key qualitative implications of our framework are driven by this endogenous state and history dependence of actions and beliefs. In particular, for state realizations y_t where the precision of initial beliefs is far from its target (high $\widehat{\sigma}_{t-1}^2(y_t)$), the agent chooses to acquire a more precise current signal η_t and hence puts a bigger weight on it in the resulting action $\widehat{c}_t(y_t)$ (high $\alpha_t^*(y_t)$). In contrast, for state realizations close to the position of previous signals η_{t-k}, the precision of initial beliefs is high (low $\widehat{\sigma}_{t-1}^2(y_t)$) and the agent finds it optimal to not acquire much additional information. At such 'familiar' states the optimal $\alpha_t^*(y_t)$ is relatively small, and the resulting action will be primarily driven by the beginning-of-period beliefs

$\widehat{c}_{t-1}(y_t)$. By equation (10), the optimal $\alpha_t^*(y_t)$ may even become zero when the accumulation of prior reasoning information around y_t is sufficiently strong, as highlighted by Corollary 2.

This basic feature of the reasoning intensity choice gives rise to behavior that appears to follow a 'habitual' form in familiar parts of the state space, where prior deliberation has reduced uncertainty about the optimal action sufficiently so that $\alpha_t^*(y_t)$ is small. However, agents do not mechanically follow such past-experience based heuristics – beliefs are likely to be revised if the state y_t moves into unfamiliar territory, where uncertainty over the optimal action is relatively high. In that case, the agent chooses a high precision of the current period signal, and thus a high $\alpha_t^*(y_t)$, leading to behavior driven by the outcome of the new reasoning effort, rather than past deductions. In this way, the model is endogenously generating a System 1/System 2 type of decision-making (an established psychological phenomenon, e.g. Stanovich and West (2000)), with heuristic-like behavior in familiar circumstances (akin to System 1), with a change to a deliberative approach in unfamiliar circumstances (System 2).

Towards observable implications. Overall, the main qualitative features of our framework center around the results that (a) in more familiar circumstances agents have lower incentives to reason anew about what is the best course of action, and (b) due to their noisy perceptions they may end up with different views on the optimal action. While these properties are grounded in extensive evidence from behavioral economics, neuroscience and psychology, the particular implications about observable behavior crucially depend on what is the distribution of the state variable, and hence what is the familiar part of the state space.

To get some intuition, consider first the case where y_t follows an exogenous distribution independent of the reasoning errors. Since both the prior and the signals are centered at the true c^*, the updated beliefs in equation (4) become

$$\widehat{c}_t(y_t) = \widehat{c}_0(y_t) + \sum_{k=0}^{t} \omega_{t-k}(y_t)(\eta_{t-k} - \widehat{c}_0(y_t)) = c^*(y_t) + \sum_{k=0}^{t} \omega_{t-k}(y_t)\varepsilon_{t-k}.$$

Because reasoning errors are idiosyncratic, they wash out on average, and therefore, while there is still noise in observed behavior (i.e. at any given instance $\widehat{c}_t(y_t) \neq c^*(y_t)$), the *typical* behavior, i.e. $\int \widehat{c}_t(y_t) d\varepsilon^t$, is not systematically different from the full-information $c^*(y_t)$. This finding echoes a common argument that the rational model might be a good approximation of aggregate behavior, even if individual agents are making mistakes.

In contrast, our mechanism implies that observed behavior can be systematically altered when the history of mistakes affects the evolution of the state - for example, consumption choices affect the accumulation of wealth through the budget constraint. In that case, there is a correlation between the path of errors ε_{t-k} and that of the encountered states y_{t-k}. In turn, this affects the optimal reasoning intensity $\sigma_{\eta,t}^2$ and hence generates a correlation between the *weight* put on a signal η_{t-k} in the update of beliefs and the error realization of that signal.

As a result, the typical estimate, and thus behavior, is no longer guaranteed to equal $c^*(y_t)$, as the signal errors do not necessarily wash out. The particular observable implications, measured by an outside analyst as the *joint* distribution of actions and states, depend on the broader economic structure and the way errors feed into the law of motion of the state. To examine this feedback in detail, in the next section we consider an application of our framework to a canonical consumption-savings problem, a mechanism which is often at the heart of macroeconomic models and naturally features an endogenous state – wealth.

4 A consumption-savings model with costly reasoning

To explore the feedback mentioned above and emphasize, in particular, the effects of heterogeneity stemming from the reasoning errors, we study consumption-savings behavior subject to our reasoning friction in the setting of incomplete markets.

To this end, we consider an otherwise standard Aiyagari (1994) economy populated by a continuum of ex-ante identical households, indexed by i, that share the same preferences in the form of a non-satiable and concave utility $u(c_{i,t})$ of consumption. Each household inelastically supplies her stochastic endowment of labor $s_{i,t}$ at a constant wage w. These income shocks $s_{i,t}$ are iid across time and agents and drawn from a time-invariant distribution $\mathcal{S}$ with a mean of one. The agents' ability to reduce their consumption exposure to this risk is limited - there is only one asset, in the form of a homogeneous physical capital that earns a constant rental rate $\widetilde{r}$ and depreciates at rate $\delta \in (0,1)$. The resulting budget constraint is

$$c_{i,t} + a_{i,t} = (1+r)a_{i,t-1} + ws_{i,t},$$

where $r \equiv \widetilde{r} - \delta$, $a_{i,t-1}$ is the amount of capital held by agent i at the end of period $t-1$ and $a_{i,t}$ is the current choice of savings. As in standard models of incomplete markets, this optimal asset choice is subject to a borrowing constraint $\underline{a} \geq 0$ such that

$$a_{i,t} \geq -\underline{a}.$$

The aggregate production function is also standard - it takes as input the average capital $K = \int a_i di$ and employment $H = \int s_{i,t} di$, and produces $K^\alpha H^{1-\alpha}$, with $\alpha \in (0,1)$. The role of this side of the economy is to determine the rental rate and the wage from the usual firm first-order conditions $\widetilde{r} = \alpha K^{\alpha-1} H^{1-\alpha}$ and $w = (1-\alpha)K^\alpha H^{-\alpha}$, respectively. Given the assumed inelastically supplied labor and i.i.d. labor supply shocks s_{it}, we have $H = 1$.

4.1 Decision problem

We first describe the optimal policy function under full information, which the households try to estimate subject to the reasoning friction developed in Section 3. The assumption of i.i.d. exogenous income shocks $s_{i,t}$ means that the sufficient state for the agent's decision is the available "cash-on-hand" defined as

$$y_{i,t} \equiv (1+r)a_{i,t-1} + ws_{i,t}.$$

The key property here is that the future state $y_{i,t+1}$ is determined partly endogenously from the current choice of consumption $c_{i,t}$, as well as by the random realization of $s_{i,t+1}$.

Each agent is interested in solving the same general problem

$$V(y_{i,t}) = \max_{c_{i,t},a_{i,t}} u(c_{i,t}) + \beta \mathbb{E}_t V(y_{i,t+1}), \tag{11}$$

subject to the budget constraint $a_{i,t} + c_{i,t} = y_{i,t}$ and the borrowing limit $a_{i,t} \geq -\underline{a}$.

The consumption policy function that solves the problem in (11) can be written as

$$\widetilde{c}^*(y_{i,t}) = \min(y_{i,t} + \underline{a}, c^*(y_{i,t})), \tag{12}$$

where the kink at $y_{i,t} + \underline{a}$ arises from the borrowing limit. The policy $c^*(y_{i,t})$ gives the optimal action taking into account future borrowing constraints, but ignoring today's constraint.

Reasoning friction

As in the general framework of Section 3, households do not have free cognitive access to the full-information policy function c^*, but perfectly observe their financial resources $y_{i,t}$, the borrowing constraint $\underline{a}$, and estimate the unknown function c^* via costly reasoning signals

$$\eta_{i,t} = c^*(y_{i,t}) + \varepsilon_{i,t}, \;\; \varepsilon_{i,t} \sim N(0, \sigma_{\eta,i,t}). \tag{13}$$

Agents have the common time-0 prior that $c^* \sim \mathcal{GP}(\widehat{c}_0, \widehat{\sigma}_0)$, which as discussed earlier is centered at the truth, i.e. $\widehat{c}_0 = c^*$, and has the covariance function $\widehat{\sigma}_0$ as in Corollary 1.[18]

An agent i at time t makes two choices: reasoning intensity, by selecting the desired signal-noise variance $\sigma^2_{\eta,i,t}$, and consumption level $c_{i,t}$. In choosing the precision of the reasoning signal $\eta_{i,t}$, the agent optimally trades off its cost and benefit, as described in detail

[18] The underlying full-information policy c^* is not affected by the reasoning friction. Because the signals $\eta_{i,t}$ in equation (13) are drawn from that policy, the interpretation here is one of a bounded rational agent that is 'naive' (in the terminology of O'Donoghue and Rabin (1999)).

in Section 3, and optimally selects a signal precision that generates the posterior variance

$$\widehat{\sigma}^2_{i,t}(y_{i,t}) = \min\left[\kappa, \widehat{\sigma}^2_{i,t-1}(y_{i,t})\right]. \tag{14}$$

By Proposition 1, the optimal choice of $\sigma^2_{\eta,i,t}$, together with beginning-of-period beliefs $\widehat{c}_{i,t-1}(y_{i,t})$ and $\widehat{\sigma}_{i,t-1}(y_{i,t})$, lead to the following conditional expectation of $c^*(y_{i,t})$:

$$\widehat{c}_{i,t}(y_{i,t}) = \widehat{c}_{i,t-1}(y_{i,t-1}) + \alpha_{i,t}(y_{i,t})(c^*(y_{i,t}) + \varepsilon_{i,t} - \widehat{c}_{i,t-1}(y_{i,t-1})), \tag{15}$$

where the optimal weight put on the current reasoning signal $\eta_{i,t}$ can be expressed as

$$\alpha_{i,t}(y_{i,t}) = \max\left[1 - \kappa/\widehat{\sigma}^2_{i,t-1}(y_{i,t}), 0\right]. \tag{16}$$

Lastly, taking into account the borrowing limit is straightforward as the agent perfectly observes $y_{i,t} + \underline{a}$, and hence the actual consumption is[19]

$$c_{i,t} = \min(y_{i,t} + \underline{a}, \widehat{c}_{i,t}(y_{i,t})).$$

Heterogeneity and equilibrium

Agents in our economy are heterogeneous for two reasons. The first is the idiosyncratic income shocks $s_{i,t}$, which are a common feature of standard models too. The second reason is that agents obtain stochastic histories of reasoning errors $\varepsilon_{i,t}$, leading to different information sets $\{\eta_i^t\}$ and perceived optimal decision rules $\widehat{c}_{i,t}(y)$. Therefore, the distribution of agent types in the costly reasoning model is richer than in the standard full information Aiyagari (1994) model. In particular, an agent's type at time t is characterized by the following: (i) prior conditional variance and consumption *functions* $\widehat{\sigma}^2_{i,t-1}$ and $\widehat{c}_{i,t-1}$; (ii) observed cash on hand $y_{i,t} = (1+r)a_{i,t-1} + ws_{i,t}$; and (iii) current period reasoning error $\varepsilon_{i,t}$. We denote the set of objects that determine an agent state at time t as $\tau_{i,t} \equiv (\widehat{\sigma}_{i,t-1}, \widehat{c}_{i,t-1}, y_{i,t}, \varepsilon_{i,t})$.

Moreover, the heterogeneity in *conditional actions* (as opposed to states) in our costly reasoning model is also different compared to the full-information model, where the agents act under the same policy function $c^*(y)$ and only differ in their actions due to transitory differences in their states $y_{i,t}$. In our model instead, agents also differ in their effective policy function $\widehat{c}_{i,t}(y)$, and this systematically changes the properties of the average behavior and also the wealth distribution in the long-run.

[19]The Gaussian noise in signals suggests that extreme signals can lead to a negative estimate: $\widehat{c}_{i,t}(y) < 0$. In the numerical implementation we prevent this by imposing another constraint that $c_{i,t} > 0$. In practice, we find that this is not a problem at our calibration.

In other words, even though the reasoning errors are iid across agents, the interaction between the errors and the endogenous law of motion of individual states implies that the reasoning errors matter in the aggregate. To make this point concrete, we denote by $\lambda_t(\tau)$ the time-t probability distribution over the agent types $\tau_{i,t}$ and note that the constrained-optimal behavior of the agents induces a law of motion for $\lambda_t(\tau)$. We are interested in characterizing the properties of this distribution at the stationary equilibrium, which we define below.

Definition 1. *A stationary equilibrium is a full-information policy function $c^*(y)$, a probability distribution $\lambda(\tau)$ and positive real numbers (K, r, w), such that*

(1) the time-invariant prices (r, w) satisfy

$$r = \alpha K^{\alpha-1} - \delta; \;\; w = (1-\alpha)K^{\alpha}.$$

(2) the policy function $c^*(y)$ solves the full-information problem in equation (11)

(3) the reasoning choice $\sigma^2_{\eta,i,t}$ and the consumption choice $c_{i,t}$ satisfy Proposition 1, while conditional beliefs $\widehat{c}_{i,t}(y)$ and $\widehat{\sigma}_{i,t}(y, y')$ follow Lemma 2.

(4) given $y_{i,t}$ and a consumption choice $c_{i,t}$, cash on hand evolves as $y_{i,t+1} = (1 + r)(y_{i,t} - c_{i,t}) + ws_{i,t+1}$, where $s_{i,t+1}$ is an iid draw from the time-invariant distribution $\mathcal{S}$.

(5) the distribution $\lambda(\tau)$ is time-invariant, with a law of motion induced by (1)-(4).

(6) aggregate capital equals the average of the households' asset decisions

$$K = \int_{\tau} [y - c(\tau)] \, d\lambda(\tau).$$

4.2 Selection effects in updating beliefs

To build intuition for the properties of the stationary equilibrium, here we illustrate the basic mechanism by following the sequence of choices and beliefs evolution of two agents, who are identical except for the realization of their reasoning errors, $\varepsilon_{i,t}$ – i.e. they face identical income shock realizations $s_{i,t}$ and initial conditions $a_{i,0}$.

At time-1 both agents choose the same reasoning effort, which by equation (14) is such that $\sigma^2_{\eta,i,1} = \kappa/\alpha_1^*$. Thus, the time-1 reasoning signals the agents draw have the identical distribution $\eta_{i,1} = c^*(y_1) + \varepsilon_{i,1}$, $\varepsilon_{i,1} \sim N(0, \kappa/\alpha_1^*)$. Here $\alpha_1^* \equiv \alpha_1(y_1) = 1 - \frac{\kappa}{\sigma_c^2}$ is the resulting signal-to-noise ratio, which does not depend on the value of y_1 since $\widehat{\sigma}^2_0(y) = \sigma^2_c$ for all y.

Period 1: ex-post heterogeneous beliefs

At time-1 agents are ex-ante identical and choose the same reasoning effort, but make different consumption choices due to the idiosyncratic realizations of the reasoning errors:

$$c_{i,1} = \min[y_1 + \underline{a}, c^*(y_1) + \alpha_1^* \varepsilon_{i,1}].$$

At this stage observed behavior across agents only differs due to the iid shock $\varepsilon_{i,1}$, which would wash out on average. However, the reasoning signals do not only generate stochastic choice, but also lead to systematic differences in the policy function estimates $\widehat{c}_{i,1}(y)$ across the state space, which leads to persistent differences in the *savings* rates and thus affects the future wealth accumulation of the two agents. In particular, using equation (2) agent i's conditional estimate of the policy function for any value of y is

$$\widehat{c}_{i,1}(y) = c^*(y) + \exp(-\psi(y - y_1)^2)\alpha_1^* \varepsilon_{i,1}.$$

The key qualitative differences arise from a difference in the signs of $\varepsilon_{i,t}$. To showcase this, for the purposes of this example we assume that agent 1 receives a signal with a positive error ($\varepsilon_{1,1} \equiv \varepsilon > 0$), while agent 2 receives a signal with a mirror-image negative error ($\varepsilon_{2,1} = -\varepsilon < 0$).[20] As the reasoning errors $\varepsilon_{i,1}$ makes the two agents over- or under-estimate, respectively, optimal consumption at *all* states y, we label agent 1 as a "consumer" or C, and similarly label agent 2 a "saver" or S.

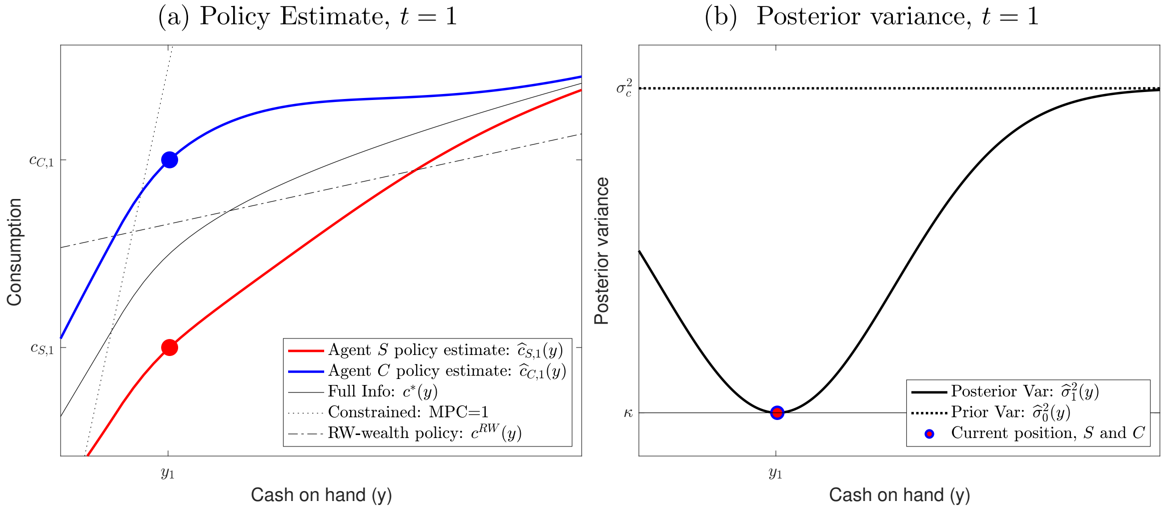
To illustrate, in Figure 2 panel (a) we plot the updated estimates of the policy function $\widehat{c}_{i,1}$, for agent C (in blue) and agent S (in red), as induced by the reasoning signals $\eta_{i,1}$. The blue and red circles represent the respective actions taken at time-1, $c_{i,1}$, and also mark the position of y_1 in this example. We also plot the full-information policy c^* (solid black line), and the dotted line plots the action implied by the borrowing constraint, i.e. $c(y_{i,t}) = y_{i,t} + \underline{a}$.

Figure 2, panel (a) shows that agent C updates beliefs upward and ends up with an "over-consumption" bias relative to the (common) prior belief $c^*(y)$ throughout the state space, while the opposite is true for agent S. The plot also showcases that the shifts in the estimated policy functions $\widehat{c}_{C,1}(y)$ and $\widehat{c}_{S,1}(y)$, relative to the prior beliefs $c^*(y)$, are strongest at cash-at-hand values y close to y_1 – e.g. notice that the estimates $\widehat{c}_{i,1}(y)$ converge to $c^*(y)$ for large y. Again, this reflects the local nature of the reduction of uncertainty.

To see this local reduction in uncertainty directly, consider the resulting posterior

[20] To simplify exposition and intuition, we assume throughout this illustration that ε is small enough so that agent 1 does not hit the borrowing constraint with his time-1 action.

Figure 2: Conditional beliefs at $t = 1$



variance function $\widehat{\sigma}^2_{i,1}$, which is the same for both agents due to their identical choice of $\sigma^2_{\eta,1}$:

$$\widehat{\sigma}^2_{i,1}(y) = \widehat{\sigma}^2_1(y) = \sigma^2_c(1 - \alpha^*_1 \exp(-2\psi(y - y_1)^2)). \tag{17}$$

This function is plotted in Figure 2, panel (b) and has a characteristic U-shape that visualizes the state and history dependence of posterior uncertainty. The state-dependence is embodied in the fact that $\hat{\sigma}^2_1(y)$ is not a constant function, but varies with the value of the state y. The history dependence is due to the fact that $\widehat{\sigma}^2_1(y)$ is increasing in the distance between y and y_1, the state at which the time-1 signal is most informative.

Policy estimates and savings rates

The state and history dependence of uncertainty interacts with the endogenous dynamics of $y_{i,t}$ because cash-on-hand depends not only on concurrent income shocks, but also on the effective savings rate of an agent as determined by his consumption choices.

In particular, we define the savings rate as the expected change in cash-on-hand

$$\mathbb{E}_t(y_{i,t+1}) - y_{it} = (1 + r)(y_{i,t} - \widehat{c}_{i,t}(y_{i,t})) + w. \tag{18}$$

To visualize how any given policy estimate $\widehat{c}_{i,t}(y)$ affects the savings rate, we define the counter-factual policy function, $c^{RW}(y)$, that implies a savings rate of zero. If an agent was to follow that consumption rule, then his cash-on-hand will be a random walk, i.e. $\mathbb{E}_t(y_{i,t+1}) = y_{i,t}$. We can back this policy out, by setting equation (18) to zero and obtain:

$$c^{RW}(y_{i,t}) = \frac{r}{1+r} y_{i,t} + \frac{1}{1+r} w, \tag{19}$$

We plot this object with the dash-dot line in Figure 2. If an agent follows a consumption rule that lies above $c^{RW}(y)$ then that agent's wealth is expected to fall (i.e. has a downward drift on average), and if an agent's consumption policy lies below, then he is actively saving and assets are expected to grow on average. In particular, using equations (18) and (19):

$$\mathbb{E}_t(y_{i,t+1}) - y_{it} = (1+r)(c^{RW}(y_{i,t}) - \widehat{c}_{i,t}(y_{i,t})).$$

Hence, the cash-on-hand value at which a consumption policy crosses the $c^{RW}(y)$ line from below (i.e. an *upcrossing*) is also the *steady-state* level of wealth that an agent following that policy is tending towards. This cash-on-hand value is also known as the "target buffer-stock" in the terminology of Carroll (2004), and it is a stable point that cash-on-hand reverts to, because whenever current financial resources are above that level, then the consumption policy is above $c^{RW}(y)$, hence wealth has a downward drift and vice versa.

Under full information, there is a unique steady-state level of cash-on-hand that all agents tend toward, which we call $\bar{y}^*$, given by the intersection of the solid black line ($c^*(y)$) and the dash-dot line ($c^{RW}(y)$) in Figure 2. In that model, any wealth heterogeneity is only due to transitory income shock realizations, not due to differences in systematic savings rates.

In contrast, in our model agents generally have different steady-state wealth levels $\bar{y}_{i,t}$ due to the dispersion in their policy function estimates $\widehat{c}_{i,t}(y)$. For example, from Figure 2 we see that the over-consumption bias in agent C's beliefs and the under-consumption bias in agent S's beliefs lead to differences in their long-run wealth levels, as implied by their respective intersections with $c^{RW}(y)$, with the "consumer" tending towards a lower steady-state wealth: $\bar{y}_{C,t} < \bar{y}^* < \bar{y}_{S,t}$. Thus, our model displays wealth heterogeneity both due to luck (i.e. idiosyncratic income shocks) and due to differences in saving rates, which result here from differential consumption policy estimates.

Period 2: selection effects and systematic heterogeneity

The time-1 differential savings rates of our two example-agents result in different time-2 wealth levels, even though the new, time-2 income shock is the same for both, specifically:

$$y_{C,2} = (1+r)(y_1 - c_{C,1}) + ws_2 < (1+r)(y_1 - c_{S,1}) + ws_2 = y_{S,2}. \tag{20}$$

Importantly, on top of generating additional heterogeneity, the differential saving rates also interact with the state dependent optimal reasoning choice, as summarized by Proposition 2.

Proposition 2. *The optimal reasoning intensity and the weight of the new signal in updating beliefs are both increasing in distance from location of the previous reasoning signal:*

$$\frac{\partial\sigma^2_{\eta,i,2}}{\partial||y_{i,2}-y_1||} < 0 \text{ and } \frac{\partial\alpha_{i,2}(y_{i,2})}{\partial||y_{i,2}-y_1||} > 0.$$

Therefore, agent C reasons more than agent S, i.e. $\alpha_{C,2}(y_{C,2}) > \alpha_{S,2}(y_{S,2})$, if and only if

$$s_2 < 1 + \frac{(1+r)}{w}(c^*(y_1) - c^{RW}(y_1)) \equiv \bar{s}$$

Proof. Details are in Appendix A. □

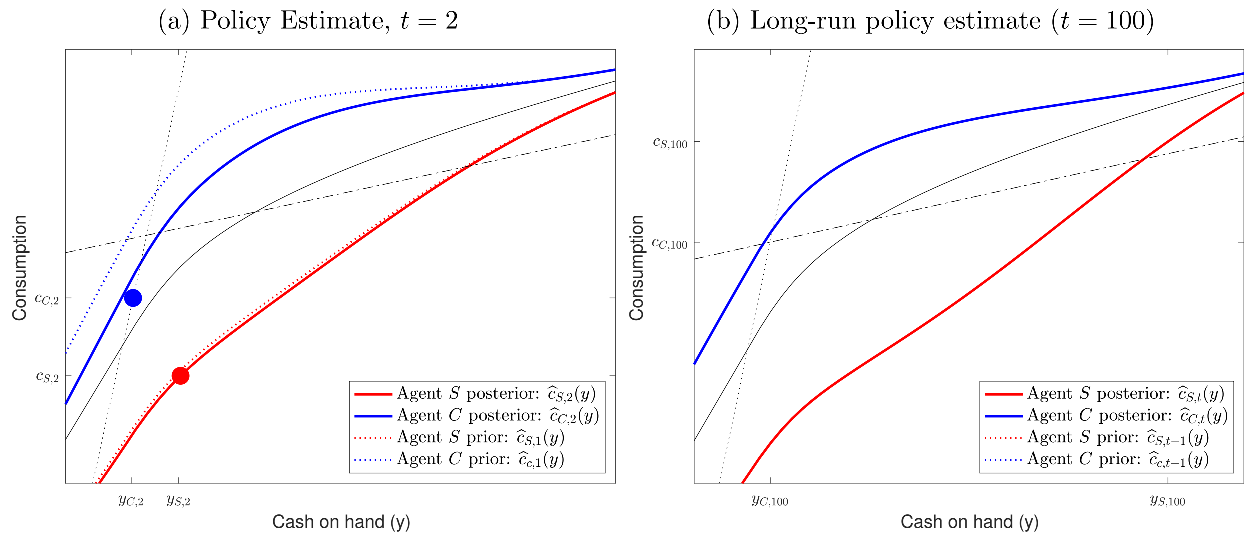
The first result in Proposition 2 follows directly from Proposition 1 and Lemma 3. By Proposition 1, the optimal variance of the reasoning error is $\sigma^2_{\eta,i,2} = \frac{\kappa\widehat{\sigma}^2_1(y_{i,2})}{\widehat{\sigma}^2_1(y_{i,2})-\kappa}$ and the resulting signal to noise ratio is $\alpha_{i,2}(y_{i,2}) = 1 - \frac{\kappa}{\widehat{\sigma}^2_1(y_{i,2})}$. Given that the variance function $\widehat{\sigma}^2_1(y)$ is lowest at y_1 and U-shaped around it (e.g. Figure 2 and Lemma 3), an agent i chooses to reason more the larger is the distance $||y_{i,2} - y_1||$. Intuitively, if an agent finds herself facing a significantly different level of cash-on-hand than the level at which she reasoned previously, then uncertainty about the current optimal action is higher and will warrant further reasoning than otherwise.

The change in wealth, however, is not simply a function of the size of the exogenous income shock, but also depends on the savings rate, and this is what underpins the second result of Proposition 2. Since agent C's has a negative savings rate and thus a negative drift in wealth, we can show that he experiences a larger shift in his state variable, i.e. $||y_{C,2} - y_1|| > ||y_{S,2} - y_1||$, if and only if the time-2 income shock is low enough – $s_{i,2} = s_2 < \bar{s}$.

Such "low" realizations of current income essentially compound the effects of the "over-consumption" bias of agent C and his resulting low savings rate, and thus pushes his cash-on-hand even further away from y_1 than otherwise. On the other hand, such a "low" income shock acts to counter-balance the "under-consumption" bias of agent S, hence limiting the change in her cash-on-hand. Thus, in this situation, agent C faces higher uncertainty about his time-2 optimal action as compared to agent S, i.e. $\widehat{\sigma}^2_1(y_{C,2}) > \widehat{\sigma}^2_1(y_{S,2})$, and hence agent C chooses to reason with a higher intensity.

As a result, given $s_2 < \bar{s}$, agent C updates his beliefs by more, as seen in panel (a) of Figure 3, where we plot the time-2 updated beliefs $\widehat{c}_{i,2}(y)$ (solid line) together with the previous period estimates $\widehat{c}_{i,1}(y)$ (dashed lines). Since in this case $\alpha_{C,2}(y_{C,2}) > \alpha_{S,2}(y_{S,2})$ (as per Proposition 2), the change in beliefs of agent C is substantially bigger than the revision in agent S's beliefs, which hardly change (red dashed line vs solid red line) as the latter chooses to acquire little new information. Naturally, the reverse behavior would emerge when $s_2 > \bar{s}$.

Figure 3: Conditional beliefs evolution



In that case, agent S faces a more unfamiliar part of the state space than agent C, for whom an unexpectedly large shock s_2 effectively moves him back towards the familiar state y_1.

Learning traps and long-run behavior

These results showcase that there is a selection in *when* the two agents choose to revise beliefs. The differential savings rates interact with new income shock realizations so that agents with positive savings rates are more likely to significantly revise beliefs after a positive income shock has pushed their wealth significantly above past wealth levels, and vice versa.

This has important implications for the long-run beliefs. Take agent S for example. On average this agent is likely to experience a drift up in wealth, and accumulate new information mainly at such higher wealth levels. Since the new signals are unbiased and her initial policy function estimate is low (which characterizes her as a "saver"), the new information will tend to revise her belief up, towards the true $c^*(y)$.

Importantly, however, uncertainty reduction is local, and the new signals will mainly update the estimated policy locally to the higher wealth levels at which the new signals arrive. Thus, rather than a uniform shift of beliefs up, the revisions will instead change the estimated policy at higher levels of cash-on-hand y, while largely preserving the low estimate in the neighborhood of the initial wealth y_1. Hence, even as the target wealth level $\bar{y}_{S,t}$ changes as new information arrives (since then the intersection point of her estimate $\widehat{c}_{S,t}(y)$ with $\widehat{c}^{RW}(y)$ changes) the evolving beliefs are still likely to result in a target wealth $\bar{y}_{S,t}$ above the initial wealth y_1, and thus agent S maintains her positive drift in assets.

As agent S builds wealth, she reaches the vicinity of her target $\bar{y}_{S,t}$. In that neighborhood, the stochastic nature of income shocks are bound to eventually push her cash-on-hand level $y_{S,t}$ above $\bar{y}_{S,t}$. By the characteristic property of $\bar{y}_{S,t}$ as a target, the agent then exhibits a negative savings rate, and thus a downward drift in asset. Critically, this downward drift tends to move the state *back* into a part of the state space that is now "*familiar*", in the sense that the agent has experienced such lower levels of wealth in the past, and has already reasoned up to the optimal amount there. Because of this familiarity, she is unlikely to choose to reason much again, and will instead behave according to the policy estimate she enters the period with. Thus, as the agent has now reduced uncertainty to his target levels on both sides of her current target wealth level $\bar{y}_{S,t}$, she settles in a stable stochastic steady state, where assets fluctuate around that intersection point and beliefs are not likely to change by much anymore (unless the agent experiences large shocks).[21]

We label this situation a "*learning trap*". Its defining feature is a sequence of reasoning signals that establishes stable wealth dynamics in the neighborhood of those same signals' positions, and thus a high likelihood for the agent's wealth to remain within this "familiar" part of the state space. As wealth fluctuates in that low uncertainty region, the agent has little incentive to reason further. But without new signals beliefs remain the same, perpetuating the consumption behavior that keeps wealth stable within this region to begin with. While all agents are likely to eventually fall in such a learning trap and largely stop updating beliefs, the heterogeneity in agents' histories of signals and income shocks will lead to significant cross-sectional dispersion in the eventual stable target wealth levels $\bar{y}_{i,t}$.

For example, consider the case of agent C instead. The interaction of initial negative savings rate and state dependent reasoning choice will, on average, lead to revision of beliefs that preserves a low target wealth level. Eventually, income shocks would also push this agent to reason on both sides of that low target level, which will then get established as a stable steady-state and become his familiar region of the state space. Importantly, an outsider that analyzes this agent's behavior will observe a pattern of *habitually* (and surprisingly, from the perspective of c^*) high level of consumption at that persistently low level of wealth.

Overall, for some agents their crossings and target wealth levels are low and close (or even at) the borrowing constraint, while for others these levels are significantly higher than the full-information steady state wealth $\bar{y}^*$. We illustrate this observation in panel (b) of Figure 3, where we show the resulting estimates of agent C and S after 100 periods. While agent S has a long-run belief that implies a target wealth level significantly above $\bar{y}^*$,

[21] Technically, this stability of wealth does not occur with probability one because of the exogenous income shocks $s_{i,t}$, which can in principle take a large enough value (i.e. positive in the case of this saver agent) to push the agent outside of the familiar region implied by the upward crossing. These large shocks may thus lead the agent to further update her beliefs, changing the target wealth $\bar{y}_{S,t}$, even if those updates may be small in nature. We quantify these statements in our numerical analysis.

agent C has retained a policy function which implies a target level significantly below $\bar{y}^*$. In fact, because agent C's policy estimate is above c^{RW} up until it hits the constraint, that target wealth level is in fact the constraint itself, a property that in Section 5 we connect to the empirical puzzle of the lack of savings by poor agents. Moreover, the Figure also illustrates the insight that once the agents settle in their learning traps they do not reason much anymore, i.e. the dotted lines are virtually indistinguishable from the solid lines.[22]

Two key agent types: (1) hand-to-mouth & (2) rich with high MPC

Two key qualitative features of the long-run beliefs are worth nothing. First, agent C settles around a stochastic steady state level $\bar{y}_C$ which is close to the constraint. Hence, this type of agent will exhibit behavior that looks akin to "*hand-to-mouth*" persistently. In contrast, the full-information model will only generate hand-to-mouth agents due to bad luck, in the sense of a sequence of low income shocks. Such hand-to-mouth status is temporary, as agents will aggressively save to move back towards $\bar{y}^*$ (notice that $c^*(y)$ is significantly below $c^{RW}(y)$ near the constraint, signifying a high savings rate under full-information). In contrast, in our model hand-to-mouth behavior is a *persistent* characteristic, and one that will be helpful in distinguishing between models when we discuss the quantitative results.

Second, consider the rich agents and their marginal propensity to consume (MPC), or formally the local *slope* of the consumption policy function, measuring the consumption change after an unexpected income shock. In the full information case, by the well known result that the slope of $c^*(y)$ converges to that of $c^{RW}(y)$ as y gets big, the MPCs of rich agents is very low – close to $r/(1+r)$.

In contrast, in our model, because local *steepness* makes wealth dynamics stable, the slope of the policy estimate $\widehat{c}_{i,t}(y)$ around their target wealth level $\bar{y}_{i,t}$ ends up being high. Thus, in our model rich agents are characterized by both high target wealth levels *and* by *high MPCs*. Their typical *past* savings behavior is to accumulate assets, summarized by the high observed steady-state level $\bar{y}_{i,t}$. Their typical *current* behavior is the high slope of the estimated $\widehat{c}_{i,t}(y)$, which means these rich agents aggressively (i) consume out of positive income shocks and (ii) reduce their consumption when hit by negative shocks, a behavior that keeps their wealth stable. In this section we have used the example of agent S to illustrate this joint dynamics of wealth accumulation (see panel (a) in Figure 3) and eventual high local MPC (in panel b). As we quantify and argue in Section 5 this result of high MPC for rich agents is of large significance for this class of models.

Both of these characteristic qualitative features of our model, dispersion in target

[22] Visually, the prior (dotted red line) in panel (b) is different than the solid red line of panel (a), indicating that some reasoning has occurred from the earlier time 2 to this illustrated current period.

wealth levels and excess sensitivity of consumption, differentiate the framework from the standard approach of imperfect observation of the state. In that case, as analyzed for example by Luo (2008), consumption under-reacts to income shocks (i.e. lower MPC than in the full-information model). The intuition is straightforward – agents observe y imprecisely, hence they respond weakly to income shocks and thus consumption exhibits inertia, a key result in the literature on inattention. Moreover, even though agents in that framework are confused about the actual value of y, they still all tend to the same long-run stochastic steady state, hence do not display dispersion in target wealth levels. As we show below, both of our model-implied features are borne out in the data.

To fully evaluate and quantify the properties of our learning mechanism we now turn to a numerical implementation of the stationary equilibrium of our economy, and describe the resulting ergodic joint distribution of states, actions and beliefs.

5 Numerical analysis

Whenever possible, we follow the standard parametrization considered in Aiyagari (1994). Households have log-utility with a discount factor of $\beta = 0.96$. The i.i.d. labor income shock is drawn from a log-normal distribution $\ln(s_{i,t}) \sim N(-\frac{\sigma_s^2}{2}, \sigma_s^2)$, with $\sigma_s = 0.2$, and there is no borrowing allowed, i.e. $\underline{a} = 0$. On the production side, the capital share and the annual depreciation are set to standard values of $\alpha = 0.36$ and $\delta = 0.08$, respectively.

Next, we turn to the parameters governing the reasoning friction. First, to help with the otherwise slow and long history-dependent evolution of beliefs of individual agents in the computation of the stationary steady state, we introduce a form of discounting of past reasoning signals. We opt for the tractable modeling assumption that agents face i.i.d. Poisson information shocks, where with probability θ an agent's history of accumulated reasoning signals becomes obsolete and that agent's beginning-of-period beliefs reset to the time-0 prior. One interpretation of this discounting scheme is based on viewing agents as finitely lived, where conditional on death, the agent transfers his assets and the resulting continuation utility off to the offspring. However, the transfer of reasoning information about the optimal policy is imperfect across generations, which for simplicity we assume leads to full discounting of past information. In our parametrization we set $\theta = 0.02$, so that the economy is continuously repopulated with agents that on average re-start their learning problem every 50 years.

Second, as also discussed in Section 3, we aim to reduce the degrees of freedom intrinsic in specifying the prior beliefs, through restrictions that resemble the rational-expectations idea of utilizing "model-consistent" priors, as follows. First, the common prior mean function $\widehat{c}_0(y)$ is set equal to the full-information policy function $c^*(y)$. This leaves us with the parameters σ_c^2 and ψ of the covariance function $\widehat{\sigma}_0(y, y')$, which govern the uncertainty around $\widehat{c}_0(y)$. We set these equal to what an econometrician would estimate if he were to observe simulations from the model. In particular, we look for a fixed point such that given the values of $\{\sigma_c^2, \psi\}$, if an econometrician uses the resulting ergodic distribution of reasoning signals $\eta_{i,t}$ as data, he would recover the same values of $\{\sigma_c^2, \psi\}$ as used to simulate $\eta_{i,t}$. This essentially restricts the assumption on prior uncertainty to be model consistent, and also has a connection with the practice of estimating hyper-parameters in Bayesian statistics. More details on this fixed-point procedure are given in the Appendix.

Employing this model-consistent priors strategy, we calibrate the remaining degree of freedom, namely the marginal cost of reasoning κ, by exploiting the tendency for agents in our model to settle in "learning traps". Among other things, these traps link our mechanism to one particularly challenging fact - while empirically the bottom 20% in the US wealth distribution have roughly zero net assets, standard model predicts that very few agents should be in that position due to strong precautionary saving motives (see Krueger et al. (2016) for a discussion). This insight motivates us to set κ to target this moment, given the fixed-point restriction over $\{\sigma_c^2, \psi\}$ and the other parameters described above.

Putting everything together, the resulting calibration for the reasoning parameters is $\{\sigma_c^2, \psi, \kappa\} = \{0.77, 0.05, 0.48\}$. Those values suggest that agents indeed face non-trivial amount of uncertainty in the optimal policy ($\sigma_c^2 > 0$) and its shape ($\psi > 0$). The rest of this section discusses the implications of this model parameters in terms of the resulting reasoning behavior and the joint distribution of observables, i.e. wealth and consumption.

5.1 Reasoning properties in the stationary equilibrium

As previously discussed in Section 4.2, reasoning slows down when wealth dynamics become locally stable to a "familiar", low-uncertainty region of the state space. We now detail the characteristics of the resulting "learning traps" in the ergodic distribution, and in the process quantify some of the qualitative statements we have made in that Section 4.2.[23]

To operationalize the definition of what constitutes a "learning trap", we first compute the target wealth level $\bar{y}_{i,t-1}$, as implied by the point at which the policy function estimate that an agent i enters period t with, $\widehat{c}_{i,t-1}(y)$, crosses the counter-factual policy rule $c^{RW}(y)$:

$$\widehat{c}_{i,t-1}(\overline{y}_{i,t-1}) = c^{RW}(\overline{y}_{i,t-1}), \tag{21}$$

[23]To compute the stationary distribution we iteratively simulate an economy with 10,000 periods and 5,000 agents, and search for the values of the interest rate r and wage w which satisfy the definition of stationary equilibrium in Section 4. We compute the reported moments over the last 5,000 periods of the simulation.

Second, we define the agent i at time t to be *in* a "learning trap" *if* the uncertainty function he enters the period with evaluated at the crossing point $\overline{y}_{i,t-1}$ is such that

$$\widehat{\sigma}^2_{i,t-1}(\overline{y}_{i,t-1}) \leq \kappa. \tag{22}$$

In this case, the prior uncertainty over the optimal course of action is low enough at $\overline{y}_{i,t-1}$ so that at time t the agent would optimally choose not to reason further at this state and set the optimal reasoning intensity to zero, $\alpha_{i,t}(\overline{y}_{i,t-1}) = 0$.[24] In contrast, if condition (22) is reversed, then we refer to an agent i at time t as *not in* (or, out of) a learning trap.

The intuition behind our operational definition of learning traps is the following. First, the crossing point $\overline{y}_{i,t-1}$ gives agent's i hypothetical state value such that if the agent enters the period with cash-on-hand equal to $\overline{y}_{i,t}$ and acts according to her prior policy estimate, then her cash-on-hand remains unchanged on average, i.e. $\mathbb{E}_{t-1}(y_{i,t}) = \overline{y}_{i,t-1}$. Second, if in that situation the agent would also choose not to reason further, then her policy estimate does not change either: $\widehat{c}_{i,t}(y) = \widehat{c}_{i,t-1}(y)$. Finally, due to the local nature of uncertainty reduction and the accumulation of prior information, if condition (22) holds it is likely to also hold in a neighborhood around $\bar{y}_{i,t-1}$ (per the insight on spillovers in Corollary 2). In that case, uncertainty is low not only exactly at $\bar{y}_{i,t-1}$ but also more generally around it and thus the agent will find it optimal to not update beliefs for a range of y close to $\bar{y}_{i,t-1}$, leading to her state and beliefs to jointly stabilize, in a stochastic steady-state sense, around $\overline{y}_{i,t-1}$.[25]

We evaluate this intuitive behavior by providing moments in Table 1 on the long-run characteristics of such learning traps. We first describe these moments for all agents and then we will also further condition on agent's wealth entering period t, the key state variable.

Ergodic properties of learning traps

In Panel (A) of Table 1 we first report that about 70% of agents in the ergodic distribution are in such learning traps. Moreover, on average, the first time an agent i enters a learning trap is one-third into the agents' typical lifetime. Once they enter into this state, the probability of a typical agent to ever exit that first trap is roughly 30%. Thus, learning traps are ubiquitous in the ergodic distribution and most agents do not even exit the very first one they encounter.

In Panel (B) of Table 1 we report moments on the reasoning choices, cash-on-hand behavior and MPCs for agents currently in and out of a learning trap. First, we show that

[24] As exemplified in Corollary 2, the inequality in equation (22) may be strict when the knowledge spillovers present in the accumulation of prior reasoning information around $\overline{y}_{i,t-1}$ are sufficiently strong.

[25] A particular implication is that on average the actual cash-on-hand $y_{i,t}$ is indeed close to $\overline{y}_{i,t-1}$ for people defined as being in a "learning trap". We show that this is indeed true below. We did not choose an operational definition of learning trap based on a distance metric between $y_{i,t}$ and $\overline{y}_{i,t}$ because that would require an ad-hoc assumption on what is a "small" distance.

our operational definition of learning traps is indeed picking up agents who find it optimal to do very little additional reasoning in the current period – for agents inside the trap updated beliefs $\widehat{c}_{i,t}(y_{i,t})$ put a typical weight of just $\alpha_{i,t} = 0.002$ on new reasoning signals $\eta_{i,t}$. This update is significantly smaller than for agents outside the trap, who have a typical $\alpha_{i,t} = 0.043$. To put these magnitudes in context, we note that the largest reasoning intensity occurs in the first period of any agent's life (at $\alpha = 1 - \kappa/\sigma_c^2 = 0.376$), when the prior uncertainty over $c^*(y)$ is at its highest. Compared to this initial learning, agents' learning eventually slows down but it does so significantly more for agents who find themselves in a learning trap.

Second, we find that for agents inside a trap find themselves at an average distance from their steady-state cash-on-hand, $|y_{i,t} - \overline{y}_{i,t-1}|$, that is much smaller than for agents outside of a learning trap: 0.48 vs 4.05 respectively. This highlights that our operational definition of a learning traps indeed picks up agents that are both close to their target wealth-level and find that region "familiar" and hence see little need for further reasoning.[26] In contrast, agents outside a learning trap are far away from their target wealth level $\bar{y}_{i,t-1}$, and thus still transitioning and, on average, face higher uncertainty as their wealth keeps drifting.

Third, the notion of agents settling in a learning trap we described in Section 4.2 also implies that they should not only stay close to $\overline{y}_{i,t}$ (which can move over time as the policy estimate $\widehat{c}_{i,t}(y)$ gets updated), but should also remain in the same overall neighborhood of the state space. And indeed, because agents in a learning trap only adjust beliefs slightly in a typical period, their crossing point also barely moves – in Table 1 we find that the mean absolute change $|\overline{y}_{i,t} - \overline{y}_{i,t-1}|$ for agents inside a trap is 0.04, dramatically smaller than the 0.629 for agents currently outside a current trap, whose beliefs are still evolving.

Fourth, our endogenous reasoning mechanism makes agents settling into learning traps a probabilistic statement. As indicated for example by Proposition 2, the local nature of uncertainty reduction leads to less intense reasoning in familiar states. Since part of the current cash-on-hand $y_{i,t}$ is determined stochastically, through the income shock $s_{i,t}$, it follows that the extent to which $y_{i,t}$ remains in a previously familiar territory is a stochastic event, even conditional on assets $a_{i,t-1}$, prior beliefs $\widehat{c}_{i,t-1}(y)$ and $\widehat{\sigma}^2_{i,t-1}(y)$. In Table 1 we report that the probability of being outside a learning trap at $t+1$, conditional on being in one at t is positive, but nevertheless very small, at 0.8%.[27] While exiting a trap is thus very unlikely, the probability of entering one at $t+1$, conditional on currently being out of it, is significantly

[26] Even if agents would have exactly converged to this stochastic steady state in period $t-1$, their current state $y_{i,t}$ would move around it simply due to the current exogenous income shocks $ws_{i,t}$, with a standard deviation of 0.23. Thus, in terms of magnitudes, this exogenous variation accounts for about half of the typical difference $|y_{i,t} - \overline{y}_{i,t-1}|$ of 0.48 for agents inside the trap, with the other half being the result of the endogenous wealth accumulation in the ergodic distribution

[27] Note that this moment looks one period ahead, and conditions on a trap encountered at any point in time, and thus it is different from the exiting moment report in Panel (A).

larger, at 6.5%. Put together, these two conditional probability define the model's implied ergodic transition matrix of agents moving in and out of a current learning trap.

Finally, in concluding Section 4.2 we noted that the mechanism produces two qualitative features that directly speak to observable variables that have been of particular interest to macroeconomists and can help connect the model to the data. First, the model produces a mass of agents that are *persistently* "Hand-to-Mouth" (HtM), as those agents that settle in learning traps at target wealth levels $\bar{y}_{i,t}$ close to (or even at) the borrowing constraint. Second, the manner in which agents away from the constraint establish stable wealth dynamics is via consumption policy estimates that are significantly steeper locally than $r/(1+r)$, thus the model is predicted to generate surprisingly high MPCs for wealthy, unconstrained agents.

To evaluate these predictions and further understand the inter-workings of the mechanism, next we split the agents in our simulation by their wealth entering period t. In particular, to facilitate links with the empirical evidence, we follow Zeldes (1989) and Aguiar et al. (2020) and define an agent as being in a HtM status if her beginning-of-period net worth $a_{i,t-1}$ is less than two months of labor earnings, and to be non-HtM otherwise.[28]

Moments	All agents		HtM		Non-HtM	
(A) Frequency						
Unconditional prob to be in a trap	0.702		0.878		0.653	
Mean first time in trap/life length	0.317		0.224		0.338	
Prob to ever exit the first trap	0.29		0.105		0.331	
(B) Conditional on being *in* or *out* of a trap						
	In	*Out*	*In*	*Out*	*In*	*Out*
Mean reasoning $\alpha_{i,t}$	0.002	0.043	0.001	0.062	0.002	0.041
Mean deviation $\lvert y_{i,t} - \overline{y}_{i,t-1} \rvert$	0.482	4.049	0.431	1.049	0.5	4.342
Mean change $\lvert \overline{y}_{i,t} - \overline{y}_{i,t-1} \rvert$	0.041	0.628	0.008	0.101	0.053	0.68
Prob switch out/in at $t+1$ (%)	0.81	6.48	0.09	8.10	1.07	6.42
MPC at time t	0.354	0.152	0.791	0.825	0.177	0.109

Table 1: An agent at time t is in (or out) of a learning trap if condition (22) holds (or reversed). The two columns 'All agents' report ergodic moments conditional on being in/out of a trap for all agents. The two columns 'HtM' report the corresponding in/out moments only for agents whose $a_{i,t-1} < w/6$, while the 'Non-HtM' columns report moments for the rest of the agents.

Learning traps across HtM status

The insights of how learning traps manifest themselves differently for HtM or non-HtM agents relate to the intuition of section 4.2. First, we report in Panel (A) of Table 1 that

[28] With i.i.d. risk, in our annual model this threshold is $w/6$, where w is the stationary equilibrium wage.

the frequency of learning traps is more pronounced for the former type (88% of which are inside a trap vs 65% for non-HTM), experience earlier their first trap (22% into their typical life) and are significantly less likely to exit it (at 10%) than non-HtM agents. The reason is two-fold – near the constraint the optimal policy agents learn about is steeper and hence it is more likely to establish a steep upward-crossing estimate $\widehat{c}_{i,t}(y)$ for low y, and at the same time, the constraint itself can help stabilize wealth dynamics by acting as an absorbing state.

As introduced earlier, a distinctive implication of our mechanism are the surprisingly high MPCs of agents that have settled around high target wealth levels $\bar{y}_{i,t}$. Table 1 quantifies this by showing that for the group of non-HtM agents the average MPC is 0.18 when inside a learning trap, but only 0.11 when outside of one. This showcases that a crucial feature of settling inside a learning trap is a consumption policy estimate with a significantly higher local slope than $r/(1+r)$. Thus, the model indeed generates a surprisingly high MPC for wealthy, unconstrained agents, which we argue in Section 5.2 is both an important empirical feature, and one that sharply differentiates our model from the bulk of the literature.

Lastly, another interesting implication of the model is that there is a significant amount of heterogeneity in both MPCs and saving rates even conditional on current wealth levels. For example, the last two columns of Table 1 shows that the MPCs of the *non-HtM* differ significantly between agents in and out of learning traps. Moreover, the fact that agents outside of a learning trap are still transitioning towards their eventual (and dispersed) target wealth level generates significant heterogeneity in saving rates.[29]

5.2 Joint distribution of wealth levels and consumption functions

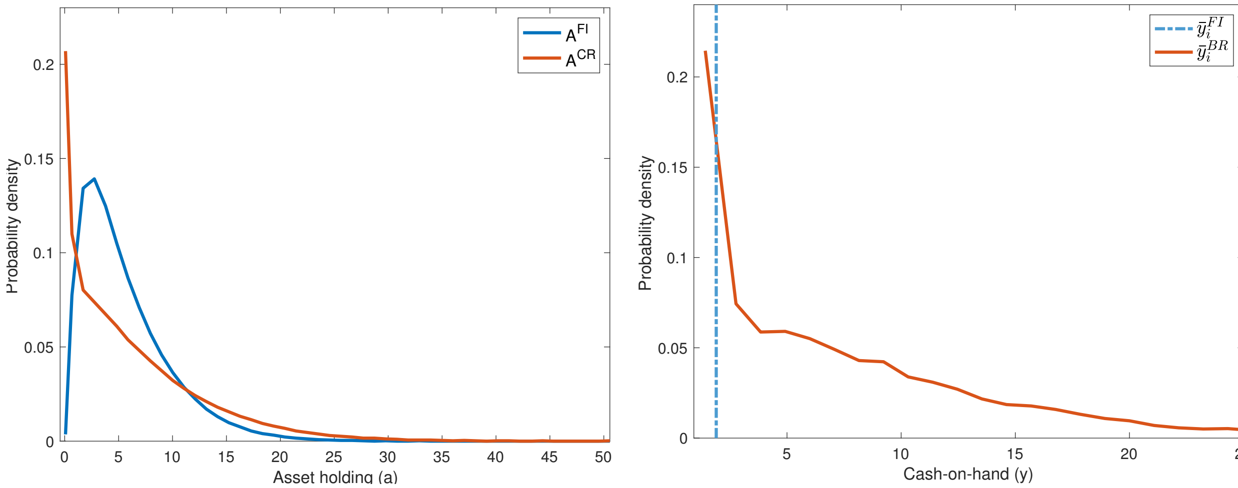
We now discuss several important observable implications of the mechanism. In particular, we argue that even in the context of an otherwise simple model a-la Aiyagari (1994), our costly reasoning (CR) mechanism significantly improves upon its full-information (FI) counterpart along two key dimensions: (1) more frequent and persistent hand-to-mouth (HtM) status, and (2) higher marginal propensities to consume (MPC), especially for wealthy, unconstrained agents.[30] In the process, the CR model also delivers larger wealth inequality overall.

Wealth distribution

We start by briefly discussing our mechanism's implication for wealth heterogeneity. Panel (a) of Figure 4 plots the stationary distribution of assets a_i in the benchmark CR economy (red line) and the counter-factual FI economy (blue line). A striking difference between the

[29] Recent studies by Lewis et al. (2019) and Bach et al. (2017), which document significant heterogeneity in consumption behavior orthogonal to wealth levels, provides suggestive evidence of these features of the model.

[30] The FI model has the same sequence of income shocks, but there is no uncertainty over c^* (i.e. $\sigma_c^2 = 0$).



(a) Wealth distribution

(b) $\bar{y}_i$ distribution

Figure 4: Wealth distribution.

two distributions is the large mass of low wealth agents in the CR model. Another notable qualitative difference is the larger density of rich agents in the CR model compared to the FI economy, exemplified by the slower decay of the right tail in the asset distribution. A key reason for the increased wealth dispersion, both in the left and right tails, is the dispersion in target wealth levels $\bar{y}_i$, shown in panel (b). Naturally, in the FI case this is a degenerate distribution with a single mass point, as all agents tend towards the same steady-state wealth. In our model, however, $\bar{y}_i$ is heterogeneous across agents, and its dispersion is similar to the overall heterogeneity in wealth (comparing panels (a) and (b)).

As a summary statistic for wealth inequality, we find that our benchmark model produces a significantly higher Gini coefficient (at 0.58), bringing the model half-way from its FI counter-part (at 0.39) to the US data (at 0.77, as reported by Krueger et al. (2016) for PSID in 2006). The model is still some ways short of the data, but this is not surprising given that we have intentionally chosen to present our mechanism in a stylized setting, with very simple income and asset structures. And consistent with the intuition above, a significant portion of this dispersion is due to the dispersion in target-wealth levels - the Gini coefficient of the distribution of $\bar{y}_i$ is 0.49, or 84% of its value for the overall wealth distribution.

Hence, target wealth levels play an important role in both generating poor and rich agents in our model, consistent with empirical work such as Bernheim et al. (2001), Ameriks et al. (2003), or Hendricks (2007), which emphasizes such 'unobserved' heterogeneity. Through the lenses of the learning traps implied by our model, this heterogeneity reflects persistent differences in the perception of the (otherwise common) policy function.

Ergodic beliefs

With this ergodic wealth distribution in mind, we now characterize the typical cross-sectional consumption behavior. We do so by showcasing the typical shapes of the agents' estimated consumption policy functions, evaluated in the neighborhood of the state that agents find themselves in the ergodic distribution. As we have argued earlier, it is precisely the endogenous selection of beliefs through time that makes this *joint* ergodic distribution of (state, beliefs) interesting and fit the data better than the FI model.[31] To help summarize the key insights, we average policy estimates across two groups of agents, *HtM* and *non-HtM*, as defined earlier. This dichotomy helps leverage some key commonalities in the consumption behavior of agents within those two types in our model, and is also useful when connecting to the data and other models, as we describe below. In particular, we compute

$$\widehat{c}_{group}(x) \equiv \int_{i \in group} \widehat{c}_i(y_i + x) di, \tag{23}$$

where the *group* defines being currently either in a *HtM*-status or *non-HtM* status, respectively, and y_i is the cash-on-hand value for agent i in the stationary distribution.[32]

Thus, $\widehat{c}_{group}$ is the average policy estimate as a function of deviations from the ergodic value of cash-on-hand for agent i – i.e. the function $\widehat{c}_{group}(x)$ evaluated at $x = 0$ given the average action of agents in that specific group at their own individual level of y_i. By varying x, we extract the typical *shape* of the policy function around y_i for agents inside the specific group. We also similarly compute the average shape of the optimal $c^*_{group}(x)$ and the PIH policy $c^{RW}_{group}(x)$ in the neighborhood of y_i, by substituting in the functions $c^*(y)$ and $c^{RW}(y)$, respectively, for $\widehat{c}_i(y)$ inside the integral of equation (23).[33]

The constructs $\widehat{c}_{group}(x)$ will pick up the key systematic patterns of consumption behavior across agents in a given group. If the noise in estimates around the typical cash-on-hand values y_i is random across agents, then $\widehat{c}_{group}(x)$ would also equal $c^*(y)$. However,

[31] As we detail in the Appendix, the cross-sectional average of policy function estimates, $\int \widehat{c}_i(y)di$, equals the true optimal policy $c^*(y)$ for *any* y. This occurs because, for any given y, some agents over-estimate, while others under-estimate, the optimal consumption $c^*(y)$, depending on their own history of income shocks and signal realizations. Thus, this result highlights that there is no mechanical average bias in the estimation procedure. Importantly, however, this unconditional average policy estimate *is not* representative of the typical consumption behavior. Indeed, due to the interaction between wealth and reasoning errors, the agents that are close to any given level of the state y at the stationary distribution are not randomly selected, as emphasized throughout this section and previously in section 4.2.

[32] We generally drop the subscript t to refer to the cross-sectional stationary distribution. Note also that $y_i \neq \bar{y}_i$, but as seen in Table 1, most agents are indeed close to their stochastic steady state $\bar{y}_i$ on average.

[33] The Appendix shows similarly constructed average typical policy functions $\widehat{c}_{group}(x)$ by wealth quintile instead. This more granular grouping approach yields very much the same results, hence for simplicity we focus the main text on the dichotomy between *HtM* and *non-HtM*.

Figure 5: Ergodic Policy Estimates

(a) Average estimated policy function for HtM (b) Average estimated policy function for non-HtM

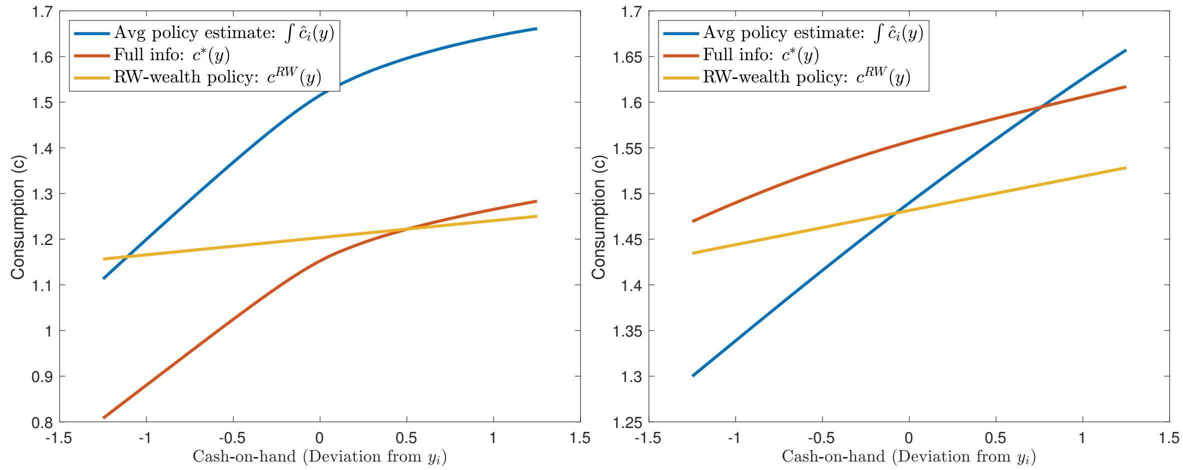


Figure 5 shows that this is not the case.[34]

Large and persistent mass of hand-to-mouth agents

First, in panel (a) of Figure 5 we plot the typical ergodic policy function for the HtM group. A striking feature is that the typical estimate of our boundedly rational agents differs from $c^*(y)$ primarily in its level, not shape, and in particular we note the sharp ranking

$$\widehat{c}_{HtM}(x) > c^{RW}_{HtM}(x) > c^*_{HtM}(x). \tag{24}$$

Thus, these low-wealth agents are not only typically significantly over-estimating optimal consumption, but are also, on average, de-accumulating assets (since $\widehat{c}_{HtM}(x) > c^{RW}_{HtM}(x)$, implying a negative saving rate).

This showcases that the typical HtM agent in our model looks very much like the example-agent C from Section 4.2. These agents have accumulated a sequence of signals that lead to an excessively high estimate of the consumption policy function, and thus a low saving rate. Hence, these agents also remain in this HtM status persistently, as their low saving rates (or equivalently low target wealth $\bar{y}_{i,t}$) keep their assets low. As their assets cycle in this now 'familiar' environment, they see no further need for reasoning, and thus become 'habitually' high consumption/low-asset type of agents. On the other hand, because the shape of the estimated policy function is similar to $c^*(y)$, their effective MPCs are very similar to that of a full-information agent that also finds himself in a HtM situation. The

[34] The x-axis is in terms of deviations from the ergodic value of cash-on-hand, and we plot a range of +/- five standard deviations of income shocks (i.e. $w * \sigma_l = 0.24$) around y_i.

difference is that the systematically low saving rates of our *HtM* agents ensure that this is both a persistent state and also one that is prevalent in the stationary distribution.

In contrast, under full-information agents save aggressively when close to the constraint (i.e. $c^*(y)$ is significantly lower than $c^{RW}(y)$ when assets are low), hence in that model *HtM* status is due to a long sequence of low income shocks, which is both a low probability and transitory status, as agents quickly build up their wealth once income shocks mean-revert. This well-known feature of the standard incomplete markets model makes it challenging for it to generate a large fraction of hand-to-mouth agents in the stationary distribution (as emphasized for example by Krueger et al. (2016) and Aguiar et al. (2020)).

Importantly, as argued convincingly by the recent work of Aguiar et al. (2020), the large fraction of poor, "hand-to-mouth" households in the data appears to be primarily driven by systematically lower savings rate among those households, rather than transitory outcome of bad income shocks or persistent differences in income processes. In Table 2, panel A, we demonstrate that due to the bounded rationality mechanism discussed so far, our model is able to quantitatively match several of these key features of the data on *HtM* households.

First, in the stationary distribution of the costly reasoning model 23% of agents are *HtM* which matches the data up to a rounding error. In contrast, in the full-information counter-part of our model, the ergodic mass of *HtM* agents is just 1% – more than an order of magnitude smaller than that in the data.

Crucially, both in the data and in our model *HtM* status is not only prevalent, but also very persistent at the individual level. We connect to the empirical results Aguiar et al. (2020) by first regressing the expected *future* consumption growth of agents on their current *HtM* status, a regression that we replicate in our simulations by estimating:

$$\Delta \ln c_{i,t+2} = \beta_0 + \beta_1 HtM_{t,i} + \varepsilon_{i,t+2} \tag{25}$$

In the data, Aguiar et al. (2020) find that this univariate regression yields a coefficient β_1 that is virtually zero, signifying that there is no difference in the average consumption growth of current *HtM* agents and those with higher wealth. As they note, this runs counter to the standard model's implication that current *HtM* agents are on average expected to save out of the constraint quickly, and thus experience high future consumption growth. Indeed, consistent with this intuition, we find that in the FI version of the model consumption growth conditional on current *HtM* status is 1.5% higher than otherwise.

In contrast, our costly reasoning model matches the data (as we report in Table 2, panel A) because in our model the heterogeneity in wealth is driven to a large extent by differences in target wealth levels. Thus, the majority of the agents in our simulation are both close to their target wealth levels and intend to stay there, hence on average the *HtM* and *non-HtM* agents in our model have similar consumption growth rates, as they are not experiencing systematically different average growth rates in assets.

The insight of *HtM* agents having different target wealth levels leads Aguiar et al. (2020) to augment the regression in (25). In particular, as they do in the data, we define the *propensity* to be *HtM* as the fraction of time that a given agent i in our simulation finds himself in a hand-to-mouth situation ($FracHtM_i$) and add this as a regressor to estimate

$$\Delta \ln c_{i,t+2} = \beta_0 + \beta_2 HtM_{t,i} + \gamma FracHtM_i + \varepsilon_{i,t+2} \tag{26}$$

Aguiar et al. (2020) find that the coefficient on current *HtM* status is now significantly positive, $\beta_2 = 0.025$, while the coefficient on the propensity to be *HtM* on average is significantly negative, $\gamma = -0.038$. Additional results in that paper, show that these differences are unlikely to be due to differences in income processes or idiosyncratic risk, and thus conclude they are likely due to a mechanism where the *HtM* agents have a low target wealth level, hence unless they are close to it (as indicated by current *HtM* status) they have low savings rates, and thus a low consumption growth.

Indeed, this is exactly what happens in our model. Due to the differential target wealth levels, which essentially makes *HtM* status a *persistent* agent characteristic, our model is able to match the regression coefficients very well, delivering a $\beta = 0.035$ and $\gamma = -0.048$. On the other hand, the full-information model is sharply at odds with this empirical evidence, again because it lacks this differential target wealth levels mechanism.

The puzzling lack of saving by constrained agents

We just argued that the low savings rate of constrained agents in our model helps explain their persistently low wealth. The same mechanism rationalizes a related challenging empirical fact, documented by Ganong and Noel (2019). The major puzzle they pose is why constrained agents do not save out of income which is known to be available only temporary. Put differently, at the predictable moment when that income is no longer available these agents' consumption drops sharply, reflecting an apparent lack of saving and thus consumption smoothing. While standard models of liquidity constraints can explain why agents cannot borrow enough to smooth consumption, Ganong and Noel (2019) make a convincing argument that such models cannot explain the lack of *saving* in anticipation of a drop in income.

Our mechanism can instead rationalize this observation. The key intuition is that the majority of agents near the constraint in our model are there persistently, due to a high estimate of the *level* of the policy function and a resulting low long-run wealth level. In other

Moments	Data	Benchmark	Full info
	(1)	(2)	(3)
(A) Constrained agents			
Fraction of Hand-to-Mouth (HtM)	0.23	0.23	0.01
β_1 univariate regression	0	0.001	0.015
β_2 multivariate regression	0.025	0.035	0.015
γ multivariate regression	-0.038	-0.048	-0.010
$\Delta_{c,t+1}$	-0.12	-0.137	0.001
(B) MPC			
Average (aggregate) MPC	0.2-0.6	0.29	0.05
Average MPC | top 20% of assets	0.17	0.15	0.04
Average MPC | *non-HtM*		0.15	0.04
Average MPC | *HtM*		0.83	0.08

Table 2: We report moments from data in column (1) and the stationary distribution in our benchmark costly reasoning model in column (2). In our counterfactuals we keep parameters at their benchmark values but set $\sigma_c^2 = 0$ in column (3), to construct the full-information counter-part of our model. The data moments in Panel (A) are documented by Aguiar et al. (2020) and Ganong and Noel (2019), as we discuss in the text. In Panel (B) the range of credible estimates of the aggregate MPC is from Carroll et al. (2017), and estimate of the MPC of the rich from McDowall (2020).

words, in our model, such agents not only display a high MPC or elasticity of consumption, but also a habitual tendency to consume a lot, and thus dis-save, on average. As a result, these agents do not appear to have the same strong saving incentive as their full-information counterparts, even in the face of predictable income declines, as we quantify below.

To align with the empirical setting of Ganong and Noel (2019), we design the following experiment at the stationary equilibrium of our model. First, we look at the subset of agents relatively close to the constraint, which have an average MPC of 0.6, similar to the observed drop of spending out of unexpected income shocks for the agents that Ganong and Noel (2019) analyze.[35] Second, we replicate their observation that at the exhaustion of the UI benefits labor income can be expected to fall by an average of 40%, as a percent of its steady-state. In our model, with iid shocks, we recover this predictable fall at $t+1$ by simulating a time t labor income shock of size $+40\%w$. Given the iid nature of shocks, agents thus expect an average percentage change in labor income from t to $t+1$ of $\Delta_{s,t+1} = -40\%$. Third, we collect these agents' resulting consumption decisions $c_{i,t} = \widehat{c}_{i,t}(y_{i,t})$, the ensuing cash-on-hand evolution $y_{i,t+1} = (1+r)(y_{i,t} - c_{i,t}) + ws_{i,t+1}$, and consumption actions next period $c_{i,t+1} = \widehat{c}_{i,t+1}(y_{i,t+1})$.

[35]This means picking out the agents in the bottom 30% of the wealth distribution, which amounts to largely focusing on the *HtM* agents.

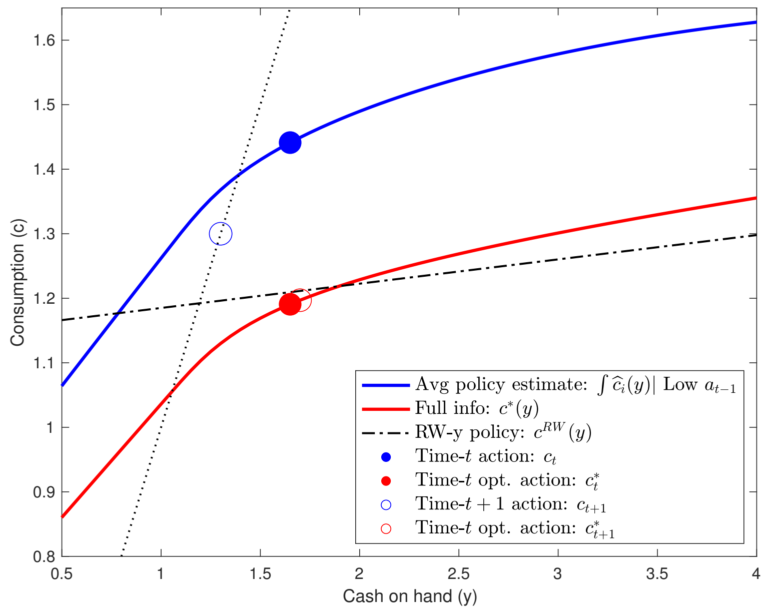


Figure 6: Lack of saving by bounded-rational constrained agents

We then compute the average percentage change in consumption between t and $t+1$, as income falls (predictably) by 40%:

$$\Delta_{c,t+1} = \int \frac{c_{i,t+1} - c_{i,t}}{c_{i,t}} di.$$

Finally, we also collect the counter-factual consumption and savings actions implied by the fully-rational policy $c^*_{i,t} = c^*(y_{i,t})$, conditional on the same initial cash-on-hand levels. The $t+1$ optimal consumption choices are thus $c^*_{i,t+1} = c^*(y^*_{i,t+1})$, where the $t+1$ cash-on-hand in that case is $y^*_{i,t+1} = (1+r)\left(y_{i,t} - c^*_{i,t}\right) + ws_{i,t+1}$, reflecting the counter-factual time-t optimal consumption choice. We similarly compute the average percent change in $c^*_{i,t+1}$.

Our key result here is that the boundedly rational agents experience on average a change in consumption between t and $t+1$ of $\Delta_{c,t+1} = -13.7\%$. This result is very puzzling from the perspective of the full-information policy function, which would instruct agents to save most of that temporary income and implement an almost perfect consumption smoothing between t and $t+1$, resulting in $\Delta^*_{c,t+1} = 0.09\%$. In contrast to actions taken under that full-information, our agents thus save too little out of their temporarily available liquidity at time t. This means that at $t+1$, when the average labor income is predicted to fall dramatically from time t, many agents are back to being close to the constraint and therefore having to cut back significantly on consumption. Remarkably, as we report in Table 2, this large drop in consumption is quite close to the average 12% drop in spending documented by Ganong and Noel (2019) at the exhaustion of the UI benefit.

The intuition for the lack of saving of these agents close to the constraint relates back to our analysis of the empirical moments reported by Aguiar et al. (2020), and is best understood by contrasting the level and the slope of the effective consumption policy function.

We use Figure 6 to visually illustrate the argument. The blue solid line plots the average policy function estimate for the sample of boundedly rational agents near the constraint that we analyze here, and the red line plots the optimal policy c^*. The key difference is that the typical $\widehat{c}_{i,t}(y)$ is significantly *higher* than c^*, even though the slopes are quite similar throughout. Thus, given an unexpected shock to income, both the poor boundedly rational and the poor optimal agents display the same MPC or consumption sensitivity. However, for any level of liquidity, the boundedly rational agents consume a higher fraction of the available resources and hence save significantly less on average.

Given the size of injected liquidity, the typical consumption action $c_t = \int c_{i,t} di$ of our constrained agents is the blue filled-in circle, sitting on the typical policy estimate. As we have seen in Figure 5 and illustrated by the example-agent C in Figure 3, this choice of c_t is significantly above the one implied by c^{RW} policy (by about 25%), implying that these agents save considerably too little to maintain the same cash-on-hand at $t+1$ as of time t. In fact, the large over-consumption relative to $c^{RW}(y_t)$ implies that the boundedly rational agents experience a sharp decline in cash-on-hand at time $t+1$, generating a drop between the typical c_{t+1} (the blue empty circle) and c_t of 14%. In contrast, under the counterfactual full-information, the typical optimal consumption action c_t^* (red filled-in circle) is close to its annuity value $c^{RW}(y_t)$, thus keeping cash-on-hand stable between t and $t+1$ and smoothing consumption through the predictable drop in income. In other words, while both agents similarly increase consumption upon the initial unexpected positive income shock in this exercise (reflecting high MPC), the optimal agent still consumes a relatively small fraction of his overall financial resources (even though consumption increased sharply), while the boundedly rational agent follows his habitual tendency to consume too much and dis-save.

Put together, our costly reasoning model, grounded in the psychological and neuroscience evidence of noisy, resource-rational reasoning based on episodic memory, formalizes and proposes a novel narrative for why agents do not easily leave the constraint. In this view, otherwise ex-ante identical agents have noisy cognitive access to the full-information consumption policy c^* that would optimally imply a high saving decision around the constraint. For some agents, this noisy perception of that best course of action is to consume more aggressively than that. In turn, in the ergodic distribution these agents are over-represented in the left tail of the wealth distribution, and both display a high MPC and habitually high level of consumption, and thus persistently low wealth.[36]

[36]This argument also suggests that extending the model to allow for borrowing ($\underline{a} > 0$) would still imply

High MPCs of unconstrained wealthy agents

A standard approach to deliver our previous key properties, of a mass of agents that have low wealth perpetually, is to consider preference heterogeneity, and in particular heterogeneity in time-discount rates (e.g. Aguiar et al. (2020), Carroll et al. (2017)). As further emphasized by Krueger et al. (2016), such preference heterogeneity is usually seen as crucial for a model's ability to generate the resulting large mass of poor agents and, through it, empirically relevant aggregate MPCs that speak to the propagation of macroeconomic shocks and policy.

However, an issue often left unaddressed in this approach is that in the data the wealthy, unconstrained households also have surprisingly high MPC. The literature is careful to note that being "constrained" may be interpreted in terms of the overall net worth, as in the one asset economy studied here, or more specifically in terms of liquid wealth in a model that differentiates assets by liquidity, like in Kaplan et al. (2014). Still, a significant body of recent evidence, like Lewis et al. (2019), Fagereng et al. (2020), Gelman (2020), McDowall (2020), point out that even for agents with high liquid wealth the MPC level is significantly higher than implied by standard models. For example, in a model of discount-rate heterogeneity, the 'saver' (high β) types accumulate wealth and become the typical rich agents, but their stronger incentive to smooth consumption also leads, everything else constant, to lower MPCs when they are in a *non-HtM* circumstance, i.e. rich and highly liquid.[37]

In this context, our model is consistent with this puzzling evidence on high MPCs of rich and (liquidity) unconstrained agents, while at the same time delivering the systematically different target wealth levels and *HtM* behavior emphasized by the prior empirical literature. To showcase this unconstrained behavior, in panel (b) of Figure 5 we plot the typical policy function of the group of *non-HtM* agents. Overall, the resulting typical policy is foreshadowed by the example-agent S from Section 4.2. First, we can see that the typical policy estimate $\widehat{c}_{non-HtM}(x)$ is mostly below $c^*(y)$, consistent with the endogenous selection of these wealthy agents as under-estimating the latter. Through this perspective, at their typical ergodic cash-on-hand, rich agents consume less than implied by $c^*(y)$. Second, we can also see that the typical ergodic cash-on-hand of these wealthy agents is quite close to their long-run target wealth level $\bar{y}_i$, as implied by the fact that the typical policy estimate $\widehat{c}_{non-HtM}(x)$ is roughly centered around the intersection with $c^{RW}(y)$. This is to be expected by the fact that, as we

that a large mass of agents will sit near to that constraint, and thus continue to experience high MPCs. In contrast, in standard models allowing for borrowing tends to imply that agents move enough away from the constraint, so that even those with approximately zero assets (i.e. still measured as *HtM*), are characterized by significantly lower MPCs than in a model with $\underline{a} = 0$. See Carroll et al. (2017) and Aguiar et al. (2020) as examples of models that produce such lower MPCs for what otherwise look like *HtM* agents.

[37]More generally, recent state-of-the-art models relying on preference heterogeneity, such as Aguiar et al. (2020), while typically aimed and successful at accounting for *HtM* behavior, have a difficult time jointly explaining the high MPC of the (liquidity) unconstrained, rich agents.

saw from Table 1, two-thirds of the *non-HtM* agents are in a learning trap.

But as discussed earlier, consumption behavior around the individual steady-state wealth $\bar{y}_i$ is necessarily characterized by a MPC that is significantly higher than the slope of the $c^{RW}(y)$ function, which is equal to $r/(1+r)$. This is because $\bar{y}_i$ is indeed a stable, steady-state wealth level if and only if the consumption function is sufficiently steep in the neighborhood of $\bar{y}_i$, so that, for example, upon a positive income shock the agent consumes a large fraction of it, thus keeping wealth stable.

The upshot of this qualitative insight is that the average MPC within the group of *non-HtM* agents in our model is significantly higher than $r/(1+r) \approx 0.04$, as reported in panel (B) of Table 2. At our calibration, the average MPC of *non-HtM* agents is 0.15 which is almost four times larger than that, and consequently also four times larger than the average MPC of the full-information model (since the slope of $c^*(y)$ converges to $r/(1+r)$ for large y). Precise empirical estimates of the MPCs of wealthy agents are hard to obtain, but recent studies by McDowall (2020) and Gelman (2020) imply that the MPC of the wealthiest 20% of agents is around 0.17, very close to what our model implies (without it being targeted).

Aggregate MPC

Overall, our model successfully generates large, empirically relevant aggregate MPC thanks to both of the forces emphasized earlier – (1) the compositional effect of being able to generate a large (and persistent) fraction of *HtM* agents, and (2) the implied high MPC for the wealthy, unconstrained agents. Both of these channels matter for the average (or aggregate) MPC. In particular, our model delivers an average MPC of 0.29, which is well in-line with the empirical estimates – those range from 0.2 to 0.6 according to the review in Carroll et al. (2017). In contrast, in the FI version of the model the mean MPC is counter-factually low and just 0.05.

Decomposing the difference in the average MPC between our model and the FI benchmark $(0.29 - 0.05 = 0.24)$ in terms of the contribution of the two channels, we find that the second, namely the high MPCs of the rich agents, account for 35% of the difference. Hence, while our model shares the insight of Krueger et al. (2016) that the mass of poor agents is crucial for generating empirically relevant MPCs, our quantitative analysis shows that the surprisingly high MPCs of wealthy agents also contributes significantly.

5.3 Policy implications

In conclusion, we stress that in our model the pattern of "mistakes" is endogenous. One immediate exercise that highlights this endogenous nature is to compare our benchmark model to one where agents know c^* but make idiosyncratic "trembling hand" mistakes. We discuss this experiment in Appendix C. Its key conclusion is that while such mistakes generate micro volatility and heterogeneity, they tend to wash out over the long-run and are unable to generate the systematic differences in behavior obtained by our model in Section 5.2.

More fundamentally, our model also offers a novel mechanism through which observed behavior may change even if the underlying properties of the economy, otherwise sufficient to describe behavior under full-rationality, have not. We showcase below this property and its policy implications by contrasting the effects of a fiscal stimulus payment received (i) while the economy is at its stationary equilibrium, to a case where (ii) the stimulus is received at a time where additional information arrives that lowers agents' confidence in their previous reasoning. While the arrival of this information does not affect the counterfactual full-rational model's response to the stimulus, since there agents are assumed to have no doubt about $c^*(y)$, it matters in our model.[38] In particular, this sudden lack of confidence *reduces* the effectiveness of the stimulus policy, as it endogenously prompts agents to reason more intensely than otherwise, revise beliefs and thus abandon their "business-as-usual", high MPC behavior.

To illustrate these effects, in panel (a) of Figure 7 we plot the response of aggregate consumption to a one-time 1% increase in average income – corresponding, for example, to a lump-sum fiscal stimulus payment to all agents. We plot the response of aggregate consumption in our benchmark CR economy with the blue line, and the counter-factual response in the FI economy with a yellow line. As foreshadowed by the high MPCs our model delivers, there is a significant aggregate response, with aggregate consumption increasing by 0.25% on impact, as compared to just a 0.04% increase in the FI economy.[39]

Next we consider the effect of the same fiscal stimulus payment when received at a time when intangible information, for example reflecting the perception of a "new normal", leads agents to acknowledge that their previously accumulated reasoning about the optimal $c^*(y)$ is now obsolete.[40] The aggregate response of consumption to the stimulus in this case is roughly cut in half, as can be seen from the red line in panel (a) of Figure 7.

The fall in the effectiveness of the stimulus is a direct consequence of the endogenous nature of the behavioral "mistakes" of our agents. As agents choose now to reason more than otherwise, their current behavior puts more weight on the new reasoning signals, which are centered at c^*, while the destruction of past information resets the beginning-of-period beliefs to c^*. Hence, intuitively, the average behavior temporarily changes to resemble more

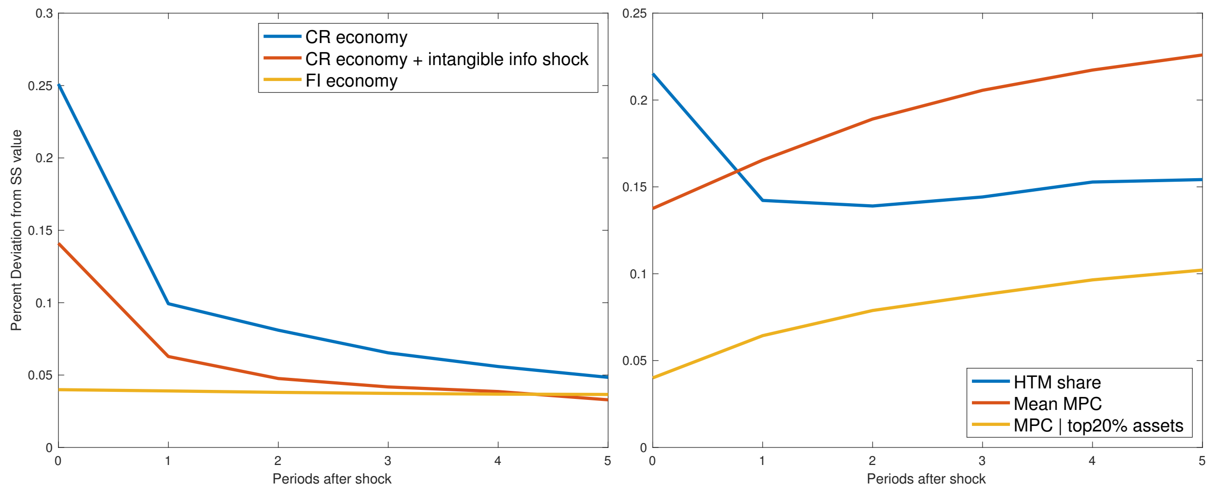
[38] Thus, this intangible information experiment isolates the effect of our model of bounded rationality, by keeping the agents' preferences, endowment process, and available financial and technological constraints constant, and thus the underlying full-information $c^*(y)$ as unchanged.

[39] Since there is no fundamental change to the structure of the economy, after this one-time shock it converges back to the same stationary distribution.

[40] The experiment is simple and stark - this shock resets all agents' beliefs back to the time-0 prior. A case where only some, but not all, previous reasoning information is rendered useless works qualitatively similar.

Figure 7: Response of economy following an intangible information shock

(a) Aggregate effect of stimulus (b) Marginal Propensity to Consume



closely that under c^*, with the differential effects being particularly large for the rich agents, as unconstrained households have steep policy functions either way.

As we see in panel (b) of Figure 7, both of the two key behavioral properties that we emphasized in Section 5.2 are now changing. First, there is a lower mass of *HtM* agents (blue line), as the renewed reasoning leads some to accumulate more savings. Second, this new, endogenous reasoning also leads to a significant behavioral change for the rich agents – for example the MPCs of the wealthiest 20% falls almost four times, to 0.04 (yellow line). Altogether, there is a sharp reduction in the aggregate MPC, which is essentially halved and falls to 0.15, as we can see in panel (b) of Figure 7 (red line).

Overall, this illustration highlights that our model may carry important lessons for the design of policy interventions. Thus, the model also points to caution, defining the classic Lucas (1976) critique, in not mechanically extrapolating agents' behavior, here imperfect due to bounded rationality, as fixed across possibly different environments.

6 Conclusion

This paper is motivated by a long-standing interest in the economics literature of relaxing the typically convenient, but otherwise extreme, assumption of decision-makers having free cognitive access to their optimal policy function. In this context, the first contribution of the paper is methodological in nature, as we propose a framework to model costly reasoning that is (i) tractable and portable across specific economic models, and (ii) well grounded in a broad neuroscience, experimental and computational literature. The second contribution is applied. We show that the proposed costly reasoning framework is a parsimonious and novel mechanism that generates rich and intuitive joint dynamics of beliefs and actions in a standard incomplete markets model. These dynamics can help bring the model closer to the data, and also hold important lessons for policy makers. In this context, our mechanism opens up the possibility to develop a concrete framework to test the role of costly reasoning in driving macroeconomic outcomes and policies. Recent work, such as DAcunto et al. (2019) who find field evidence for cognitive frictions in how agents map macroeconomic expectations to their implied optimal actions, is a promising such direction.

References

Aguiar, M., M. Bils, and C. Boar (2020): "Who are the Hand-to-Mouth?" NBER Working Paper No. 26643.

Aiyagari, S. R. (1994): "Uninsured idiosyncratic risk and aggregate saving," *The Quarterly Journal of Economics*, 109, 659–684.

Alaoui, L. and A. Penta (2016): "Cost-Benefit Analysis in Reasoning," Working Paper.

Ameriks, J., A. Caplin, and J. Leahy (2003): "Wealth accumulation and the propensity to plan," *The Quarterly Journal of Economics*, 118, 1007–1047.

Aragones, E., I. Gilboa, A. Postlewaite, and D. Schmeidler (2005): "Fact-Free Learning," *The American Economic Review*, 95, pp–1355.

Bach, L., L. E. Calvet, and P. Sodini (2017): "From saving comes having? Disentangling the impact of saving on wealth inequality," Swedish House of Finance Research Paper.

Ballinger, T. P. and N. T. Wilcox (1997): "Decisions, error and heterogeneity," *The Economic Journal*, 107, 1090–1105.

Bernheim, B. D., J. Skinner, and S. Weinberg (2001): "What accounts for the variation in retirement wealth among US households?" *American Economic Review*, 91, 832–857.

Bordalo, P., N. Gennaioli, and A. Shleifer (2020): "Memory, Attention, and Choice," *The Quarterly Journal of Economics*, 135, 1399–1442.

Bornstein, A. M., M. W. Khaw, D. Shohamy, and N. D. Daw (2017): "Reminders of past choices bias decisions for reward in humans," *Nature Communications*, 8, 1–9.

Bornstein, A. M. and K. A. Norman (2017): "Reinstated episodic context guides sampling-based decisions for reward," *Nature neuroscience*, 20, 997–1003.

Caplin, A., M. Dean, and D. Martin (2011): "Search and satisficing," *The American Economic Review*, 101, 2899–2922.

Carroll, C. (2004): “Theoretical foundations of buffer stock saving,” NBER WP 10867.

Carroll, C., J. Slacalek, K. Tokuoka, and M. N. White (2017): “The distribution of wealth and the marginal propensity to consume,” *Quantitative Economics*, 8, 977–1020.

Carvalho, L. and D. Silverman (2019): “Complexity and Sophistication,” NBER WP 26036.

Chase, W. G. and H. A. Simon (1973): “Perception in chess,” *Cognitive psychology*, 4, 55–81.

Conlisk, J. (1996): “Why bounded rationality?” *Journal of Economic Literature*, 34, 669–700.

De Nardi, M. and G. Fella (2017): “Saving and wealth inequality,” *Review of Economic Dynamics*, 26, 280–300.

Deck, C. and S. Jahedi (2015): “The effect of cognitive load on economic decision making: A survey and new experiments,” *European Economic Review*, 78, 97–119.

DellaVigna, S. (2009): “Psychology and economics: Evidence from the field,” *Journal of Economic literature*, 47, 315–72.

Dew-Becker, I. and C. G. Nathanson (2019): “Directed attention and nonparametric learning,” *Journal of Economic Theory*, 181, 461–496.

DAcunto, F., D. Hoang, M. Paloviita, and M. Weber (2019): “IQ, expectations, and choice,” *The Review of Economic Studies*, forthcoming.

Evans, J. S. B. (2003): “In two minds: dual-process accounts of reasoning,” *Trends in cognitive sciences*, 7, 454–459.

Fagereng, A., M. B. Holm, and G. J. J. Natvik (2020): “MPC heterogeneity and household balance sheets,” *AEJ: Macroeconomics*, forthcoming.

Gabaix, X. (2014): “A sparsity-based model of bounded rationality,” *The Quarterly Journal of Economics*, 129, 1661–1710.

——— (2019): “Behavioral inattention,” in *Handbook of Behavioral Economics: Applications and Foundations 1*, Elsevier, vol. 2, 261–343.

Gagnon-Bartsch, T., M. Rabin, and J. Schwartzstein (2020): “Channeled Attention and Stable Errors,” Harvard, mimeo.

Ganong, P. and P. Noel (2019): “Consumer spending during unemployment: Positive and normative implications,” *American Economic Review*, 109, 2383–2424.

Gelman, M. (2020): “What drives heterogeneity in the marginal propensity to consume? temporary shocks vs persistent characteristics,” *Journal of Monetary Economics*.

Gershman, S. J. and N. D. Daw (2017): "Reinforcement learning and episodic memory in humans and animals: an integrative framework," *Annual review of psychology*, 68, 101–128.

Gershman, S. J., E. J. Horvitz, and J. B. Tenenbaum (2015): "Computational rationality: A converging paradigm for intelligence in brains, minds, and machines," *Science*, 349, 273–278.

Gilboa, I. and D. Schmeidler (1995): "Case-based decision theory," *The Quarterly Journal of economics*, 110, 605–639.

——— (2001): *A theory of case-based decisions*, Cambridge University Press.

Girshick, A. R., M. S. Landy, and E. P. Simoncelli (2011): "Cardinal rules: visual orientation perception reflects knowledge of environmental statistics," *Nature neuroscience*, 14, 926–932.

Griffiths, T. L., F. Lieder, and N. D. Goodman (2015): "Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic," *Topics in cognitive science*, 7, 217–229.

Hendricks, L. (2007): "Retirement wealth and lifetime earnings," *International Economic Review*, 48, 421–456.

Hey, J. D. (2001): "Does repetition improve consistency?" *Experimental Economics*, 4, 5–54.

Hintikka, J. (1975): "Impossible Possible Worlds Vindicated," *Journal of Philosophical Logic*, 4, 475.

Ilut, C., R. Valchev, and N. Vincent (2020): "Paralyzed by Fear: Rigid and Discrete Pricing under Demand Uncertainty," *Econometrica*, 88, 1899–1938.

Johnson, E. J. and R. Ratcliff (2014): "Computational and process models of decision making in psychology and behavioral economics," in *Neuroeconomics: Decision Making and the Brain*, ed. by P. Glimcher and E. Fehr, New York: Academic Press, 35–48.

Kaelbling, L. P., M. L. Littman, and A. W. Moore (1996): "Reinforcement learning: A survey," *Journal of artificial intelligence research*, 4, 237–285.

Kahneman, D. (2011): *Thinking, fast and slow*, Macmillan.

Kalayci, K. and M. Serra-Garcia (2016): "Complexity and biases," *Experimental Economics*, 19, 31–50.

Kaplan, G. and G. L. Violante (2014): "A model of the consumption response to fiscal stimulus payments," *Econometrica*, 82, 1199–1239.

Kaplan, G., G. L. Violante, and J. Weidner (2014): "The wealthy hand-to-mouth," *Brookings papers on economic activity*, 77–153.

Klaes, M., E.-M. Sent, et al. (2005): "A conceptual history of the emergence of bounded rationality," *History of political economy*, 37, 27–59.

Klein, G. A., R. Calderwood, and A. Clinton-Cirocco (1986): "Rapid decision making on the fire ground," in *Proceedings of the human factors society annual meeting*, Sage Publications Sage CA: Los Angeles, CA, 576–580.

Kozlowski, J., L. Veldkamp, and V. Venkateswaran (2020): "The tail that wags the economy: Beliefs and persistent stagnation," *Journal of Political Economy*, 128.

Krueger, D., K. Mitman, and F. Perri (2016): "Macroeconomics and household heterogeneity," in *Handbook of Macroeconomics*, Elsevier, vol. 2, 843–921.

Kueng, L. (2018): "Excess sensitivity of high-income consumers," *The Quarterly Journal of Economics*, 133, 1693–1751.

Levy, W. B. and R. A. Baxter (2002): "Energy-efficient neuronal computation via quantal synaptic failures," *Journal of Neuroscience*, 22, 4746–4755.

Lewis, D. J., D. Melcangi, and L. Pilossoph (2019): "Latent heterogeneity in the marginal propensity to consume," FRB of New York Staff Report No. 902.

Lian, C. (2020): "Mistakes in future consumption, high MPCs now," MIT, Working Paper.

Lieder, F. and T. L. Griffiths (2020): "Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources," *Behavioral and Brain Sciences*, 43.

Lipman, B. L. (1991): "How to decide how to decide how to...: Modeling limited rationality," *Econometrica*, 1105–1125.

——— (1999): "Decision theory without logical omniscience: Toward an axiomatic framework for bounded rationality," *The Review of Economic Studies*, 66, 339–361.

Liu, W., J. C. Principe, and S. Haykin (2011): *Kernel adaptive filtering: a comprehensive introduction*, vol. 57, John Wiley & Sons.

Lucas, C. G., T. L. Griffiths, J. J. Williams, and M. L. Kalish (2015): "A rational model of function learning," *Psychonomic bulletin & review*, 22, 1193–1215.

Lucas, R. E. (1976): "Econometric policy evaluation: A critique," in *Carnegie-Rochester conference series on public policy*, vol. 1, 19–46.

Luo, Y. (2008): "Consumption dynamics under information processing constraints," *Review of Economic Dynamics*, 11, 366–385.

Mackowiak, B., F. Matejka, and M. Wiederholt (2020): "Rational inattention: A disciplined behavioral model," CEPR Discussion Paper No. 15408.

Maćkowiak, B. and M. Wiederholt (2009): "Optimal sticky prices under rational inattention," *The American Economic Review*, 99, 769–803.

——— (2015): "Business cycle dynamics under rational inattention," *The Review of Economic Studies*, 82, 1502–1532.

Manzini, P. and M. Mariotti (2014): "Stochastic choice and consideration sets," *Econometrica*, 82, 1153–1176.

Matějka, F. and A. McKay (2014): "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *The American Economic Review*, 105, 272–298.

McDowall, R. A. (2020): "Consumption behavior across the distribution of liquid assets," .

Mosteller, F. and P. Nogee (1951): "An experimental measurement of utility," *Journal of Political Economy*, 59, 371–404.

O'Donoghue, T. and M. Rabin (1999): "Doing it now or later," *American Economic Review*, 89, 103–124.

Olafsson, A. and M. Pagel (2018): "The liquid hand-to-mouth: Evidence from personal finance management software," *The Review of Financial Studies*, 31, 4398–4446.

Parker, J. A. (2017): "Why Don't Households Smooth Consumption? Evidence from a $25 Million Experiment," *American Economic Journal: Macroeconomics*, 9, 153–83.

Plonsky, O., K. Teodorescu, and I. Erev (2015): "Reliance on small samples, the wavy recency effect, and similarity-based learning." *Psychological review*, 122, 621.

Powell, W. B. (2007): *Approximate Dynamic Programming: Solving the curses of dimensionality*, vol. 703, John Wiley & Sons.

Rasmussen, C. E. and C. K. Williams (2006): *Gaussian processes for machine learning*, vol. 1, MIT press Cambridge.

Ratcliff, R. and G. McKoon (2008): "The diffusion decision model: theory and data for two-choice decision tasks," *Neural computation*, 20, 873–922.

Reis, R. (2006): "Inattentive consumers," *Journal of Monetary Economics*, 53, 1761–1800.

Rubinstein, A. (1998): *Modeling bounded rationality*, MIT press.

Shenhav, A., S. Musslick, F. Lieder, W. Kool, T. L. Griffiths, J. D. Cohen, and M. M. Botvinick (2017): "Toward a rational and mechanistic account of mental effort," *Annual review of neuroscience*, 40, 99–124.

Simon, H. A. (1955): "A behavioral model of rational choice," *The Quarterly Journal of Economics*, 69, 99–118.

——— (1956): "Rational choice and the structure of the environment." *Psychological review*, 63, 129.

——— (1976): "From substantive to procedural rationality," in *25 years of economic theory*, Springer, 65–86.

Sims, C. A. (1998): "Stickiness," in *Carnegie-Rochester Conference Series on Public Policy*, vol. 49, 317–356.

——— (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50, 665–690.

Sloman, S. A. (1996): "The empirical case for two systems of reasoning," *Psychological bulletin*, 119, 3.

Stanovich, K. E. and R. F. West (2000): "Individual differences in reasoning: Implications for the rationality debate?" *Behavioral and brain sciences*, 23, 645–665.

Stevens, L. (2020): "Coarse pricing policies," *The Review of Economic Studies*, 87, 420–453.

Todd, P. M. and G. Gigerenzer (2003): "Bounding rationality to the world," *Journal of Economic Psychology*, 24, 143–165.

Tulving, E. (1972): "Episodic and semantic memory," in *Organization of memory*, ed. by E. Tulving and W. Donaldson, New York: Academic, 381–403.

——— (2002): "Episodic memory: From mind to brain," *Annual review of psychology*, 53, 1–25.

Tversky, A. (1969): "Intransitivity of preferences," *Psychological review*, 76, 31.

Wei, X.-X. and A. A. Stocker (2015): "A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts," *Nature neuroscience*, 18, 1509.

Wiederholt, M. (2010): "Rational Inattention," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf and L. E. Blume, Palgrave Macmillan, vol. 4.

Woodford, M. (2003): "Imperfect Common Knowledge and the Effects of Monetary Policy," *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, 25.

——— (2014): "Stochastic choice: An optimizing neuroeconomic model," *The American Economic Review*, 104, 495–500.

——— (2019): "Modeling Imprecision in Perception, Valuation, and Choice," *Annual Review of Economics*, 12.

Zeldes, S. P. (1989): "Consumption and liquidity constraints: an empirical investigation," *Journal of Political Economy*, 97, 305–346.

Appendix

A Proofs

Lemma 1. *If $\theta_k \overset{iid}{\sim} N(\mu_k, \sigma_c^2)$, equation (1) implies that $c^* \sim \mathcal{GP}(\widehat{c}_0, \widehat{\sigma}_0)$, with*

$$\widehat{c}_0(y) = \sum_{k=1}^{N} \mu_k \phi_k(y); \qquad \widehat{\sigma}_0(y, y') = \sigma_c^2 \sum_{k=1}^{N} \phi_k(y)\phi_k(y')$$

Proof. Given that $\theta_k \overset{iid}{\sim} N(\mu_k, \sigma_c^2)$, it immediately follows that for any pair of real scalars, $y, y' \in \mathbb{R}$, the vector $\left[\sum_{k=1}^{N} \theta_k \phi_k(y), \sum_{k=1}^{N} \theta_k \phi_k(y')\right]'$ has the following joint Gaussian distribution:

$$\begin{bmatrix} \sum_{k=1}^{N} \theta_k \phi_k(y) \\ \sum_{k=1}^{N} \theta_k \phi_k(y') \end{bmatrix} \sim N\left(\begin{bmatrix} \sum_{k=1}^{N} \mu_k \phi_k(y) \\ \sum_{k=1}^{N} \mu_k \phi_k(y) \end{bmatrix}, \begin{bmatrix} \sigma_c^2 \sum_{k=1}^{N} (\phi_k(y))^2 & \sigma_c^2 \sum_{k=1}^{N} \phi_k(y)\phi_k(y') \\ \sigma_c^2 \sum_{k=1}^{N} \phi_k(y)\phi_k(y') & \sigma_c^2 \sum_{k=1}^{N} (\phi_k(y'))^2 \end{bmatrix} \right),$$

Assuming a complete set of basis functions $\{\phi_k\}_{k=1}^{N}$ that is big enough to achieve an arbitrarily good approximation of c^* so that

$$c^*(y) \approx \sum_{k=1}^{N} \theta_k \phi_k(y) \sim \mathcal{GP}(\widehat{c}_0, \widehat{\sigma}_0)$$

where

$$\widehat{c}_0(y) = \sum_{k=1}^{N} \mu_k \phi_k(y)$$

$$\widehat{\sigma}_0(y, y') = \sigma_c^2 \sum_{k=1}^{N} \phi_k(y)\phi_k(y')$$

$\square$

Lemma 2. *Given the time-0 prior belief $c^* \sim \mathcal{GP}(\widehat{c}_0, \widehat{\sigma}_0)$, conditional beliefs are given by $c^* \big| \{\eta^t, y^t\} \sim \mathcal{GP}(\widehat{c}_t, \widehat{\sigma}_t)$ with moments evolving according to the recursive expressions*

$$\widehat{c}_t(y) = \widehat{c}_{t-1}(y) + \frac{\widehat{\sigma}_{t-1}(y, y_t)}{\widehat{\sigma}_{t-1}^2(y_t) + \sigma_{\eta,t}^2}(\eta_t - \widehat{c}_{t-1}(y_t)), \tag{27}$$

$$\widehat{\sigma}_t(y, y') = \widehat{\sigma}_{t-1}(y, y') - \frac{\widehat{\sigma}_{t-1}(y, y_t)\widehat{\sigma}_{t-1}(y', y_t)}{\widehat{\sigma}_{t-1}^2(y_t) + \sigma_{\eta,t}^2} \tag{28}$$

where $\widehat{c}_t(y) \equiv E_t(c^*(y)|\eta^t)$ and $\widehat{\sigma}_t(y, y') \equiv \mathrm{Cov}(c^*(y), c^*(y')|\eta^t)$ are the posterior mean and covariance functions. Lastly, $\widehat{\sigma}_t^2(y) \equiv \widehat{\sigma}_t(y, y)$ denotes the posterior variance at a given y.

Proof. We prove this by way of induction. Consider the first update of beliefs at $t = 1$. By the definition of the Gaussian Process distribution, and the fact that $\eta_1 = c^*(y_1) + \varepsilon_1$ where ε_1 is Gaussian scalar and independent of the Gaussian Process c^*, it follows that for any $y, y' \in \mathbb{R}$

$$\begin{bmatrix} c^*(y) \\ c^*(y') \\ \eta_1 \end{bmatrix} \sim N\left(\begin{bmatrix} \widehat{c}_0(y) \\ \widehat{c}_0(y') \\ \widehat{c}_0(y_1) \end{bmatrix}, \begin{bmatrix} \widehat{\sigma}_0^2(y) & \widehat{\sigma}_0(y,y') & \widehat{\sigma}_0(y,y_1) \\ \widehat{\sigma}_0(y',y) & \widehat{\sigma}_0^2(y') & \widehat{\sigma}_0(y',y_1) \\ \widehat{\sigma}_0(y_1,y) & \widehat{\sigma}_0(y_1,y') & \widehat{\sigma}_0^2(y_1) + \sigma_{\eta,1}^2 \end{bmatrix} \right),$$

where we have used the short-hand notation $\sigma_0^2(y) \equiv \sigma_0(y,y)$.

By the standard property of multivariate Gaussian distributions (and given that y_1 is a known, deterministic scalar), the conditional distribution $\begin{bmatrix} c^*(y) \\ c^*(y') \end{bmatrix} \Big| \eta_1$ is also Gaussian:

$$\begin{bmatrix} c^*(y) \\ c^*(y') \end{bmatrix} \Bigg| \eta_1 \sim N\left(\begin{bmatrix} \widehat{c}_1(y) \\ \widehat{c}_1(y') \end{bmatrix}, \begin{bmatrix} \widehat{\sigma}_1^2(y) & \widehat{\sigma}_1(y,y') \\ \widehat{\sigma}_1(y',y) & \widehat{\sigma}_1^2(y') \end{bmatrix} \right),$$

where by standard Bayesian updating formulas

$$\begin{bmatrix} \widehat{c}_1(y) \\ \widehat{c}_1(y') \end{bmatrix} = \begin{bmatrix} \widehat{c}_0(y) \\ \widehat{c}_0(y') \end{bmatrix} + \begin{bmatrix} \widehat{\sigma}_0(y,y_1) \\ \widehat{\sigma}_0(y',y_1) \end{bmatrix} \frac{\eta_1 - \widehat{c}_0(y_1)}{\widehat{\sigma}_0^2(y_1) + \sigma_{\eta,1}^2}$$

$$\begin{bmatrix} \widehat{\sigma}_1^2(y) & \widehat{\sigma}_1(y,y') \\ \widehat{\sigma}_1(y',y) & \widehat{\sigma}_1^2(y') \end{bmatrix} = \begin{bmatrix} \widehat{\sigma}_0^2(y) & \widehat{\sigma}_0(y,y') \\ \widehat{\sigma}_0(y',y) & \widehat{\sigma}_0^2(y') \end{bmatrix} - \begin{bmatrix} \widehat{\sigma}_0(y,y_1) \\ \widehat{\sigma}_0(y',y_1) \end{bmatrix} \frac{1}{\widehat{\sigma}_0^2(y_1) + \sigma_{\eta,1}^2} \begin{bmatrix} \widehat{\sigma}_0(y,y_1) \\ \widehat{\sigma}_0(y',y_1) \end{bmatrix}'$$

which is simply equations (27) and (28) in matrix form, evaluated at $t = 1$. Thus, $c^*|\eta_1 \sim \mathcal{GP}(\widehat{c}_1, \widehat{\sigma}_1)$, where the functions $\widehat{c}_1$ and $\widehat{\sigma}_1$ are defined above. This confirms the result for $t = 1$.

For the induction step, assume that equations (27) and (28) hold for $t-1$ and that $c^*|\eta^{t-1} \sim \mathcal{GP}(\widehat{c}_{t-1}, \widehat{\sigma}_{t-1})$. Now consider the update at time t, again it follows that for any $y, y' \in \mathbb{R}$, the joint distribution $[c^*(y), c^*(y'), \eta_t]|\eta^{t-1}$ is Gaussian, with means given by the $\widehat{c}_{t-1}$ function and a variance-covariance matrix fully characterized by the $\widehat{\sigma}_{t-1}$ function. Then, by steps similar to those above

$$\begin{bmatrix} c^*(y) \\ c^*(y') \end{bmatrix} \Bigg| \eta^t \sim N\left(\begin{bmatrix} \widehat{c}_t(y) \\ \widehat{c}_t(y') \end{bmatrix}, \begin{bmatrix} \widehat{\sigma}_t^2(y) & \widehat{\sigma}_t(y,y') \\ \widehat{\sigma}_t(y',y) & \widehat{\sigma}_t^2(y') \end{bmatrix} \right),$$

where by the same standard Bayesian updating formulas for any $y, y' \in \mathbb{R}$

$$\begin{bmatrix} \widehat{c}_t(y) \\ \widehat{c}_t(y') \end{bmatrix} = \begin{bmatrix} \widehat{c}_{t-1}(y) \\ \widehat{c}_{t-1}(y') \end{bmatrix} + \begin{bmatrix} \widehat{\sigma}_{t-1}(y,y_t) \\ \widehat{\sigma}_{t-1}(y',y_t) \end{bmatrix} \frac{\eta_t - \widehat{c}_{t-1}(y_t)}{\widehat{\sigma}_{t-1}^2(y_t) + \sigma_{\eta,t}^2}$$

$$\begin{bmatrix} \widehat{\sigma}_t^2(y) & \widehat{\sigma}_t(y,y') \\ \widehat{\sigma}_t(y',y) & \widehat{\sigma}_t^2(y') \end{bmatrix} = \begin{bmatrix} \widehat{\sigma}_{t-1}^2(y) & \widehat{\sigma}_{t-1}(y,y') \\ \widehat{\sigma}_{t-1}(y',y) & \widehat{\sigma}_{t-1}^2(y') \end{bmatrix} - \begin{bmatrix} \widehat{\sigma}_{t-1}(y,y_t) \\ \widehat{\sigma}_{t-1}(y',y_t) \end{bmatrix} \frac{1}{\widehat{\sigma}_{t-1}^2(y_t) + \sigma_{\eta,t}^2} \begin{bmatrix} \widehat{\sigma}_{t-1}(y,y_t) \\ \widehat{\sigma}_{t-1}(y',y_t) \end{bmatrix}'$$

$\square$

Lemma 3. *Let* $\alpha_t(y) \equiv \frac{\widehat{\sigma}_{t-1}(y,y_t)}{\widehat{\sigma}^2_{t-1}(y_t)+\sigma^2_{\eta,t}}$ *be the weight put on* η_t *in the time-t estimate* $\widehat{c}_t(y)$.

- *If* $\psi = 0$, *then* $\alpha_t(y)$ *is just a constant – i.e.* $\alpha_t(y) = \alpha_t$ *for all* $y \in \mathbb{R}$, *and thus*

$$\widehat{c}_t(y) = \widehat{c}_0(y) + \sum_{k=1}^{t} \alpha_k \prod_{j=k+1}^{t} (1-\alpha_j) u_k$$

where $u_k = \eta_k - \widehat{c}_0(y_k)$ *is the deviation of signal* η_k *from the time-0 prior mean belief.*

- *If* $\psi > 0$, *then the informativeness of the signal* η_t *is state-dependent –* $\frac{\partial \alpha_t(y)}{\partial y} \neq 0$ *– and hence the shape of the time-t estimate* $\widehat{c}_t$ *differs from the time-0 prior, i.e.:*

$$\frac{\partial\left(\widehat{c}_t(y) - \widehat{c}_0(y)\right)}{\partial y} \neq 0$$

The effect of the information in η_t *is also local to* y_t, *since* $\lim_{||y-y_t||\to\infty} \alpha_t(y) = 0$

Proof. Consider the updated conditional estimate $\widehat{c}_t(y)$, where by Lemma 2 for any $y \in \mathbb{R}$:

$$\widehat{c}_t(y) = \widehat{c}_{t-1}(y) + \alpha_t(y)(\eta_t - \widehat{c}_{t-1}(y_t))$$

Iterating backwards, it follows that

$$\widehat{c}_t(y) = \widehat{c}_0(y) + \sum_{k=1}^{t} \alpha_k(y) \prod_{j=k+1}^{t} (1-\alpha_j(y)) u_k \tag{29}$$

where $u_k = \eta_k - \widehat{c}_0(y_k)$ is the deviation of signal η_k from the prior mean belief.

To prove the two parts of the Lemma, we will proceed by induction. First, if $\psi = 0$, for $t = 1$ we have

$$\alpha_1(y) = \frac{\widehat{\sigma}_0(y, y_1)}{\widehat{\sigma}^2_0(y_1) + \sigma^2_{\eta,1}} = \frac{\sigma^2_c}{\sigma^2_c + \sigma^2_{\eta,1}} \text{ for all } y$$

In this case, $\alpha_1(y) = \alpha_1$ is just a constant. Moreover, the updated covariance function is

$$\widehat{\sigma}_1(y, y') = \widehat{\sigma}_0(y, y') - \alpha_1 \widehat{\sigma}_0(y, y') = \widehat{\sigma}^2_1 = \sigma^2_c(1-\alpha_1)$$

which is again a constant independent of y and y'. Now consider the induction step; assuming that

$$\alpha_k(y) = \alpha_k \text{ for all } y \text{ and } k < t,$$

$$\widehat{\sigma}_{t-1}(y, y') = \widehat{\sigma}^2_{t-1} \equiv \sigma^2_c \prod_{k=1}^{t-1} (1-\alpha_k),$$

it follows that the effective signal-to-noise ratio for the time t signal is again a constant invariant to y:

$$\alpha_t(y) = \alpha_t \equiv \frac{\widehat{\sigma}_{t-1}(y, y_t)}{\widehat{\sigma}^2_{t-1}(y_t) + \sigma^2_{\eta,t}} = \frac{\widehat{\sigma}^2_{t-1}}{\widehat{\sigma}^2_{t-1} + \sigma^2_{\eta,t}}$$

Similarly, the resulting posterior variance at t is also invariant to y:

$$\widehat{\sigma}_t(y, y') = \widehat{\sigma}_{t-1} - \alpha_t \widehat{\sigma}_{t-1} = \widehat{\sigma}_t^2 = \widehat{\sigma}_{t-1}^2(1 - \alpha_t)$$

Hence, $\alpha_{t+1}(y)$ is also a constant invariant to y and so on. Thus, for any time t the effective signal-to-noise ratio is invariant to y, hence the conditional estimate is simply a constant shift away from the time-0 prior, with the value of that shift given by a weighted average of signal surprises:

$$\widehat{c}_t(y) = \widehat{c}_0(y) + \sum_{k=1}^{t} \alpha_k \prod_{j=k+1}^{t} (1 - \alpha_j) u_k$$

To prove the second part, when $\psi > 0$, we need to show that $\frac{\partial \alpha_t(y)}{\partial y} \neq 0$ almost everywhere and that $\lim_{||y-y'|| \to \infty} \alpha_t(y) = 0$. We will do both by induction, and the key is the evolution of the conditional covariance $\widehat{\sigma}_{t-1}(y, y')$. Starting with the case of $t = 1$, for any pair $y, y' \in \mathbb{R}$

$$\widehat{\sigma}_0(y, y') = \sigma_c^2 \exp(-\psi(y - y')^2)$$

which is decreasing in the distance $||y - y'||$, and $\frac{\partial \widehat{\sigma}_0(y,y')}{\partial y} \neq 0$ except for $y = y'$. The updated covariance function is

$$\widehat{\sigma}_1(y, y') = \sigma_c^2 \exp(-\psi(y - y')^2) - \frac{\sigma_c^4 \exp(-\psi\left((y - y_1)^2 + (y' - y_1)^2\right))}{\sigma_c^2 + \sigma_{\eta,1}^2}$$

hence similarly $\frac{\partial \widehat{\sigma}_1(y,y')}{\partial y} \neq 0$ outside of a measure 0 set and $\lim_{||y-y'|| \to \infty} \widehat{\sigma}_1(y, y') = 0$.

On the other hand, given that

$$\alpha_1(y) = \frac{\sigma_c^2 \exp(-\psi(y - y_1)^2)}{\sigma_c^2 + \sigma_{\eta,1}^2}$$

we clearly have $\frac{\partial \alpha_1(y)}{\partial y} \neq 0$ except for when $y = y_1$, which is measure 0. Also, $\lim_{||y-y_1|| \to \infty} \alpha_1(y) = 0$.

For the induction step, assume that $\frac{\partial \widehat{\sigma}_{t-1}(y,y')}{\partial y} \neq 0$ except for possibly on a set of measure 0, and that $\lim_{||y-y'|| \to \infty} \widehat{\sigma}_{t-1}(y, y') = 0$. The updated covariance function is

$$\widehat{\sigma}_t(y, y') = \widehat{\sigma}_{t-1}(y, y') - \frac{\widehat{\sigma}_{t-1}(y, y_t)\widehat{\sigma}_{t-1}(y', y_t)}{\widehat{\sigma}_{t-1}^2(y_t) + \sigma_{\eta,1}^2}$$

thus, $\frac{\partial \widehat{\sigma}_t(y,y')}{\partial y} \neq 0$ except for possibly on a set of measure 0, and $\lim_{||y-y'|| \to \infty} \widehat{\sigma}_t(y, y') = 0$.

Thus, for any arbitrary t, since $\alpha_t(y) = \frac{\sigma_{t-1}(y,y_t)}{\sigma_{t-1}^2(y_t) + \sigma_{\eta,t}^2}$,

$$\frac{\partial \alpha_t(y)}{\partial y} \neq 0$$

$$\lim_{||y-y_t|| \to \infty} \widehat{\alpha}_t(y) = 0$$

Lastly, by equation (29),

$$\widehat{c}_t(y) - \widehat{c}_0(y) = \sum_{k=1}^{t} \alpha_k(y) \prod_{j=k+1}^{t} (1 - \alpha_j(y)) u_k$$

and since $\frac{\partial \alpha_t(y)}{\partial y} \neq 0$, it follows that

$$\frac{\partial \left(\widehat{c}_t(y) - \widehat{c}_0(y)\right)}{\partial y} \neq 0$$

□

Proposition 1. *The optimal signal noise variance is given by*

$$\sigma^{*2}_{\eta,t} = \begin{cases} \frac{\kappa \widehat{\sigma}^2_{t-1}(y_t)}{\widehat{\sigma}^2_{t-1}(y_t) - \kappa} & , \text{ if } \widehat{\sigma}^2_{t-1}(y_t) \geq \kappa \\ \infty & , \text{ if } \widehat{\sigma}^2_{t-1}(y_t) < \kappa \end{cases}$$

and this in turn implies the time-t action

$$c_t = \widehat{c}_t(y_t) = \widehat{c}_{t-1}(y_t) + \alpha^*_t(y_t)(c^*(y_t) + \varepsilon_t - \widehat{c}_{t-1}(y_t)),$$

where the optimal weight put on the new reasoning signal, $\alpha^*_t(y_t)$ depends on the current state y_t and the history $\{y^{t-1}, \sigma^{t-1}_\eta\}$ of past signals' location and precision:

$$\alpha^*_t(y_t) \equiv \frac{\widehat{\sigma}^2_{t-1}(y_t)}{\widehat{\sigma}^2_{t-1}(y_t) + \sigma^{*2}_{\eta,t}} = \max\left[1 - \frac{\kappa}{\widehat{\sigma}^2_{t-1}(y_t)}, 0\right].$$

Proof. The agent's reasoning decision is governed by

$$\min_{\widehat{\sigma}^2_t(y_t)} \widehat{\sigma}^2_t(y_t) + \kappa \ln\left(\frac{\widehat{\sigma}^2_{t-1}(y_t)}{\widehat{\sigma}^2_t(y_t)}\right). \tag{30}$$

s.t.

$$\widehat{\sigma}^2_t(y_t) \leq \widehat{\sigma}^2_{t-1}(y_t),$$

The first-order condition implies that the optimal posterior variance choice is

$$\widehat{\sigma}^2_t(y_t) = \min\left[\kappa, \widehat{\sigma}^2_{t-1}(y_t)\right] \tag{31}$$

By Lemma 2,

$$\widehat{\sigma}^2_t(y_t) = \frac{\widehat{\sigma}^2_{t-1}(y_t)\sigma^2_{\eta,t}}{\widehat{\sigma}^2_{t-1}(y_t) + \sigma^2_{\eta,t}}$$

Using this expression and equation (31), we obtain the expression for the optimal reasoning noise variance $\sigma^2_{\eta,t}$.

Similarly, using the solution for $\sigma^2_{\eta,t}$, Lemma 2 and the fact that $\eta_t = c^*(y_t) + \varepsilon_t$, it follows directly that

$$c_t = \widehat{c}_t(y_t) = \widehat{c}_{t-1}(y_t) + \max\left[1 - \frac{\kappa}{\widehat{\sigma}^2_{t-1}(y_t)}, 0\right](c^*(y_t) + \varepsilon_t - \widehat{c}_{t-1}(y_t)),$$

□

Proposition 2. *The optimal reasoning intensity and the weight of the new signal in updating beliefs are both increasing in distance from location of the previous reasoning signal:*

$$\frac{\partial \sigma^2_{\eta,i,2}}{\partial ||y_{i,2} - y_1||} < 0 \text{ and } \frac{\partial \alpha_{i,2}(y_{i,2})}{\partial ||y_{i,2} - y_1||} > 0.$$

Therefore, agent C reasons more than agent S, i.e. $\alpha_{C,2}(y_{C,2}) > \alpha_{S,2}(y_{S,2})$, if and only if

$$s_2 < 1 + \frac{(1+r)}{w}(c^*(y_1) - c^{RW}(y_1)) \equiv \bar{s}$$

Proof. Consider the time-1 posterior variance function (expressed as a function of the distance $||y - y_1||$):

$$\widehat{\sigma}^2_1(y) = \sigma^2_c(1 - \frac{\sigma^2_c}{\sigma^2_c + \sigma^2_{\eta,1}} \exp\left(-2\psi||y - y_1||^2\right)$$

Note that this function is the same for both agents $i \in \{S, C\}$. Moreover,

$$\frac{\partial \widehat{\sigma}^2_1(y)}{\partial ||y - y_1||} = \frac{\sigma^4_c}{\sigma^2_c + \sigma^2_{\eta,1}} \exp\left(-2\psi||y - y_1||^2\right) 4\psi||y - y_1||^2 > 0$$

except for the knife-edge case $||y - y_1|| = 0$ when this derivative is zero. Using the expression for the optimal signal-noise variance $\sigma^2_{\eta,2}$ from Proposition 1:

$$\sigma^2_{\eta,i,2} = \frac{\kappa \widehat{\sigma}^2_1(y_{i,2})}{\widehat{\sigma}^2_1(y_{i,2}) - \kappa}$$

outside of the measure-zero case $y_{i,2} = y_1$. Then, it follows directly that

$$\frac{\partial \sigma^2_{\eta,i,2}}{\partial ||y_{i,2} - y_1||} = \frac{\kappa}{\widehat{\sigma}^2_1(y_{i,2}) - \kappa} \frac{\partial \widehat{\sigma}^2_1(y_{i,2})}{\partial ||y - y_1||} \underbrace{\left(1 - \frac{\widehat{\sigma}^2_1(y_{i,2})}{\widehat{\sigma}^2_1(y_{i,2}) - \kappa}\right)}_{<0} < 0$$

Similarly, outside of the measure-zero case $y_{i,2} = y_1$:

$$\alpha_{i,2}(y_{i,2}) = 1 - \frac{\kappa}{\widehat{\sigma}^2_{t-1}(y_{i,2})}$$

and thus

$$\frac{\partial \alpha_{i,2}(y_{i,2})}{\partial ||y - y_1||} = \frac{\kappa}{(\widehat{\sigma}^2_{t-1}(y_t))^2} \frac{\partial \widehat{\sigma}^2_1(y)}{\partial ||y - y_1||} > 0$$

For the last part of the proposition, note that

$$\alpha_{C,2}(y_{C,2}) - \alpha_{S,2}(y_{S,2}) = \frac{\kappa(\sigma_c^2 - \kappa)\left[\exp(-2\psi(y_{C,2} - y_1)^2) - \exp(-2\psi(y_{S,2} - y_1)^2)))\right]}{(\sigma_c^2 - (\sigma_c^2 - \kappa)\exp(-2\psi(y_1 - y_{C,2})^2))(\sigma_c^2 - (\sigma_c^2 - \kappa)\exp(-2\psi(y_1 - y_{S,2})^2))}$$

Hence, $\alpha_{C,2}(y_{C,2}) > \alpha_{S,2}(y_{S,2})$ if and only if

$$\exp(-2\psi(y_{C,2} - y_1)^2) > \exp(-2\psi(y_{S,2} - y_1^2))$$

$$\Longleftrightarrow$$

$$(y_{C,2} - y_1)^2 < (y_{S,2} - y_1)^2$$

substituting in the law of motion for cash-on-hand:

$$y_{i,2} = (1+r)(y_1 - \widehat{c}_{i,1}(y_1)) + ws_2 = (1+r)(y_1 - c^*(y_1) - \alpha_1\varepsilon_{i,1}) + ws_2$$

it follows that $\alpha_{C,2}(y_{C,2}) > \alpha_{S,2}(y_{S,2})$ if and only if

$$s_2 < \frac{(1+r)c^*(y_1) - ry_1}{w} + \frac{(1+r)\alpha_1}{2w}\frac{\varepsilon_{C,1}^2 - \varepsilon_{S,1}^2}{\varepsilon_{C,1} + |\varepsilon_{S,1}|}$$

and using the definition of $c^{RW}(y)$ and our assumption that $\varepsilon_{C,1} = -\varepsilon_{S,1}$, it follows that $\alpha_{C,2}(y_{C,2}) > \alpha_{S,2}(y_{S,2})$ if and only if

$$s_2 < 1 + \frac{(1+r)}{w}(c^*(y_1) - c^{RW}(y_1))$$

$\square$

B Calibration Strategy

To calibrate the learning-related parameters $\{\sigma_c^2, \psi, \kappa\}$ (given values for the rest of the parameters, which are discussed in Section **??**) we look for a fixed point, such that (i) an econometrician trying to estimate the distribution of the reasoning signals $\eta_{i,t}$ would indeed recover the actual values of $\{\sigma_c^2, \psi\}$ used for the simulation and (ii) the model implies zero net-assets for the bottom 20% of the wealth distribution (matching the PSID data).

We are motivated to look for a fixed point in $\{\sigma_c^2, \psi\}$ in order to ensure that agents have model consistent priors, in the sense that the prior beliefs properly capture the features of the true $c^*(y)$ policy function. Intuitively, we would like to parameterize the prior in an "efficient" way so that it does not imply any ex-ante biases that impede the estimation process.

To do so, we consider the problem of an agnostic econometrician who attempts to estimate the function $c^*(y)$ out of a data set that constitutes the reasoning signals $\eta_{i,t}$ from the simulation of our model. The econometrician uses the same Bayesian methods as the agents in our model, and has a Gaussian Process prior over the unknown function c^*. The econometrician is "agnostic", however, in the sense that instead of treating the prior distribution of c^* as a primitive, he treats the prior as a "hyper-parameter" that is optimized over during the estimation procedure. In this way, the econometrician looks for the "optimal" or most efficient prior for the data at hand (which is the simulated reasoning signals from the ergodic distribution of the model).

In our calibration strategy, we want to ensure that the agents in our model indeed hold those same "optimal" priors. To do so, we look for a fixed point between the assumed parameterization of the agents' priors, i.e. $\{\sigma_c^2, \psi\}$, and the values for the prior hyper-parameters the agnostic econometrician recovers. – which we label $\{\tilde{\sigma}_c^{*,2}, \tilde{\psi}^*\}$.

In particular, this econometrician

1. is given the history of reasoning signals η_i^t for all agents in the simulated economy as a data set which we denote $\boldsymbol{\eta}$

2. is aware of the structure of the signals, i.e. that

$$\eta_{i,t} = c^*(y_{i,t}) + \varepsilon_{i,t} \ , \ \varepsilon_{i,t} \sim N(0, \sigma^2_{\eta,i,t}).$$

3. Observes $y_{i,t}$ and $\sigma^2_{\eta,i,t}$, which we collect in the observed vector $\mathbf{y}$, but is uncertain about the function $c^*(y)$ and holds the following Gaussian Process prior over it:

$$c^* \sim \mathcal{GP}(0, \tilde{\sigma}_0)$$

As is standard practice in Bayesian statistics, the econometrician assumes that the prior mean is the constant zero function – in this sense, he is "agnostic" about the particular functional form of c^*. The econometrician's prior is also parameterized by a squared exponential function, just as the agents in our model:

$$\tilde{\sigma}_0(y, y') = \tilde{\sigma}_c^2 \exp(-\tilde{\psi}(y - y'^2)$$

To differentiate with the agents' covariance function, we label the parameters of the econometrician's covariance function with tildes.

4. Given the collection of reasoning signals $\boldsymbol{\eta}$ the econometrician forms the posterior distribution $c^*|\boldsymbol{\eta}$, and finds the optimal hyper-parameters $\{\tilde{\sigma}_c^{*,2}, \tilde{\psi}^*\}$ by maximizing the resulting marginal likelihood of the data (as a function of $\{\tilde{\sigma}_c^2, \tilde{\psi}\}$):

$$\max_{\tilde{\sigma}_c^2, \tilde{\psi}} p(\boldsymbol{\eta}|\mathbf{y}, \tilde{\sigma}_c^2, \tilde{\psi}) = -\frac{1}{2}\mathbf{y}' K_\eta^1 \mathbf{y} - \frac{1}{2}\ln(K_\eta) - \frac{n}{2}\ln(2\pi)$$

where K_η is the covariance matrix of the econometrician's data vector $\boldsymbol{\eta}$, with (i, j) element

$$K_\eta(i, j) = \tilde{\sigma}_c^2 \exp\left(-\tilde{\psi}(\boldsymbol{y}(i) - \boldsymbol{y}(j))^2\right) + \mathbb{1}(i = j)\boldsymbol{\sigma}_\eta^2(i)$$

Note: since c^* has a Gaussian Process distribution with a squared exponential covariance function, the covariance between two data points $\boldsymbol{\eta}(i)$ and $\boldsymbol{\eta}(j)$ depends on the position of the y state values at which the two respective η signals are observed. The diagonal entries of K_η are also affected by the variance of the idiosyncratic reasoning noise $\sigma^2_{\eta,i,t}$, which the econometrician observes and takes into account.

5. This results in maximized values of the prior hyper-parameters $\tilde{\sigma}_c^2, \tilde{\psi}$:

$$\{\tilde{\sigma}_c^{*,2}, \tilde{\psi}^*\} = \arg\max_{\tilde{\sigma}_c^2, \tilde{\psi}} p(\boldsymbol{\eta}|\mathbf{y}, \tilde{\sigma}_c^2, \tilde{\psi})$$

Thus, for any given simulation of our model we can obtain the agnostic econometrician's inferred values of σ_c^2 and ψ by following steps 1-5 above. In addition, we can then vary the parameter κ, which controls the extent or magnitude of the reasoning friction, in order to hit the additional target of zero net-wealth for the poorest 20% of agents at the ergodic steady state of the model.

We use the following numerical strategy to find the necessary fixed-point in $\{\sigma_c^2, \psi, \kappa\}$:

(a) Given an initial guess for $\{\sigma_c^2, \psi, \kappa\}$ (taking rest of the parameters as given), we simulate the model using the benchmark simulation size of $T = 10,000, N = 5,000$.

(b) We discard the first half of the simulated time-series and are left with a 5000x5000 panel data set $\boldsymbol{\eta}$.

(c) This is a very large dataset, so to speed up the hyper-parameter estimation outlined above (and thus make our fixed point search feasible), we select a random samples of length $\frac{1}{\theta}$ (i.e. an average life-cycle of information) out of the full dataset $\boldsymbol{\eta}$ and perform steps 1-5 above on each of those samples.

    We repeat, with replacement, 500 times and then take the average of the resulting 500 pairs of estimated hyper-parameters which we call $\{\bar{\sigma}_c^{*,2}, \bar{\psi}^*\}$.

(d) We check whether we have achieved a fixed point defined as satisfying both

    (i) $||\{\sigma_c^2, \psi\} - \{\bar{\sigma}_c^{*,2}, \bar{\psi}^*\}|| < 1e - 5$

    (ii) Share of total assets of bottom 20% of agents $< 1e - 5$

(e) if the two conditions above are satisfied we stop and use those coefficients $\{\sigma_c^2, \psi, \kappa\}$. If not, we update the guess of the parameter values as needed and go back to step (a).

C Costly reasoning vs simple mistakes

We discuss here a counter-factual model where agents have full-information about the optimal policy function but make idiosyncratic mistakes in their actions. Namely, we consider that agents are suffering from a simple "trembling-hand" kind of control problem, where they set an approximately accurate action that is contaminated with i.i.d. noise:

$$c_{i,t}^{trmb} = c^*(y_{i,t}) + \sigma_\tau^2 \varepsilon_{i,t}. \tag{32}$$

When simulating this model, we use the exact same sequence of noise shocks $\varepsilon_{i,t}$ that affect the reasoning signals in the costly-reasoning model. Hence, the stochastic choice of agents in this alternative model is driven by the same source of exogenous disturbances that generates contemporaneous dispersion in our benchmark model. We calibrate the standard deviation of these shocks, $\sigma_\tau = 0.18$, so as to match the dispersion of actions around the full-information action in the benchmark CR model – i.e. to match $Var(c_{i,t} - c^*(y_{i,t}))$.

Moments	Data	Benchmark	Full info	Trembles
	(1)	(2)	(3)	(4)
(A) Hand-to-Mouth (HtM)				
Fraction of Hand-to-Mouth	0.23	0.23	0.01	0.02
β_1 univariate regression	0	0.001	0.015	0.028
β_2 multivariate regression	0.025	0.035	0.015	0.030
γ multivariate regression	-0.038	-0.048	-0.015	-0.030
(B) MPC				
Average (aggregate) MPC	0.2-0.6	0.29	0.05	0.06
Average MPC | top 20% of assets	0.17	0.15	0.04	0.04
Average MPC | *non-HtM*		0.15	0.04	0.04
Average MPC | *HtM*		0.83	0.08	0.04

Table C.1: We report moments from data in column (1) and the stationary distribution in our benchmark costly reasoning model in column (2). In our counterfactuals we keep parameters at their benchmark values but set $\sigma_c^2 = 0$ in column (3), to construct the full-information counter-part of our model. The data moments in Panel (A) are documented by Aguiar et al. (2020) by utilizing the PSID panel structure. In Panel (B) the range of credible estimates of the aggregate MPC is from Carroll et al. (2017), and estimate of the MPC of the rich from McDowall (2020).

In this counterfactual "trembles" model, with moments reported in column (4) of Table C.1, agents similarly display stochastic choice and make mistakes in their actions, as in the CR model. This creates some additional wealth heterogeneity compared to the FI model, but that is quantitatively negligible – the share of assets held by the top 20% of agents increases from 44% to 46%, and the fraction of HtM agents increases from 1% to 2%. Overall, the Gini coefficient remains the same at 0.39.

Similarly, while the "trembles" generate consumption volatility, on average an agent behaves as under c^*, hence the MPCs for the unconstrained agents are also low. There is a slight increase in the average MPC (0.06 vs 0.05), but that is completely due to the compositional effect of more HtM agents – the MPC of the unconstrained agents are equivalent in the "trembles" and the full-information models. This showcases that the endogeneity in the reasoning intensity choice and the implied selection effect we have discussed in Section 4.2 are crucial for obtaining the systematically higher MPCs in our benchmark model.

The key to these results is that the errors in this counter-factual are not systematic, hence they tend to wash out over the long-run. For example, an agent might under-consume a few times, but he is not likely to become a "saver" type that persistently under-consumes, like it is possible in the CR model due to the endogenous, state-dependent choice of reasoning intensity. This is exemplified by the expected consumption growth of HtM agents in the "trembles" version of the model, which is even higher than under FI (2.8% vs 1.5%), underscoring the strong mean-reversion in actions in this counter-factual model.

The fact that the ergodic implications of this "trembles" counter-factual are similar to the full-information model illustrates a standard justification in the literature that the latter model may still be a good approximation for an underlying model where agents do end up making errors. In contrast, our mechanism shows that when behavioral "mistakes" are modeled in a "resource rational" way, the joint-distribution of beliefs and actions differ qualitatively and quantitatively from its full-information version.