CESifo WORKING PAPERS

10327
2023

March 2023

Complexity and Time

Benjamin Enke, Thomas Graeber, Ryan Oprea

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10327*

Complexity and Time

Abstract

We provide experimental evidence that core intertemporal choice anomalies – including extreme short-run impatience, structural estimates of present bias, hyperbolicity and transitivity violations – are driven by complexity rather than time or risk preferences. First, all anomalies also arise in structurally similar atemporal decision problems involving valuation of iteratively discounted (but immediately paid) rewards. These computational errors are strongly predictive of intertemporal decisions. Second, intertemporal choice anomalies are highly correlated with indices of complexity responses including cognitive uncertainty and choice inconsistency. We show that model misspecification resulting from ignoring behavioral responses to complexity severely inflates structural estimates of present bias.

JEL-Codes: C910, D910, G000.

Keywords: complexity, hyperbolic discounting, present bias, bounded rationality, noise, cognitive uncertainty.

Benjamin Enke
Economics Department
Harvard University
Cambridge / MA / USA
enke@fas.harvard.edu

Thomas Graeber
Harvard Business School
Boston / MA / USA
tgraeber@hbs.edu

Ryan Oprea
Economics Department, University of California, Santa Barbara / CA / USA
roprea@gmail.com

March 10, 2023
This paper supersedes Enke's and Graeber's earlier working paper "Noisy cognition and intertemporal choice." This research was supported by the National Science Foundation under Grant SES-1949366 and was approved by Harvard and UC Santa Barbara IRB.

1 Introduction

In this paper we provide evidence that many of the core behavioral anomalies in intertemporal choice over monetary rewards – including extreme short-run impatience, hyperbolic discounting and structural estimates of present bias – are primarily driven by complexity rather than non-standard preferences or self-control problems.

As summarized by Cohen et al. (2020) and Ericson and Laibson (2019), decision makers typically depart from the predictions of the exponential discounted utility model in a number of systematic ways when asked to value time-dated payments. The most influential of these anomalies concern the hyperbolic shape of the empirical discount function. First, experimental subjects show evidence of extreme short-run impatience, discounting relatively short time intervals at a very high rate. Second, people's revealed per-period impatience strongly decreases as the length of evaluated intervals increases. Third, in structural estimations, the combination of extremely high short-run impatience and decreasing impatience is often attributed to a present bias parameter, as in $\beta - \delta$ models (Laibson, 1997; O'Donoghue and Rabin, 1999). Finally, decision makers show evidence of a front-end delay effect, decreasing their impatience when earlier and later dates are both shifted into the future by an equal delay, violating stationarity.

What produces these systematic anomalies? The most influential explanations in the literature are *motivational* in nature, rooted in non-standard preferences or internal conflicts that appear because of temporal delay in the choice problem. For instance, the literature has proposed that such anomalies are driven by non-standard discounting functions such as hyperbolic or quasi-hyperbolic preferences (e.g., Loewenstein and Prelec, 1992; Laibson, 1997), short-term temptation (Gul and Pesendorfer, 2001), or self-control problems formalized as multiple selves (Fudenberg and Levine, 2006). Other motivational explanations emphasize that the passage of time produces transaction costs or risks that the rewards being valued will not be paid out (Halevy, 2008, 2015; Andreoni and Sprenger, 2012; Chakraborty et al., 2020).

Importantly, however, intertemporal choice is not only temporal (and thus shaped by motivational factors like preferences) – it is also *complex*, requiring decision makers to engage in non-trivial cognitive information processing to integrate the inputs of a problem (e.g., payments and delays) into an output (a valuation or choice). In particular, rational theories of intertemporal choice posit that decision makers *iteratively discount*: they evaluate rewards delayed by a one-period interval by shrinking them at some rate, repeating successively for each additional period. If precisely engaging in this cognitive process is difficult and costly, decision makers may elect to (or be forced to) pursue less careful approaches to valuation instead. We will refer to such substitutions from complex, optimal valuation procedures to simpler alternatives as "complexity responses."

The theoretical literature has proposed a number of models that characterize such simpler-than-optimal procedures, ranging from heuristics (e.g., Rubinstein, 2003; Ericson et al., 2015), to noisy introspection about one's discount factor (Regenwetter et al., 2018; Lu and Saito, 2018; He et al., 2019), to cognitive noise models in which decision makers combine imprecise mental representations with a prior (Vieider, 2021; Gabaix and Laibson, 2022; Gershman and Bhui, 2020). Importantly, what these simplified procedures have in common is that they tend to produce insensitivity to differences in the length of time intervals. Heuristically speaking, in many of these models, decision makers act as if they treat different time intervals alike to some degree. Because inelasticity of the discount function (relative to the exponential benchmark) is the characteristic feature of hyperbolicity, models describing such simplified procedures are often capable of generating classical anomalies such as exaggerated short-term discounting and decreasing impatience without appeal to preferences or self-control problems. Yet, while there is some evidence that bounded rationality influences intertemporal decisions (see references below), we lack a solid understanding of whether motivational or complexity-based mechanisms are responsible for these famous empirical anomalies.

Removing scope for motivational explanations. A key idea in our paper is that we can identify the role complexity plays in intertemporal choice anomalies by entirely removing scope for motivational explanations while retaining much of what makes intertemporal choice complex. We do this by removing actual time delay from standard elicitations and inducing an exponential reward function instead. Applying the design idea of Oprea (2022) to intertemporal choice, we ask experimental subjects to value not only intertemporal payments but also "atemporal mirrors" of the same payments – immediate, deterministic payoffs that are deliberately described in such a way as to require the same kind of discounting and a similar degree of information processing as the original intertemporal choice problem. Thus, in addition to asking subjects to value a dollar amount x paid in t periods,[1] we also ask subjects to value an *immediate* payment of x, iteratively "shrunk" t times, each time by a fixed factor, δ. Here, the number of iterations t is a direct analog to the length of the time interval. An atemporal mirror is, hence, simply a standard intertemporal choice in which we *experimentally induce* an exponential time preference, so that anomalous deviations from exponential discounting cannot be rationalized by non-standard time preferences, temptation, self-control or risk. However, doing this retains much of the complexity of iteratively discounting

[1] As summarized by Cohen et al. (2020), a key discussion is whether the fungibility of money prevents the elicitation of true time preferences. This discussion is less relevant to our paper because we are not attempting to measure preferences. Indeed, a central conclusion of ours is that even putting fungibility concerns aside, intertemporal choice experiments to a great extent fail to recover preferences due to the confounding influence of complexity.

future rewards present in the original choice problem.

By comparing the way people value delayed payments to the way they value similarly complexly-described immediate payments, we can gauge to what degree complexity alone is capable of producing intertemporal choice anomalies. We do not expect subjects to precisely calculate discounted values in these atemporal mirrors any more than we expect them to pull out a calculator to inform decisions in actual intertemporal choice. Rather, our interest is in what heuristic or noisy procedures people substitute to instead of precise evaluation, and whether these substitutions produce classical anomalies.

We find that hyperbolic discounting also strongly arises in the valuation of atemporal mirrors, and to a similar degree as in true intertemporal choice. Subjects discount atemporal payments more than the experimentally induced discount factor δ specifies when payments only need to be discounted one or two times, which directly mirrors the extreme short-run impatience that arises in intertemporal choice. At the same time, subjects discount atemporal payments that need to be discounted a large number of times considerably less strongly than they should, given the induced discount factor. Thus, even though we experimentally induce a fixed discount factor, subjects' revealed per-period "impatience" strongly decreases in the number of discounting steps, replicating the hyperbolicity widely observed in intertemporal choice experiments.

Importantly, because we measure each subject's behavior in both atemporal mirrors and true intertemporal choice, we can show that behavior in the former *predicts* behavior in the latter. We find correlations across the two choice problems of 0.34, suggesting that behaviors in the two settings are likely to a great extent driven by a common behavioral mechanism – which cannot involve temporal motivations that are absent in atemporal mirrors.

Measuring and manipulating complexity responses. To provide further evidence on the link between complexity and the anomalies, we study standard intertemporal decisions and directly trace behavioral signatures of complexity responses (evidence that subjects are using simpler-than-optimal procedures). We then examine whether these signatures predict the severity of anomalies. We measure two such signatures.

First, to the degree decision makers are aware that they have substituted to a simpler-than-optimal procedure, we expect them to express less confidence in the quality of their choices. Thus, adapting the methodology of Enke and Graeber (2022) to an intertemporal choice context, we elicit subjects' cognitive uncertainty about each decision, asking them how certain they are (in percentage terms) that their stated range of valuations for a delayed payment actually contains their true valuation for that payment. Second, many recent complexity-inspired explanations root intertemporal choice anomalies in decision or cognitive noise. To the degree decision makers use such noisy decision procedures, we expect them to make inconsistent, noisy decisions. To measure this, we gather data on choice inconsistencies in repeated instances of the same decision problems.

Both cognitive uncertainty and choice inconsistency are strongly correlated with short-run impatience and hyperbolic discounting. Subjects who are more inconsistent or more cognitively uncertain are less patient over short horizons but more patient over long horizons – which matches exactly the results from the atemporal mirror experiments in which decisions are too "impatient" over few iterations but too "patient" over many ones. These signatures of complexity responses even predict the strength of anomalies within-subject: a given subject exhibits more pronounced hyperbolicity in precisely those decisions in which s/he expresses high cognitive uncertainty. Overall, our data suggest that 85% of decreasing impatience in our setting is driven by complexity responses.

To provide causal evidence that cognitive uncertainty, choice inconsistency and intertemporal choice anomalies are indeed jointly sensitive to complexity, we also study experimental treatments that attempt to increase the cognitive costs of evaluating rewards and delays. We find that increasing complexity in this way leads to a joint increase in cognitive uncertainty, choice inconsistency and anomalies, again confirming the link between complexity, the use of imperfect decision procedures and anomalies.

Complexity and estimates of present bias. Perhaps the most commonly-discussed explanation for hyperbolic discounting in the literature is *present bias*, as in the quasi-hyperbolic $\beta-\delta$ model. Present bias is a tendency to discontinuously overweight rewards received immediately relative to rewards received with any delay. Structural estimates of present bias are usually identified from the hyperbolic shape of the empirical discount function. Given our finding that hyperbolicity is largely an outgrowth of complexity, a natural question is to what degree estimates of $\beta < 1$ are also a result of complexity responses.

In the atemporal mirrors experiment, where there is no true passage of time and exponential discounting is experimentally induced, we structurally estimate $\hat{\beta} = 0.85$ – an estimate that would conventionally be interpreted as strong evidence of present bias. Similarly, in our intertemporal choice experiments, we find that subject-level estimates of present bias are strongly correlated with cognitive uncertainty and choice inconsistencies, even though from the perspective of standard preferences-based models there is no reason to expect such a correlation. These results are robust to accounting for utility curvature. This evidence suggests that structural estimates of $\beta < 1$ are in large part driven by complexity responses.

Yet, as is well-understood in the literature (Cohen et al., 2020), *structural* estimates of β only measure "true" present bias under strong structural assumptions. More directly, *causal* estimates of present bias can be derived using front-end delay designs, in which subjects violate stationarity by acting in more patient ways when both earlier and later payments are moved into the future to the same degree. In our experiments, we consistently find these front-end delay effects. Yet, we never find any indication that these effects are driven by complexity: (i) cognitive uncertainty is uncorrelated with these effects; (ii) the complexity manipulation does not amplify them; and (iii) no front-end delay effects appear in the atemporal mirrors. Thus, *causal, front-end delay* estimates of present bias do not appear to be confounded by complexity, but *structural* estimates that rely on the hyperbolic shape of discounting severely inflate the degree of true present bias.

What accounts for this discrepancy? Intuitively, it is a consequence of the fact that the empirical discount function is substantially more hyperbolic than the magnitude of non-stationarities in front-end delay designs would imply (see Cohen et al., 2020). This suggests that, in tradeoffs over money, "true" present bias (stationarity violations) is only a small part of the reason why researchers estimate large degrees of present bias in structural exercises.[2] Instead, our evidence suggests that structural estimates of present bias are severely inflated due to model misspecification from omiting complexity responses: estimated present bias picks up both true present bias and complexity-driven hyperbolicity.

Mechanism: Complexity and insensitivity. The evidence on front-end delay effects strongly suggests that stationarity violations – while present – are unrelated to complexity and explain only a fraction of the hyperbolicity of discounting. This raises the question: through what other proximal mechanism does complexity generate hyperbolicity? Motivated by the extant literature (Read, 2001), we investigate the idea that complexity generates a generic *insensitivity to the length of time intervals*, which produces an implied per-period discount rate that decreases as the length of intervals increases. This is akin to complexity-driven insensitivities that have been observed in other domains (Abeler and Jäger, 2015; Enke and Graeber, 2022; Oprea, 2022).

Evidence of insensitivity to the length of time intervals is usually gathered using *subadditivity* designs. These involve a particularly stark documentation of insensitivity in which subjects violate transitivity by making less patient choices when a single time interval is decomposed into two separate intervals. In our data, we consistently find that complexity is linked to such subadditivity effects: (i) they are are strongly correlated with cognitive uncertainty; (ii) they increase in the complexity manipulation; and (iii) they arise with equal strength in the atemporal mirrors. Across datasets, we estimate that 100% of subadditivity effects reflect complexity responses. We conclude that most

[2]Though, as we emphasize below, "true" present bias as measured by front-end delay effects is substantially larger in exercises measures primary rewards rather than financial flows.

of hyperbolic discounting in our context is a consequence of complexity-driven *insensitivity* to variation across time intervals. This is reminiscent of the mechanism described by most theoretical models of complexity-driven intertemporal choice anomalies.

Summary. Taken together, over a number of distinct empirical approaches – including atemporal mirrors, objective decision noise, cognitive uncertainty and experimental complexity manipulations – we document the same story. Many famous anomalies in intertemporal decisions over money occur because people respond to the complexity of intertemporal choice by using imperfect decision procedures that are *insufficiently sensitive* to variation in time intervals. This inelasticity, in turn, gives rise to extreme short-run impatience, hyperbolicity and subadditivity effects, and severely inflates structural estimates of present bias. These results *do not* suggest that non-standard time preferences do not exist – our results show that complexity effects are distinct from causally-identified present bias (stationarity violations). However, our results *do* suggest that, in experimental contexts like ours, complexity is the main source of hyperbolic discounting and structural estimates of present bias, both of which are often mis-attributed to non-standard preferences.

Our findings have several implications. First, they have welfare and policy implications: to the degree anomalies are complexity-driven mistakes rather than expressions of underlying motivations, models of non-standard discounting behavior, though no less descriptively valuable, lose much of their normative bite. Second, our results suggest that we can likely improve prediction, and design better policy by modeling intertemporal choice as dependent upon complexity, rather than fixed discount functions. For instance, our results illuminate how complexity responses confound the structural estimation of present bias, one of the central objects measured in behavioral economics. This suggests that in some of those field contexts in which present bias has been identified through model estimations, there may be scope for "improving" behavior by reducing complexity. Third, our results shed light on the puzzle that demand for commitment is typically considerably weaker than models of present bias would imply (Laibson, 2015) – the reason is that part of estimated present bias reflects complexity responses rather than temptation or a desire for immediate gratification. Fourth, results like ours suggest that traditional methods of recovering preferences, which often rely heavily on the idea that preferences are transparently revealed in choice, may produce misleading results because choice often primarily reflects a complexity response rather than a clean revelation of preferences. This speaks to a widely-known puzzle in lab-to-field studies, in which correlations between lab measures of time preferences and ecological behaviors such as savings behavior – while often statistically significant – are typically quantitatively extremely small (e.g., Chabris et al., 2008; Falk et al., 2018).

Our paper is organized as follows. In Section 2 we review common anomalies in intertemporal choice, explanations offered by the literature and discuss our empirical strategies for separating them. Section 3 presents our experimental design. Sections 4–6 present the results. Section 7 discusses related literature and concludes.

2 Conceptual Background and Empirical Approaches

2.1 Anomalies in Intertemporal Choice

Consider a simple intertemporal choice problem $D = (x_1, t_1; x_2, t_2)$ in which a decision maker must decide what x_1 paid at t_1 (e.g., now) makes her indifferent to earning x_2 paid at $t_2 > t_1$ (e.g., in two months). Define $\Delta t \equiv t_2 - t_1$ (all time units are in months). In the exponential discounting model (the standard benchmark in the economics literature), the decision maker discounts rewards exponentially with an annual discount factor $\delta = 1 - \gamma$, where γ is approximately the constant discount rate. Throughout the paper we treat γ not as a preference parameter but rather as a descriptive empirical measurement of the per-period impatience that is implicit in choice. We then have:

$$(1-\gamma)^{t_1/12} x_1 = (1-\gamma)^{t_2/12} x_2 \qquad \Leftrightarrow \qquad \gamma = 1 - \left(\frac{x_1}{x_2}\right)^{12/\Delta t} = 1 - e^{-12 \cdot RRR/\Delta t}, \quad (1)$$

where $RRR/\Delta t = ln(x_2/x_1)/\Delta t$ is the "interval-adjusted required rate of return" that the decision maker reveals through her choices. In the exponential model, the interval-adjusted RRR – and, hence, also γ – are constant, absent confounding factors. In what follows, we refer to the empirical measurement of γ as "implied annual impatience," which is an approximation of the average annual discount rate implied by a decision.

Since Thaler (1981), behavioral economists have gathered significant evidence that decision makers behave in ways that are incompatible with the exponential discounting model when valuing financial flows (see Cohen et al. (2020) and Ericson and Laibson (2019) for reviews). Many of the core anomalies boil down to the observation that per-period impatience varies systematically as a function of the length of the time interval, Δt, and in particular that empirically observed discounting has a hyperbolic shape.[3]

Extreme short-run impatience. Many studies show that decision makers tend to discount relatively short intervals at such an extremely steep rate that a constant discount rate would imply implausible and empirically counterfactual decisions over longer inter-

[3]There are other choice anomalies that relate not to dates, but instead to payoffs – most notably magnitude effects and gain-loss asymmetries (Cohen et al., 2020). Our experiments concern only the classic anomalies related to variation in time.

vals.[4]

Decreasing impatience. Extreme short-run impatience is a component of a more general tendency for decision makers' revealed per-period impatience to decrease as the interval Δt becomes longer.

Sub-unitary estimates of β. It is common in the literature to summarize the hyperbolic pattern of discounting described above using structural estimates of the quasi-hyperbolic $\beta-\delta$ model (Laibson, 1997). This model postulates that decision makers put weight 1 on immediately paid ($t = 0$) rewards but weight $\beta\delta^t$ on all delayed ($t > 0$) rewards. Structural estimates tend to find values of β significantly lower than 1 in most datasets, which is routinely interpreted as evidence of present bias. However, this evidence is indirect in the sense that it is typically identified from the hyperbolicity of discounting per se. By contrast, as discussed below, present bias can alternatively be estimated *causally* using front-end delay designs (or revising-earlier-choices designs as in Augenblick et al. (2015)).

We will call this triplet of phenomena the classical "hyperbolic pattern." It is generally thought to be a consequence of one (or both) of two possible behavioral tendencies. The first is *non-stationary* discounting: holding the length of an interval Δt fixed, people might discount at different rates depending on the earlier date, t_1. Non-stationarity is typically measured using "front-end delay" designs, which produce some direct evidence of non-stationarity:

Front-end delay effects. Shifting the date of both the earlier and the later payment forward by a fixed delay d often causes a decrease in impatience. This is in violation of exponential discounting, which prescribes that the starting date (t_1) does not matter for the average per-period discount rate γ. Front-end delay designs in which the earlier date is the present (i.e. $t_1 = 0$) are a direct way of measuring "present bias" – a form of non-stationarity in which people put a special premium on immediately-paid rewards.

A second possible driver of the hyperbolic pattern is that people might be partially *insensitive* to the length of a time interval, producing a decreasing instead of a constant per-period discount rate. This insensitivity is conventionally measured using "subaddi-

[4]This tendency is sometimes confused with "present bias" in part because many early studies exclusively featured problems in which the sooner payoff date was immediate (i.e. $t_1 = 0$). More recent evidence (e.g., Kable and Glimcher, 2010) shows that, even when $t_1 > 0$, decision makers exhibit short-run impatience that is so pronounced that a constant discount factor would imply implausible and empirically counterfactual medium-run discounting behavior. See our discussion in Sections 5 and 6.

tivity" designs that rule out non-stationarity by comparing how people value *nested* time delays:

Subadditivity. There is robust evidence that decision makers violate transitivity by discounting composite time intervals of the form (t_1, t_3) less than they dicount joint decisions over sub-intervals (t_1, t_2) and (t_2, t_3), with $t_1 < t_2 < t_3$.[5] Because these choice sequences span the exact same dates this behavior reflects an insensitivity to time intervals that cannot be attributed to non-stationarity.

2.2 Temporal Motivations and Complexity

A number of explanations have been offered for these deviations from the exponential model. We group these into two broad classes: *motivational* explanations that are rooted in the actual passage of time, and *complexity* explanations that are instead rooted in the costs and difficulties of reasoning about relative intertemporal rewards.

By far the dominant class of explanations offered by the literature are *motivational* in nature: explanations rooted in preferences or internal conflicts that arise due to the fact that intertemporal choices involve the elapse of time. One category is *preference-based* explanations which argue that people simply do not have exponential, dynamically consistent time preferences, leading to non-exponential discounting behavior. This includes, for instance, hyperbolic and quasi-hyperbolic preference models (e.g., Loewenstein and Prelec, 1992; Laibson, 1997; O'Donoghue and Rabin, 1999). Other authors have proposed that people have, in effect, "multiple selves" with divergent preferences at different dates, strategically vying for control (Fudenberg and Levine, 2006). Another set of models posits temptation effects (Gul and Pesendorfer, 2001). A final class of motivational explanations emphasizes how transaction costs (of collecting delayed payments) and / or the inherent riskiness of the future induce decision makers to behave in such a way as to produce anomalies (e.g., Halevy, 2008; Andreoni and Sprenger, 2012; Chakraborty et al., 2020).

In contrast to this first class, a second class of explanations argues that anomalies arise because intertemporal decision making is inherently *complex*: it requires potentially costly cognitive operations. In order to properly discount a delayed payment in the process of valuing it, an exponential decision maker must (consciously or unconsciously) (i) introspect or calculate her discount factor, δ, and then (ii) reason recursively, discounting value by δ at each step of iteration. Both processes require potentially sig-

[5] Formally, denote by $a_{i,j}$ the indifference point for the tradeoff over interval (t_i, t_j). Then, subadditivity means that there is less discounting (more patient indifference values) over the single long interval: $a_{1,3} > a_{1,2}a_{2,3}$, or, equivalently, $\gamma(a_{1,3}) > \gamma(a_{1,2}a_{2,3})$.

nificant information processing, which is likely to be costly for decision makers (such procedural costs are what the term "complexity" means, e.g., in computer science).

If the procedural costs of precise evaluation are sufficiently large relative to available cognitive resources, decision makers may substitute to less costly, imprecise or heuristic methods of valuation instead. Throughout the paper, we will refer to this causal sequence as a "complexity response": rational choice is complex and therefore costly, creating an incentive for the decision maker to use a simpler-than-rational noisy or heuristic procedure to guide her choices instead.

A small-but-growing theoretical literature has shown that many plausibly simpler-than-rational evaluation procedures are capable of generating classical intertemporal choice anomalies. Each of these complexity-inspired accounts effectively describes a different imperfect or imprecise valuation procedure that decision makers might substitute to in lieu of precise evaluation. For instance, Rubinstein (2003), Ericson et al. (2015) and Read et al. (2013) describe heuristic procedures that avoid iterative evaluation altogether. Other recent complexity-inspired papers postulate that people *do* engage in iterative reasoning, but reduce the costs of doing so by reasoning imprecisely. This broad idea has been modeled in several different ways, including noisy introspection about one's discount factor (Regenwetter et al., 2018; Lu and Saito, 2018; He et al., 2019), noisy time perception (Brocas et al., 2018; Zauberman et al., 2009) and Bayesian cognitive noise models in which decision makers combine a noisy mental representation of the problem with a prior (Gabaix and Laibson, 2022; Vieider, 2021; Gershman and Bhui, 2020). What most of these accounts have in common is that they describe decision-making procedures that are excessively insensitive to the parameters of the decision problem (e.g., the length of a time interval) relative to the exponential model. As discussed above, this kind of insensitivity can generate the hyperbolic pattern: excessive short-term discounting and decreasing impatience.

Our goal is to empirically separate these two broad classes of explanations for intertemporal choice anomalies. We propose two strategies for doing this. An advantage of both of these strategies is that neither relies on a specific model of which noisy or heuristic procedures people resort to when they find a problem complex.

2.2.1 Removing Temporal Motivations

A key observation motivating our study is that (unlike the motivational explanations described above) the complexity-based explanations do not rely in any special way on the actual elapse of time. They instead rely on the idea that intertemporal choice problems require decision makers to conduct intensive information processing. Because they do not depend on time, these sorts of explanations may also generate anomalies in decision problems that involve no actual temporal delay, but that require a structurally similar (and thus similarly complex) type of reasoning.

Building on this observation, we propose a method for separating the two classes of explanations that has the virtue that it doesn't require us to articulate any specific model of complexity response. For any intertemporal choice problem $D = (x_1, t_1; x_2, t_2)$, we can construct an immediately paid "atemporal mirror" M_D of that same problem that replaces payment dates with a sequence of "steps" of payoff discounting. In each step of discounting, the payoff from the previous step is multiplied by an exogenously provided and known fixed factor $\delta < 1$. Thus, an atemporal mirror of D pays a deterministic amount $\delta^{t_1} x_1$ or $\delta^{t_2} x_2$ immediately. Instead of, e.g., valuing a payoff "\$50 in two months," a decision maker evaluating a mirror is asked to value a payoff "\$50 shrunk by δ two times." An atemporal mirror is therefore nothing more or less than an immediate dollar payment that has been deliberately described in such a way as to require a similar kind of information processing as is required in intertemporal choice.[6]

Because atemporal mirrors involve no actual time delays, anomalies in their evaluation cannot be driven by any of the motivational explanations reviewed above. For instance, they cannot be driven by non-exponential time preferences: an atemporal mirror *induces* exponential preferences. Similarly, there is no scope for multiple selves, temptation effects or "implicit risk" to generate anomalies. However, atemporal mirrors do maintain much of the *complexity* of intertemporal choice: under exponential discounted utility theory, the cognitive act required to properly value a mirror is formally highly similar to the cognitive act required to precisely discount a future payment. As such, the suite of simpler-than-optimal decision procedures – ranging from heuristics to different types of noise – remain available to distort choice.[7]

Our strategy is to first compare the way decision makers evaluate a set of intertemporal choice problems to the way they evaluate atemporal mirrors of those same problems. To whatever extent the same anomalies arise in the evaluation of mirrors as in the evaluation of intertemporal choice, we have evidence against motivational explanations and evidence in favor of complexity explanations. Second, we correlate the magnitude of anomalies in the two types of problems across subjects to verify that anomalies in the two cases are driven by a related behavioral mechanism.

[6]This empirical strategy is similar to that used in Oprea (2022), which compares valuations of lotteries to valuations of "deterministic mirrors" of lotteries which contain no risk.

[7]If anything, the information processing required to value atemporal mirrors might be lower than in true intertemporal choice. First, people might have non-linear utility, which does not matter for a choice between two atemporal mirrors which include no scope for risk. Second, in intertemporal decisions people may not have access to their own discount factor, which the atemporal mirrors transparently provide.

2.2.2 Linking Anomalies to Signatures of Complexity Responses

A second, and complementary approach to removing temporal motivations is to directly measure auxiliary evidence of complexity responses in traditional intertemporal choice, and study whether this evidence predicts the severity of anomalies. Our strategy relies on the observation that complexity-inspired intertemporal choice theories describe noisy and / or heuristic decision procedures that serve as proximate mechanisms for classical anomalies. These kinds of complexity responses often produce auxiliary signatures that are identifiable in the data. We measure these signatures and link them to the anomalies.

The literature has proposed two empirical indicators that a decision was made using a heuristic or noisy procedure: self-reported cognitive uncertainty and choice inconsistencies. First, we ask subjects how likely it is (in percentage terms) that their choice actually complies with their true tastes and preferences. This type of simple, unincentivized self-report about the optimality of choice has been shown to be highly predictive of anomalous behaviors in several other choice settings (Enke and Graeber, 2022; Arts et al., 2020), in no small part because people often appear to have some awareness when they are using imperfectly rational procedures. Second, we look for direct evidence of inconsistent, noisy decision-making, another indication of the use of an imperfectly rational decision procedure. Following the literature (e.g. Agranov and Ortoleva, 2017, 2020; Agranov et al., 2020; Khaw et al., 2021), we classify a decision as deriving from a noisy procedure if it is *different* from other choices made in repeated elicitations of an identical decision problem.

Our empirical strategy is to study to what degree one or both of these behavioral signatures of complexity responses predict the incidence and severity of anomalies. These kinds of correlations are clearly consistent with complexity-based explanations, but are not implied in any obvious way by preference-based (or other motivational) explanations. Thus, to the degree we find that these signatures are related to anomalies, we have evidence suggesting that the anomalies are outgrowths of complexity.

An attractive feature of our multi-pronged research design is that it relies on two essentially orthogonal empirical approaches with different strengths and weaknesses: (i) stripping away time and inducing exponential preferences while maintaining a very similar degree of complexity and (ii) directly measuring signatures of complexity responses in standard intertemporal problems. To the degree the approaches identify the same anomalies as being complexity-driven, we will have encouraging converging evidence.

Sessions	Description	Subjects
Delay & Mirror	18 tasks under *Mirror* & exactly repeated under *Delay* (order of treatments randomized)	500
Delay-M	12 delay tasks with elicitations of cognitive uncertainty and choice inconsistencies	645
Voucher-M	12 delay tasks with UberEats vouchers and elicitations of cognitive uncertainty and choice inconsistencies	500

Table 1: Overview of main experiments.

3 Experimental Design

3.1 Basic Setup

Table 1 provides an overview of the experimental design. Following standard methods used in the literature, the core tasks in our experiments are *multiple price lists* that ask subjects to evaluate a payment of x_2 at a time t_2 in terms of dollars paid at an earlier date $t_1 < t_2$. An example of the subject's decision screen is shown in Appendix Figure 8. In each list, Option A is kept identical in every row, paying x_2 at date t_2. By contrast, Option B pays an amount x_1 that declines montonotically by \$2 in each row (ranging between x_2 and \$2), at date t_1. Non-negative discounting entails that subjects choose A in early rows of the list (or, with extreme preferences, never) and switch to B at some later row (we enforce single switching). The switching point between earlier and later payment yields a direct measure of the RRR and thereby implied annual impatience, γ.

We refer to these choice problems as the *Delay* treatment. In most cases we randomize (at the subject-list level) the delayed payment $x_2 \in \{\$40, \$42, \ldots \$52\}$. The experimental design includes three main types of price lists. First, "Now Lists," in which $t_1 = 0$ and t_2 varies across 1/4, 1, 2, 12, 24, 36, 48 and 84 months. Second, "Later Lists," which are identical to Now Lists except that the earlier payment is slightly delayed: $t_1 = 1$ or $t_1 = 1/4$ months. Finally, "Subadditivity/Front-End Delay (SA/FED) Lists", in which for some horizon T we assign subjects lists (t_1=0, $t_2 = T/2$), (t_1=T/2, $t_2 = T$) and (t_1=0, $t_2 = T$), maintaining a consistent x_2 across the three lists. We randomly assign T across subjects to be either 8 or 12. Dates t_1, t_2 represent months. Lists from each of these categories are included in every treatment, for every subject and randomly ordered at the subject level. Now and Later lists are used primarily to study the anomalies of extreme short-run discounting, decreasing impatience / hyperbolicity and sub-unitary β. SA/FED Lists are used to measure subadditivity and front-end delay effects.

3.2 Removing Temporal Motivations

Our first variation on this standard choice setting is to study companion problems in which we pay subjects a iteratively discounted version of the stated payoff immediately. In these tasks, discounting occurs though a known, exogenous discount factor, transforming A and B into "atemporal mirrors" of standard intertemporal choice tasks. This is framed to subjects as "shrinking" a payment t times. Each time a payment is "shrunk," it falls to $\delta < 1$ of its previous value, but a subject must reason through the consequence of this discounting in order to properly value it. The fact that atemporal mirrors are paid immediately is repeatedly emphasized to subjects in the instructions.

A choice list from treatment *Mirror* is displayed in Appendix Figure 8. Each list asks subjects an exactly analogous sequence of binary choice questions as in the corresponding list from the *Delay* treatment. Option A (kept identical in each row of the list) is a dollar payment, paid out immediately but iteratively discounted some number of times. Option B is a dollar payment that involves strictly fewer iterations, and often none, which mimics an immediate or earlier payment. For example, in one row of a list, subjects are asked to choose between "Option A: \$42 shrunk 12 times" and "Option B: \$2". We again elicit a standard switching interval to calculate the implied "annual impatience".

The mirrors we implement include a single step of discounting for each month of discounting in the *Delay* problem it mirrors. Throughout the experiment, we set the per-period $\delta = 0.96$ (based on actual estimates of the discount factor δ from the intertemporal choices made in other sessions of the design, discussed below).

Every subject participated in both *Delay* and *Mirror* in a random order. The upside of this within-subjects design is that it allows us to correlate behavior in the two types of problems across subjects. When we are not interested in correlating behavior across treatments, we take care to rule out contamination effects by only analyzing decisions from the treatment that a subject encountered first (the results are very similar when we also include the data from the second-assigned treatment, see Appendix Table 7).

Because the treatments were designed to be compared to one another, we took great pains to use an identical interface and identical numbers. However, we were also careful to strongly differentiate the two treatments from one another using clear instructions. Importantly, to minimize cross-treatment contagion, subjects first assigned to *Mirror* did not know they would later be making intertemporal choices, and vice versa.

The *Mirror* treatment is incentivized using real payments, but the *Delay* treatment is a purely hypothetical elicitation. This was unavoidable because our motivating questions in *Delay* require us to study choices regarding multi-year delays, which are infeasible to implement using real incentive schemes. We expect this design choice to have little effect on our *Delay* results, as we discuss in Section 3.4 below. Still, to whatever degree hypothetical payments lead to, e.g., less careful decision making in *Delay* than in *Mirror*, we should expect this to work *against* the complexity hypothesis we are testing when we contrast the two treatments – we would expect hypotheticals, if anything, to exaggerate anomalies in unincentivized *Delay* observations relative to incentivized *Mirror* observations. We view this, therefore, as a conservative feature of our design.

3.3 Measuring Evidence of Noisy / Heuristic Behavior

As motivated in Section 2.2, in other treatments we gather auxiliary evidence that subjects are responding to the complexity of intertemporal choice by using noisy or heuristic decision procedures. To do this, we implement treatments *Delay-M* and *Voucher-M*. In both of these treatments, we measure the following objects.

Cognitive Uncertainty. Adapting the methodology from Enke and Graeber (2022), after each choice list, we measure cognitive uncertainty (CU) as the subject's subjective probabilistic belief that their true valuation of the later payment is actually contained in their stated switching interval:

> *Your choices on the previous screen indicate that you value \$$x_2$ in t_2 somewhere between \$a and \$b in t_1. How certain are you that you actually value \$$x_2$ in t_2 somewhere between \$a and \$b in t_1?*

Participants answer this question by selecting a radio button between 0% and 100%, in steps of 5%, see Appendix Figure 9. We interpret this question as measuring the participant's awareness that their decision procedure is noisy or heuristic.[8] We are agnostic about what exact sources of complexity in the decision problem cause subjects to doubt their decisions. The measure is not incentivized. We, again, view this as a conservative feature of our design: if subjects expend mental effort in the intertemporal choices but not in the CU question, then our results will make the links between CU and intertemporal choice anomalies look smaller than they actually are.

Choice inconsistencies. A standard way of measuring the noisiness of subjects' decision procedure is choice inconsistency in repetitions of the same choice problem. In our study, each subject completes two randomly selected choice lists twice. We generate a binary indicator that equals one if the subject's decisions on the two repeated trials are

[8]We ensure that subjects do not misunderstand the question as referring to *external* uncertainty that they may not actually receive the reward. To this effect, our experiments include a comprehension check question that directly asks participants to indicate whether the CU elicitation question asks about (i) the subject's subjective probability of actually receiving the money or (ii) their certainty about their own valuation, given that they know they will receive the money with certainty.

different from each other. We verify that our results continue to hold if we instead compute the absolute difference between the two decisions as our measure of inconsistency.

We collected these two pieces of data in two different treatments. In *Delay-M* we again use hypothetical monetary payments, which allows us to study multi-year delays. However, we pair this with a second incentivized treatment to show that our findings are robust to the inclusion of incentives. The *Voucher-M* treatment is identical to the *Delay* treatments in most respects, except (i) that we actually pay subjects for their choices using UberEats food delivery vouchers[9] and (ii) we do not study delays of more than one year (for feasibility reasons). In *Voucher-M*, payments are denominated in UberEats vouchers usable starting at date t_1 or $t_2 \leq 12$, respectively; these vouchers are valid for a period of only seven days from the starting date, which minimizes fungibility concerns. Subjects again complete multiple price lists, except that all payments refer to UberEats vouchers (of value between \$40 and \$50).

Participants' vouchers were directly credited to their personal UberEats accounts within 10 hours of completion of the study, such that subjects did not have to actively claim the voucher. The vouchers were always visible in their accounts, they could just not be used before the validity period. Because participants could always view vouchers in their account within a few hours of the study regardless of the precise validity period, there is no differential payment risk across vouchers with different time delays. Participants received automatic reminders 24 hours before a voucher became valid and 24 hours before it expired.

3.4 Discussion of Implementation

Our experimental design includes two characteristics that merit comment in light of common discussions in the literature. First, our *Delay* and *Delay-M* treatments both feature hypothetical payments – a design decision that allows us to elicit choices over multi-year delays that would be difficult to assess using an incentivized design. There are strong reasons to believe that this is a benign design choice. Reviewing the literature, Cohen et al. (2020) conclude "there is little evidence of systematic differences between RRR in incentivized and unincentivized experiments." Nonetheless, we use the *Voucher-M* treatment to probe robustness to incentivizing elicitations.

Second, our experiments (like the bulk of the literature that motivates the anomalies) involve delays in monetary payments rather than in consumption. A common concern is that these experiments may not elicit true time preferences because money (unlike

[9] UberEats is a takeout delivery service that can be used for a wide array of restaurants. It is widely available throughout the United States (Curry, 2021).

consumption) is fungible.[10] This concern is less relevant for our purposes than in much of the literature because we are not attempting to measure true time *preferences* – we are only interested in measuring and explaining the kind of discounting *behavior* that is often documented even in money experiments. Indeed, a central conclusion of ours is that even putting fungibility concerns aside, intertemporal choice experiments to a great extent fail to recover preferences due to the confounding influence of complexity.

3.5 Procedures

All experiments were conducted on Prolific. Online Appendix D contains details on experimental instructions, visual display and screening questions used.

Subjects in the *Mirror & Delay* sessions were paid a \$6 base payment and had a 20% chance of being paid a bonus based on their choice from a randomly selected list and row of Mirror (or from a separate risk elicitation we included in our sessions). In *Delay-M*, subjects earned a flat \$4.50 payment. In *Voucher-M*, subjects received a \$4 base payment and voucher payments from a randomly selected list and row with 25% chance.

4 Complexity and Hyperbolic Discounting

4.1 Evidence from Atemporal Mirrors

We begin by examining whether hyperbolic discounting appears in *Mirror*, and whether it predicts the hyperbolicity that appears in *Delay*. In analyzing this data, it is important to emphasize that there are at best weak reasons to expect similar "patience" *levels* in the two treatments. In *Mirror*, subjects face an induced discount factor of 0.96; in *Delay*, choices depend on subjects' individual discount factors, which may differ from 0.96. Our focus will therefore be on comparing the *severity of anomalies*, which are derived by comparing behaviors across different decision problems, rather than comparing patience levels.

Figure 1 provides a raw overview of the data by plotting, for both the *Delay* and *Mirror* treatments, the average switching point (expressed as a percentage of the "later" payment, x_2) as a function of the time interval or the number of discounting iterations.[11] Recall that in an exponential discounting framework with linear utility, these normalized switching points correspond to $\frac{x_1}{x_2} = \delta^{\Delta t}$. For the *Mirror* treatment, we overlay the indifference point that a payoff-maximizing subject would choose given the induced "monthly

[10]This view is not universal in the literature. An alternative line of argument holds that subjects narrowly bracket their choices and treat monetary amounts in experiments as proxies for utils (Halevy, 2014; Andreoni et al., 2018; Epper et al., 2020).

[11]We approximate switching points by computing the midpoint of the switching interval.

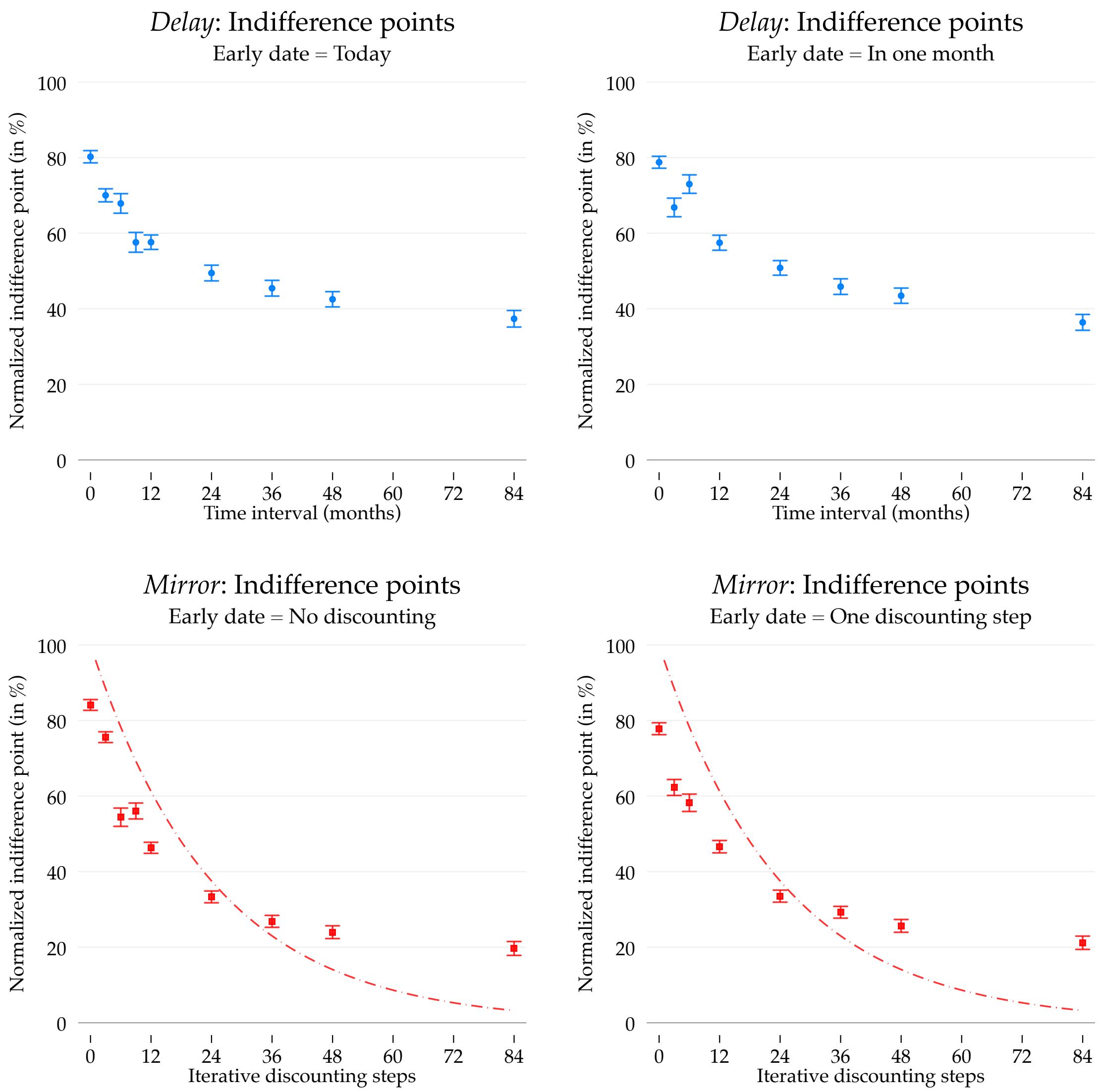

Figure 1: Average normalized indifference points by time interval (*Delay*) or number of iterations (*Mirror*). Top panels show *Delay* treatment (4,572 decisions from 254 participants). Bottom panels show *Mirror* treatment (4,428 decisions from 246 participants). In the *Mirror* panels, the dashed line represents payoff-maximizing decisions. Sample splits according to whether the earlier payment occurs today/requires no discounting. The time interval in months and the number of iterations are rounded to the nearest multiple of three. Whiskers show standard error bars, computed based on clustering at the subject level.

discount factor" ($\delta = 0.96$). The left panels plot data from Now Lists (the earlier date is immediate in *Delay* and paid with no discounting in *Mirror*); the right panels are from Later Lists (the earlier date is in one month or after one step of discounting).

Figure 2 transforms this data in a straightforward way by computing implied annual impatience, $\hat{\gamma} = 1 - (x_1/x_2)^{12/\Delta t} = 1 - e^{-RRR \cdot 12/\Delta t}$, see eq. (1).[12] For ease of illustration,

[12]A significant practical advantage to expressing decisions in terms of implied impatience γ rather than the interval-adjusted RRR is that the latter by construction produces large outliers that render visualizations and econometric analyses challenging.

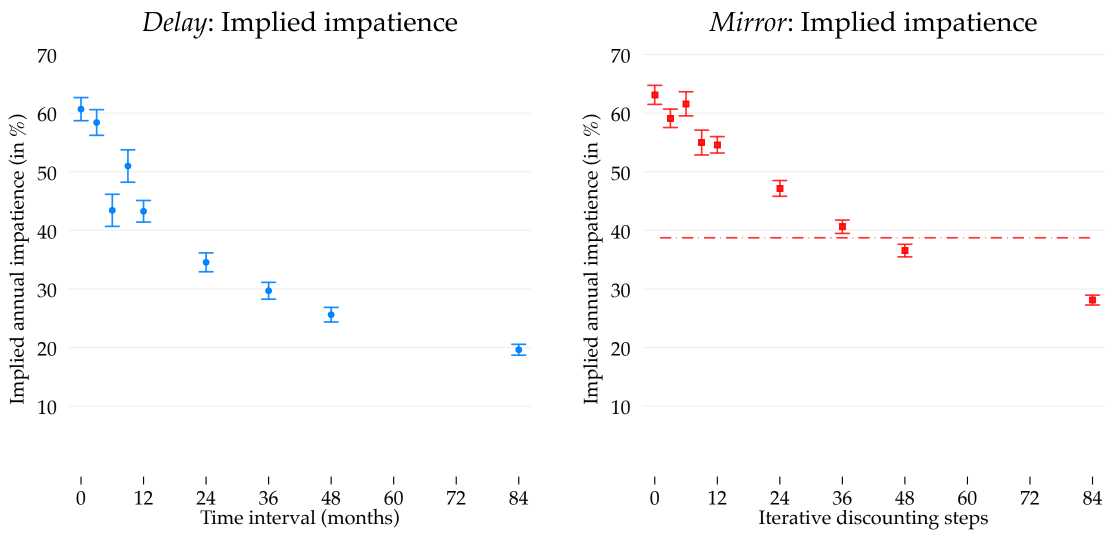

Figure 2: Average implied annual impatience $\hat{\gamma}$ by time interval (treatment *Delay*; 4,572 decisions from 254 participants) or number of iterations (treatment *Mirror*; 4,428 decisions from 246 participants). In the *Mirror* panel, the dashed line represents payoff-maximizing decisions. The time interval in months and the number of iterations are rounded to the nearest multiple of three. Whiskers show standard error bars, computed based on clustering at the subject level.

we combine Now Lists and Later Lists; the figures look very similar if separated.

Short-run impatience. It is clear from these figures that subjects in *Delay* show extremely high impatience over short horizons, both when the earlier payoff is immediate and when it is delayed by one month. Figure 2 shows that subjects discount very short horizons such as one month at an annualized rate of about 0.6.

Our first main finding is that subjects also show extreme discounting over the first few steps of discounting in *Mirror*, even though there is no delay in these problems – and even though subjects are incentivized to maximize an exponential discount function. Importantly, in *Mirror* (unlike in *Delay*) we can identify this behavior as a mistake: subjects discount payments made in $t_1 = 1$ or $t_1 = 2$ to a far greater degree than their true (experimentally induced) discount rate warrants.

Decreasing impatience. A second classical pattern visible in Figure 1 is that indifference payments are a highly compressed function of the time interval. Figure 2 shows that this compression implies that implied annual impatience is sharply and generally monotonically decreasing in the length of the interval. This pattern of decreasing impatience is a primary motivation for models of non-exponential time preferences like hyperbolic or quasi-hyperbolic discounting.

Our second main finding is that in the atemporal mirrors we similarly observe that indifference payments in Figure 1 are too compressed relative to the experimentally induced discount factor. As shown in Figure 2, this compression implies a strong decrease in implied annual impatience as the number of discounting steps increases. Once again the figure highlights that this is financially suboptimal behavior: subjects' average switch points in Figure 1 are located above the normative benchmark for few iterations but below it for many iterations.

To compare magnitudes across treatments, Appendix Table 6 presents regression evidence. In *Delay*, for each additional year, implied annual impatience decreases by 5.6 percentage points (pp). In *Mirror*, that effect is 4.8 pp, meaning that decreasing impatience in *Mirror* is 86% as strong as in *Delay*.

Result 1. *Subjects exhibit extreme short-run impatience and decreasing impatience when evaluating atemporal mirrors just as they do when evaluating delays. For mirrors, these are clear misvaluations.*

Linkage between Atemporal and Intertemporal Decisions. Next, we show that the appearance of similar anomalies in *Mirror* and *Delay* are not coincidental but are tightly linked at the individual level. To do this, unlike in the analyses above, we leverage our within-subjects design to examine the within-subject relationship between behaviors across the two treatments. If there is a common behavioral mechanism behind the anomalies across treatments (common heuristic or noisy responses to complexity), they should be correlated with each other.

We first link subjects' decisions in those choice problems that are direct mirror images of each other, such as "$40 in 6 months" vs. "$40 shrunk 6 times". Thus, we compute a correlation coefficient for (500 subjects * 18 unique problems * 2 treatments =) 18,000 observations. In doing so, we take care to net out that component of the correlation that is mechanically driven by the fact that for longer intervals or a higher number of iterations subjects should be expected to state lower valuations. Thus, we compute the partial correlation between decisions, netting out fixed effects for each unique problem type (each possible combination of t_1 and t_2). As a result, the correlation captures how similar subjects' behavior is across the two treatments, holding fixed the nature of the choice problem.

We find a partial correlation of $r = 0.34$ ($p < 0.01$), see Figure 3 for corresponding graphical evidence. This correlation is remarkably high given that the absence of time preference-based variation in the *Mirror* treatment should produce correlations *close to zero* for rational decision makers. Instead, behavior in *Mirror* produces one of the strongest predictors of intertemporal choice ever documented in the literature (Cohen et al., 2020).

A second approach to studying linkages across treatments is to focus not on each separate decision but, instead, on the magnitude of decreasing impatience: whether those

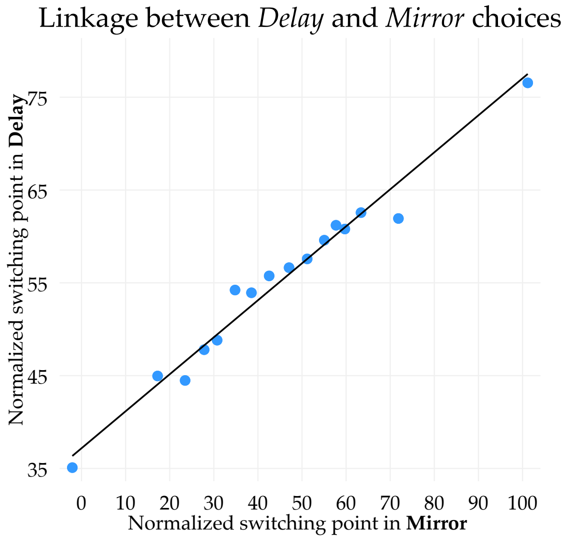

Figure 3: Binned scatter plot of normalized indifference points in structurally identical choice problems in *Delay* and *Mirror*. Partial correlation plot, controlling for fixed effects for each choice list type (each possible combination of t_1 and t_2). Based on 18,000 decisions by 500 subjects. The partial correlation is $r = 0.34$.

subjects whose impatience strongly decreases in the length of the interval in treatment *Delay* are also those whose implied per-period impatience strongly decreases in the number of iterations in treatment *Mirror*. To compute this, for each subject and treatment, we regress the annual impatience implied by a decision on the length of the time interval. The magnitude of the coefficient serves as a measure of decreasing impatience. Again, we find that, across subjects, the magnitudes of decreasing impatience in the two treatments exhibit a correlation of $r = 0.34$, $p < 0.01$.

Result 2. *Across subjects, valuation of time intervals is strongly correlated with valuation of atemporal mirrors. This suggests they are driven by a common behavioral mechanism, which cannot be temporal motivations.*

4.2 Results from Measures of Complexity Responses

Next, we turn to analyzing the data from treatments *Delay-M* and *Voucher-M*, where we elicited both cognitive uncertainty and potential choice inconsistencies.

The data from the *Delay-M* and *Voucher-M* treatments show strong *prima facie* evidence that subjects' decisions are noisy and / or heuristic in nature. In *Delay-M*, 75% of all decisions are associated with strictly positive CU and 60% of all repeated decisions show strictly positive inconsistency. In *Voucher-M*, the corresponding frequencies are 83% and 60%. These results strongly indicate wide-spread complexity responses (usage of imperfectly rational decision procedures) in the data. We now investigate how variation in these measures predicts the strength of anomalies.

The left panels of Figure 4 illustrate the raw data for the *Delay-M* treatment: the relationship between normalized indifference points (in percent) and time intervals. The panels split results based on the presence or absence of (i) measured CU in the decision (top panel) or (ii) choice inconsistency in the decision (bottom panel). The corresponding right-hand panels transform these data (as in the previous section), by computing the implied annual impatience $\hat{\gamma}$. All panels pool the data for Now and Later lists (the results are very similar looking at each of them separately). Figure 5 shows analogous results for the incentivized UberEats voucher experiments.

The top panels illustrate that decisions associated with strictly positive CU are considerably less sensitive to variation in the time interval, making them look considerably more hyperbolic. This has two direct implications. First, CU is strongly predictive of short-run impatience. For instance, in *Delay-M*, the raw correlation between normalized indifference points for one-week delays and CU is $\rho = -0.45$ both when $t_1 = 0$ and when $t_1 > 0$. In *Voucher-M*, the same correlations are $\rho = -0.39$ and $\rho = -0.45$.

The second implication of the compression of cognitively uncertain decisions is that implied annual impatience decreases much more rapidly in the time interval for uncertain than certain subjects. For instance, going from $\Delta t \approx 1$ to $\Delta t \approx 84$ months, the implied annual impatience drops by a factor of 4.5 for $CU > 0$, but only by a factor of 2 for $CU = 0$. Notably, in treatment *Delay-M*, this pattern implies that cognitively uncertain participants act as if they are *less* patient over relatively short horizons, yet *more* patient over relatively long horizons, with a crossover point at around one year.

The bottom panels of Figures 4 and 5 show analogous results for choice inconsistency. Again, we see that inconsistent decisions appear more impatient over short horizons but less impatient over long ones.

Notably, these patterns from cognitive uncertainty and choice inconsistency match exactly what we find in *Mirror*, where subjects are too "impatient" with few discounting iterations but too "patient" with many ones.

Table 2 investigates these patterns econometrically. Across specifications and datasets, we see that (i) both CU and choice inconsistency are strongly correlated with short-run impatience (columns (1) and (4)) and (ii) decreasing impatience is subantially stronger in the presence of either CU or choice inconsistencies (columns (3) and (6)).

What fraction of decreasing impatience is driven by noise / heuristics? To quantify this, we compare the magnitudes in two sub-samples: (i) decisions that are associated with no CU and no choice inconsistency vs. (ii) decisions that reflect either strictly positive CU or choice inconsistency. We examine how strongly implied annual impatience increases in the evaluated time interval, akin to the regressions in Table 2. We find that in the sample with no CU and no choice inconsistencies, the magnitude of decreasing im-

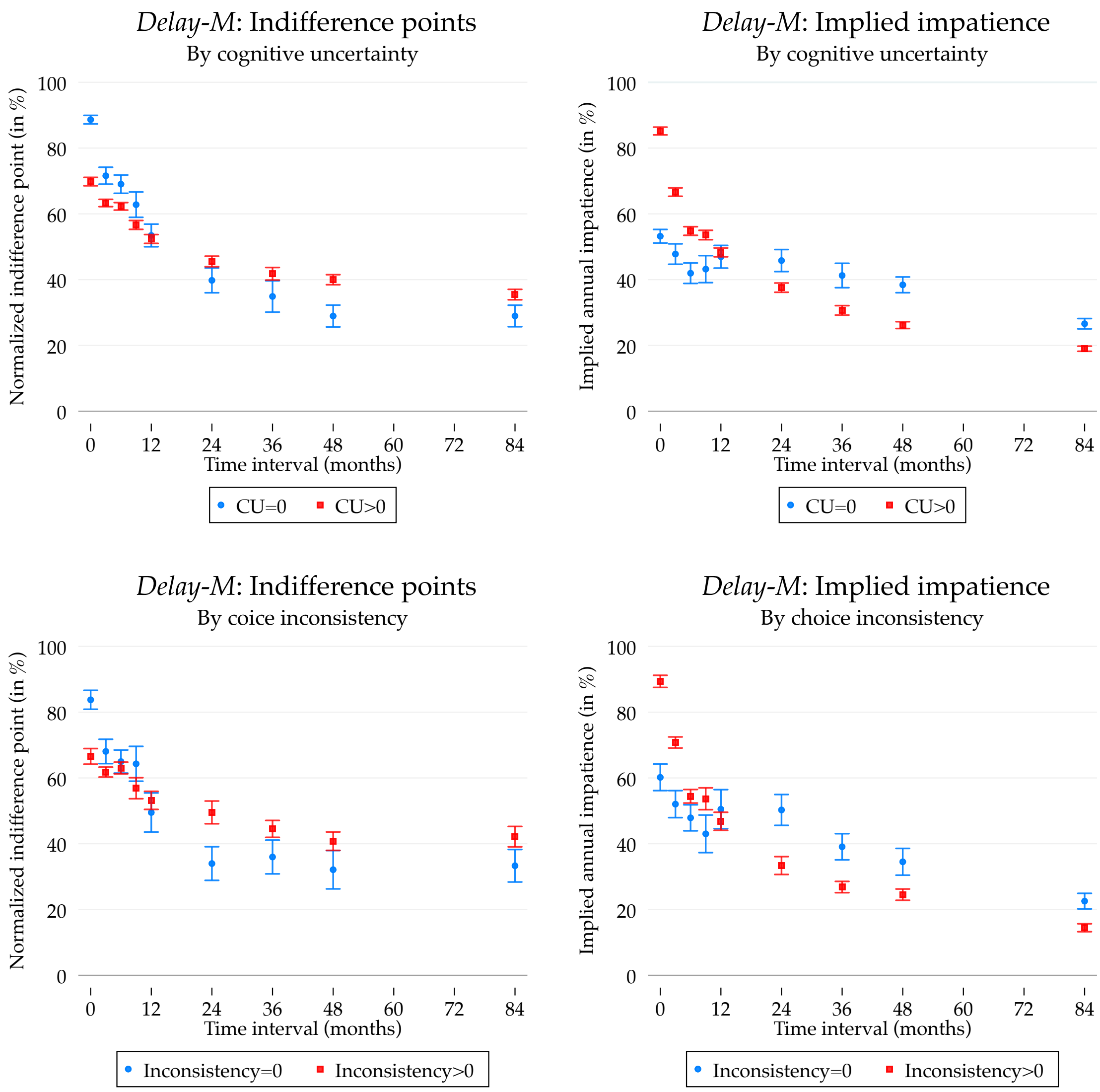

Figure 4: Normalized indifference points (left panels) and implied annual impatience (right panels) as a function of the time interval. The top panels include all decisions from *Delay-M*, and we split the sample according to whether or not a choice is associated with strictly positive CU (7,740 decisions by 645 subjects). The bottom panels include data from all decisions in *Delay-M* that were elicited twice (two repeated problems per subject for a total of 2,580 decisions from 645 subjects), and we split the sample according to whether or not decisions differed in a set of repeated choices. Time intervals are rounded to nearest multiple of three months. Whiskers show standard error bars, computed based on clustering at the subject level.

patience is only 10% of that in the comparison sample. This suggests that at least 90% of decreasing impatience is driven by noisy or heuristic complexity responses, rather than preferences. This conclusion is strikingly close, quantitatively, to the decomposition computed by comparing decreasing impatience in atemporal mirrors and time intervals in Section 4.1.

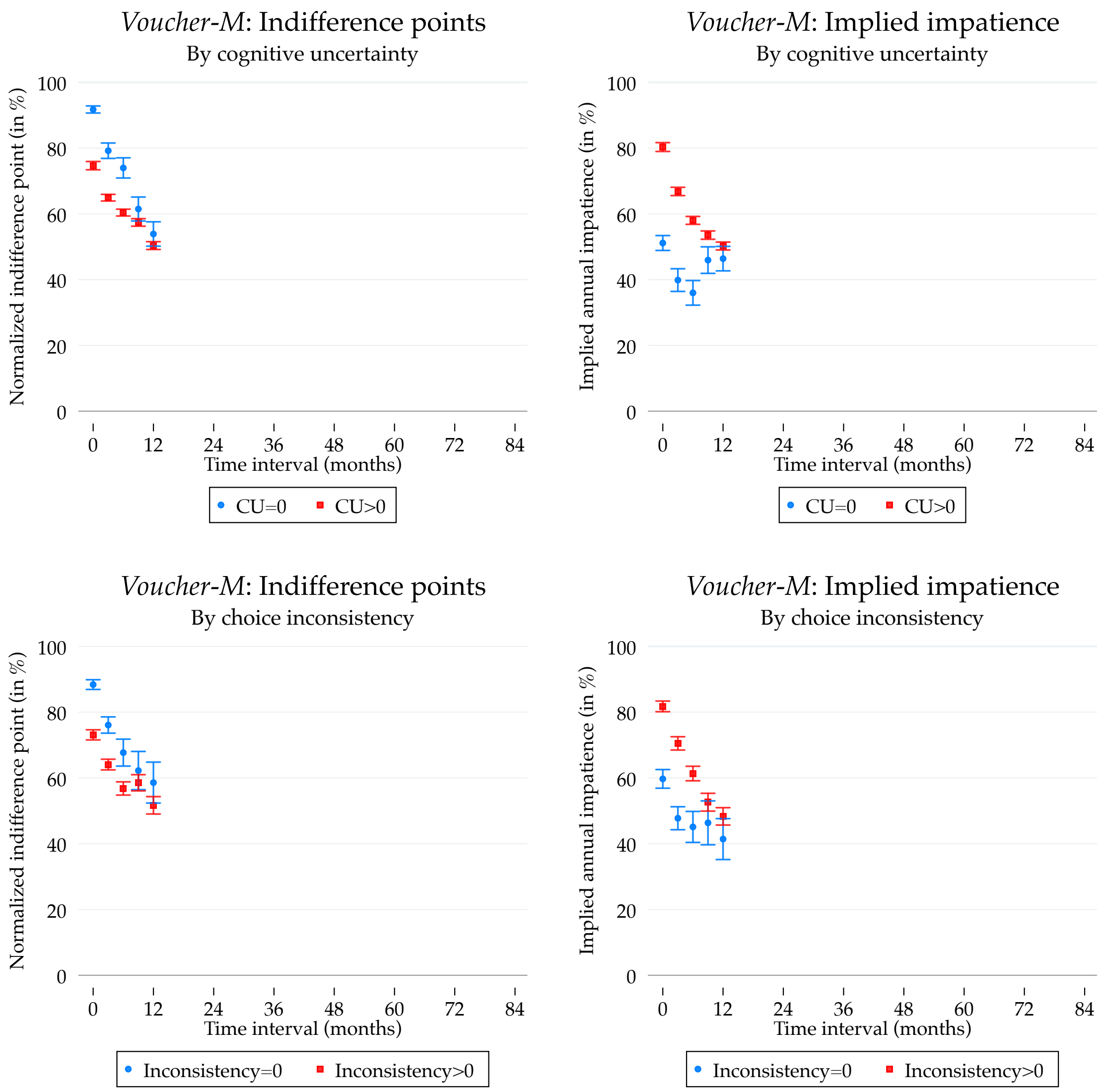

Figure 5: Normalized indifference points (left panels) and implied annual impatience (right panels) as a function of the time interval in *Voucher-M*. The top panels include all decisions, and we split the sample according to whether or not subjects indicate strictly positive CU (6,000 decisions from 500 subjects). The bottom panels include data from all decisions that were elicited twice (two repeated problems per subject, for a total of 2,000 decisions from 500 subjects), and we split the sample according to whether or not decisions differed in a set of repeated choices. Time intervals are rounded to nearest multiple of three months. Whiskers show standard error bars, computed based on clustering at the subject level.

Robustness 1: Sample splits. The results reported above do not hinge on splitting the sample into decisions with zero or strictly positive CU. To show this, we split the sample into CU quartiles. We find that the effect of the time interval on decisions continuously decreases (in absolute terms) as CU increases, see Appendix Figure 10. This shows that the results are not just driven by the particular sample splits employed above, and that variation in CU strongly predicts decisions also conditional on $CU > 0$.

Table 2: Short-run and decreasing impatience as functions of CU and choice inconsistency

	Dependent variable: Implied annual impatience (in %)					
Dataset:	*Delay-M*			*Voucher-M*		
Phenomenon:	SR imp. ($\leq$ 1m)	Decreasing impat.		SR imp. ($\leq$ 1m)	Decreasing impat.	
	(1)	(2)	(3)	(4)	(5)	(6)
Time interval		-7.61***	-3.56***		-29.1***	-20.6***
		(0.30)	(0.54)		(2.49)	(6.12)
Cognitive uncertainty	0.42***		0.24***	0.61***		0.57***
	(0.11)		(0.05)	(0.10)		(0.08)
Inconsistent decision	25.2***		11.9***	16.7***		19.6***
	(4.59)		(2.38)	(3.30)		(2.94)
Time interval × Cognitive uncertainty			-0.061***			-0.47***
			(0.01)			(0.13)
Time interval × Inconsistent decision			-4.95***			-13.0**
			(0.61)			(6.44)
Payment amount FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	344	2580	2580	766	2000	2000
R^2	0.24	0.20	0.24	0.17	0.06	0.19

Notes. OLS estimates, robust standard errors (in parentheses) are clustered at the subject level. In columns (1) and (4), the sample is restricted to time intervals of at most one month. To make the samples comparable across columns, we restrict attention to decisions for which the choice inconsistency variable is available. Time interval is in years. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Robustness 2: Within-subject variation. A potential concern is that subjects might interpret the CU question in heterogeneous ways. To show that this does not drive the results, we restrict attention to within-subject variation in CU. We normalize the CU data to have mean zero and standard deviation one for each subject, and then look at whether this pure within-subject measure of CU still predicts choices. Appendix Table 8 shows that this is the case.

Choice inconsistency in atemporal mirrors. In our *Mirror* treatment, we also repeated one randomly-selected choice list for each subject, allowing us to investigate whether choice inconsistencies are also linked to the anomalies in the evaluation of atemporal mirrors. As shown in Appendix Table 9, we find that choice inconsistencies are strongly predictive of "short-run impatience" and "decreasing impatience" in mirror valuations, just as they are of true intertemporal decisions. This further supports our interpretation that complexity leads to anomalies by inducing the use of noisy and heuristic decision procedures.

Result 3. *Short-run impatience and decreasing impatience are strongly correlated with auxiliary evidence of noisy or heuristic decision-making.*

4.3 Manipulation of Task Difficulty

Our interpretation of the correlations between anomalies and CU / choice inconcistency is that they reflect the influence complexity has on intertemporal choice. That is, we interpret the noisy/heuristic behavior we measure as a direct *response* to the fact that intertemporal choice is complex. To provide direct evidence for this linkage, we ran an additional experiment that exogenously increases the cognitive difficulty of intertemporal choice. To the degree this manipulation jointly intensifies (i) our measures of noisy/heuristic behavior and (ii) intertemporal choice anomalies, we have complementary causal evidence supporting our interpretation.

In treatment *Opaque Payments/Delays*, for a subset of subjects, ($N = 153$), we express all of the payoffs in the price list as an algebraic expression (e.g., $40 is described as "$(4*8/2)+(8*9/2)-12"). For another subset ($N = 149$), we express all dates in the price list as algebraic expressions (e.g., 1 year is described as "in (6*2/3-3) years AND (3*6/2-9) months AND (5*4/2-10) days"). These interventions are always paired with time constraints of 25 seconds to make the relevant information processing constraints more likely to bind. Both of these interventions are designed to increase the information processing required to evaluate intertemporal tradeoffs, thereby increasing the cost of rational choice (i.e., complexity).

We find that this intervention significantly increases both of our measures of boundedly rational choice. Average CU rises from 21.7% in *Delay-M* to 35.2% for *Opaque Payments/Delays*; choice inconsistency rises from 60.4% in *Delay-M* to 67.2% in *Opaque Payments/Delays* (both comparisons are statistically significant at least at the 5% level, see Appendix Table 10). These results suggest that both of our measures of noisy/heuristic decisions are sensitive to the complexity of the choice environment.

Next, we find that this manipulation simultaneously intensifies intertemporal choice anomalies. As Figure 6 shows, the decisions of subjects in *Opaque Payments/Delays* evince stronger short-run impatience and flatter long-run impatience than those of subjects in *Delay-M*, see Appendix Tables 10 and 11 for regression evidence. Thus, the exogenous manipulation of task difficulty has the same effects as the patterns we observed correlationally for choice inconsistency and CU: higher task difficulty leads to both more pronounced short-run impatience more pronounced decreasing impatience. This pattern not only matches our correlational results, but also our findings from *Mirror*.

Result 4. *Short-run impatience and decreasing impatience become significantly more pronounced when complexity is exogenously increased.*

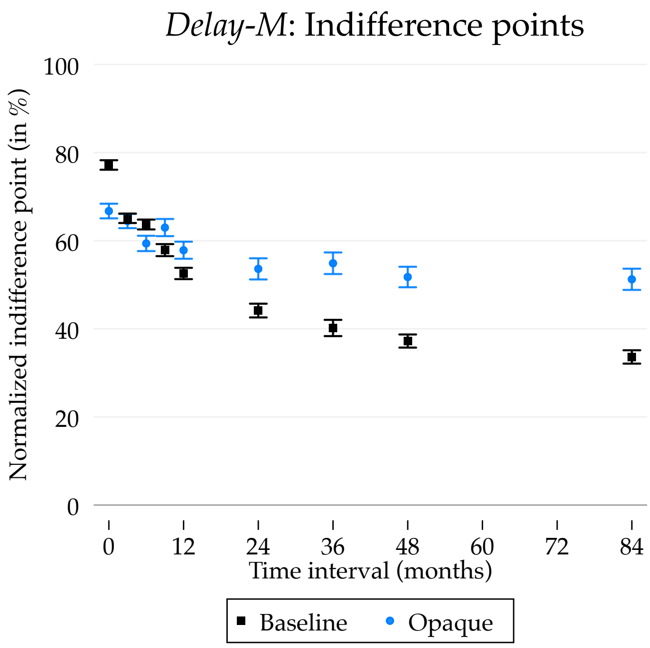

Figure 6: Normalized indifference points as a function of the time interval (rounded to nearest multiple of three months) in *Delay-M* and the two *Opaque* manipulations, pooled for ease of readability (11,364 decisions from 947 subjects). Whiskers show standard error bars, computed based on clustering at the subject level.

5 Complexity and Estimates of Present Bias

Our findings so far suggest that the hyperbolic shape of the empirical discount function is largely driven by complexity. What does this imply for our understanding of *present bias*, which is often modeled as the primitive underlying hyperbolic choice patterns?

Structural estimates of present bias. We begin by measuring present bias using the approach taken by the bulk of the literature: by structurally estimating the parameters of a $\beta-\delta$ model, as described in Section 2. Intuitively, in these model estimations, present bias is identified from the hyperbolicity of the discount function, including especially the excess degree of short-run impatience not captured by the estimated exponential discounting parameter, δ. Because we know from the previous section that this hyperbolicity is largely driven by complexity, there are strong reasons to hypothesize that structural estimates of β will, likewise, be linked to complexity. Because in this section we will frequently distinguish between structural and experimental estimates of present bias, we will denote structural estimates by $\hat{\beta}_{ST}$.

We first examine whether there is evidence for $\beta_{ST} < 1$ in our *Mirror* treatment, in which exponential discounting is experimentally induced and present-biased motivations are removed by design. Recall that in all of our experiments, a subject is asked to state an amount x_1 in t_1 that makes her indifferent to x_2 in t_2. In a $\beta - \delta$ model with linear utility, we, hence, have:

$$\delta^{t_1} \cdot x_1 = \beta_{t_1=0} \cdot \delta^{t_2} \cdot x_2 \tag{2}$$

We estimate this model at the population level, amended by a mean-zero error term. In *Mirror*, we estimate $\hat{\beta}_{ST} = 0.85$ ($s.e. = 0.01$) and $\hat{\delta} = 0.96$ ($s.e. = 0.01$).[13] Complexity alone, therefore, induces behavior that *looks like* present bias under the lens of standard estimation approaches. Intuitively, the reason for this result is that decisions in the *Mirror* treatment have a hyperbolic shape with high short-run "impatience", which gets attributed to a sub-unitary β_{ST}. Indeed, our estimates recover the true, induced δ of 0.96, suggesting that most of the distorting effects of complexity appear in the spurious estimate of β.

If complexity indeed confounds structural estimates of present bias, we should also see that – in traditional intertemporal choice experiments – decisions that are associated with stronger complexity responses (cognitive uncertainty and choice inconsistencies) are associated with more pronounced estimated present bias. To investigate this, we turn to the data from the *Delay-M* treatment. As is well-understood in the literature, individual-level heterogeneity in discount factors renders population-level estimates of β_{ST} potentially biased (Weitzman, 2001; Jackson and Yariv, 2014). Thus, we estimate eq. (2) separately for each subject.[14] Figure 7 shows a binned scatter plot of the resulting individual-level estimates of $\hat{\beta}_{ST}$ against a summary index of complexity responses (the first principal component of CU and choice inconsistency). Estimated present bias is strongly correlated with complexity responses (Spearman's $\rho = -0.28$, $p < 0.01$). Appendix Figure 12 shows that quantitatively almost identical results hold when the estimation accounts for utility curvature (measured through separate lottery choice lists at the end of the experiment). In combination with the result of "present bias" in the atemporal mirrors, this strongly suggests that conventional structural estimates of present bias to a great extent pick up complexity responses.

[13] Note that because all subjects are induced to have the same time preferences, estimates of β in atemporal mirrors do not run afoul of the aggregation concerns raised in the literature (Weitzman, 2001; Jackson and Yariv, 2014). Nonetheless, estimates at the individual level corroborate this result. As shown in Appendix Figure 11, for the majority (62%) of subjects we estimate $\hat{\beta}_{ST} < 1$.

[14] Population-level estimates deliver similar results on how β varies with complexity responses. Appendix Table 11 reports the results. For example, for $CU = 0$, we estimate $\hat{\beta}_{ST} = 0.87$, while for $CU > 0$ we get $\hat{\beta}_{ST} = 0.72$. In contrast, the estimates of δ are always very similar across the different sub-samples, suggesting that (as with our estimates from *Mirror*), complexity mostly influences the present bias term in estimates of these models. These results show that *even if aggregation was not an issue*, complexity responses would still bias the estimation of present bias.

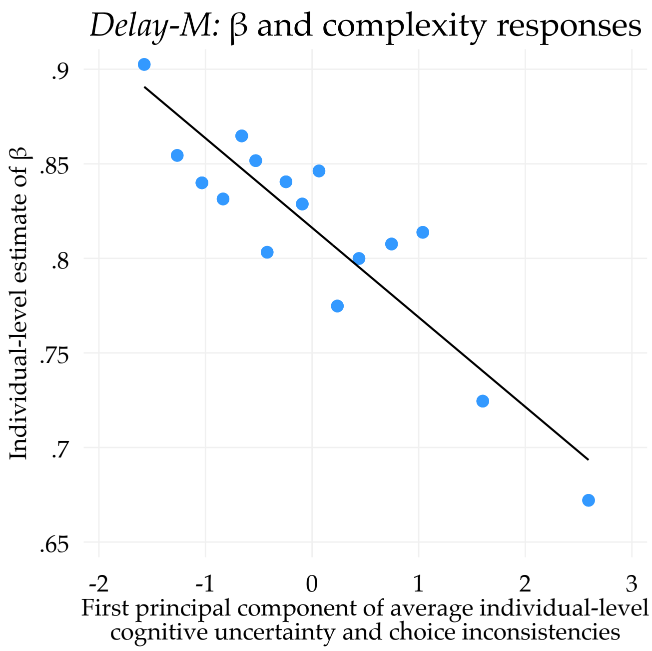

Figure 7: Binned scatter plot of individual-level $\hat{\beta}_{ST}$ in eq. (2) against first principal component of subject-level average CU and choice inconsistencies. Based on 643 subjects in *Delay-M*. The figure excludes two subjects for whom we estimate $\hat{\beta} > 2$.

Does complexity produce "true" present bias / non-stationarity? There are two potential reasons why heuristic or noisy complexity responses generate present bias in structural estimations. A first possibility is that complexity responses generate "true" present bias (i.e. stationarity violations in valuations). For example, the present may seem appealingly simple to people, contributing to a desire for immediate gratification. A second possibility, however, is that complexity produces hyperbolicity without generating stationarity violations (e.g., by producing insensitivity to variation in the size of time intervals). If this were the case, structural estimates of present bias would be confounded due to model misspecification resulting from omiting a role for these alternative effects of complexity on the discount function.

A standard way of causally identifying present bias in the literature is by measuring *front-end delay effects* (direct measurements of stationarity violations). Indeed, it is common in the literature to *define* genuine present bias through front-end delay effects (e.g., Chakraborty, 2021). In experimental documentations of these effects, subjects reveal lower discounting in evaluating $(t_1 + d, t_2 + d)$ than in (t_1, t_2), for $d > 0$. Some of our tasks feature such a front-end delay structure (with $t_1 = 0$ and d randomized between 4 and 6 months across subjects).[15]

[15] Jang and Urminsky (2022) document that front-end delay effects are typically only identifiable if the front-end delay is long. It is for this reason that we use the SA/FED lists (designed to study long 4- and 6-month delays) to look for evidence of front-end delay effects.

Table 3: Complexity and front-end delay effects

	Dependent variable: Implied annual impatience (in %)					
Phenomenon:	Front-end delay					
Treatment:	*Delay*	*Mirror*	*Delay-M*		*Voucher-M*	
	(1)	(2)	(3)	(4)	(5)	(6)
1 if front end delay	-4.24**	3.79**	-3.07***	-2.51	-4.11***	-7.23***
	(1.85)	(1.69)	(0.99)	(1.53)	(1.09)	(2.12)
Cognitive uncertainty				0.38***		0.38***
				(0.06)		(0.07)
1 if front end delay × Cognitive uncertainty				-0.058		0.070
				(0.05)		(0.07)
Payment amount FE	Yes	Yes	Yes	Yes	Yes	Yes
Task set FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	508	492	2393	2393	2337	2337
R^2	0.07	0.02	0.02	0.07	0.02	0.08

Notes. OLS estimates, robust standard errors (in parentheses) are clustered at the subject level. The data are restricted to problems that have a front-end delay structure. Task set FE are fixed effects for each pair of tasks that have a front-end delay structure. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3 summarizes the evidence on the link between complexity and front-end delay effects in our data. Columns (1) and (2) show that we find a statistically significant front-end delay effect in the *Delay* treatment but the *opposite* effect in the *Mirror* treatment. Thus, the mirror data provide no evidence that the front-end delay effect is an outgrowth of complexity. If anything, our results suggest that complexity might even work against the identification of these effects.

Columns (3)–(6) shows the results for treatments *Delay-M* and *Voucher-M*.[16] Again, we find evidence for the presence of front-end delay effects in intertemporal decisions. Importantly for our purposes, however, these effects are entirely uncorrelated with cognitive uncertainty, suggesting again that they have little to do with complexity responses. Finally, we show in Appendix Table 10 that the experimental complexity manipulation described in Section 4.3 does not amplify front-end delay effects, providing a third piece of evidence that front-end delay effects have little to do with complexity.

Thus, the atemporal mirrors, the measure of cognitive uncertainty and the experimental complexity manipulation all suggest that non-stationarity is *not* driven by complexity.

[16]Recall that we elicited only two randomly selected decisions per subject repeatedly. Given that these repeated decisions do not always occur for the choices in the SA/FED lists, we do not have access to a task-level measure of choice inconsistency that can be used to shed light on subadditivity or front-end delay effects. By contrast, the CU measure is available for each decision a subject makes.

Reconciling structural and experimental estimates. How is it possible that complexity is strongly linked to structural estimates of present bias but not to causal, experimental estimates? To address this, we first document that the magnitude of the present bias inferred from front-end delays is quantitatively considerably smaller than what is implied by the structural estimate $\hat{\beta}_{ST}$ reported above. To compute a causally-identified front-end-delay estimate (β_{FD}), we estimate eq. (2) only on those decision problems that have a front-end delay structure. In our two sets of intertemporal problems with a front-end delay structure, we estimate $\hat{\beta}_{FD} = 0.95$ and $\hat{\beta}_{FD} = 0.96$, respectively. While these estimates suggest strictly positive present bias, they are substantially larger (i.e. imply substantially less present bias) than the average individual-level structural estimate derived from estimating eq. (2) on the full dataset, $\hat{\beta}_{ST}^{ave} = 0.83$. A back-of-the-envelope calculation, hence, tentatively suggests that the structural estimate of 0.17 units of present bias can be roughly decomposed into 0.05 units attributable to true present bias (non-stationarity) and 0.12 units attributable to complexity responses.

Intuitively, the large difference between the structural estimate and the front-end delay estimate of present bias arises because the empirical discount function is substantially more hyperbolic than the magnitude of front-end delay effects would imply. Indeed, this discrepancy between the magnitude of front-end delay effects and of hyperbolic discounting is also highlighted in the review of Cohen et al. (2020). They call the coexistence of strongly decreasing impatience and relatively small front-end delay effects "contradictory patterns". Our results show that complexity is the main driver behind this discrepancy because it produces hyperbolicity but not front-end delay effects.[17]

To sum up, the severely inflated magnitude of structural estimates of present bias reflects model misspecification: conventional estimates of a $\beta - \delta$ model do not account for complexity responses, such that the complexity-induced hyperbolicity of the discount function gets spuriously attributed to β.

An important caveat is in order. Our experiments measure discounting over monetary rewards. We do not claim that the results of our back-of-the-envelope decomposition of estimated present bias into non-stationarity (e.g., a desire for immediate gratification) and complexity responses will be the same across different decision contexts and types of rewards. First, there are strong reasons to believe that in choices over primary rewards a desire for immediate gratification is larger than in choices over money (Cohen et al., 2020). It is, hence, plausible that in these contexts a larger fraction of estimated present bias reflects preferences or temptation. Second, the complexity of the choice environment – and the degree of experience the decision maker has with it – likewise plausibly affect the decomposition, with more complex environments likely generating

[17]Cohen et al. (2020) infer as much, attributing the remainder of hyperbolicity to the insensitivities described by subadditivity effects. See Section 6, below.

a larger role for complexity responses.

Result 5. *Structural estimates of present bias (that do not rely on causal experimental designs) are severely biased due to model misspecification resulting from omiting complexity responses.*

6 Complexity and Insensitivity

Our findings so far suggest that the hyperbolic shape of the empirical discount function is largely driven by complexity and the responses it inspires. In this section, we provide evidence on how complexity produces hyperbolicity. As discussed in Section 2.1, the literature has emphasized two proximal behaviors that might be responsible for hyperbolicity. The first is *non-stationary* discounting, which, in principle, might be a consequence complexity responses. However, in the previous section, we showed that there is no evidence that complexity contributes to stationarity violations.

The second proximal mechanism for hyperbolicity discussed in the literature is that people might treat intervals of different length in a non-exponential manner because they are *insensitive* to the length of the interval when evaluating delayed rewards (Read, 2001; Cohen et al., 2020). Such insensitivity, if present, would mechanically produce apparent decreasing impatience by causing subjects to treat longer and shorter time intervals too similarly to one another. Notably, similar complexity-driven insensitivities have been shown to drive anomalies in other domains in previous experimental work (e.g., Abeler and Jäger, 2015; Enke and Graeber, 2022; Oprea, 2022).

To study whether complexity induces insensitivity to time intervals, we follow the standard appraoch in the literature by measuring *subadditivity* effects. The subadditivity literature shows that impatience over a single time interval (t_1, t_3) tends to be considerably smaller than the total impatience people reveal when they are asked to make two decisions, one over interval (t_1, t_2) and one over (t_2, t_3), with $t_1 < t_2 < t_3$ (Read, 2001). The resulting transitivity violations are direct evidence of insensitivity because (i) they involve people treating shorter intervals too much like they treat a longer interval, but (ii) cannot be confounded with non-stationarity because they involve the comparison of nested intervals.

To investigate whether complexity causes the insensitivities to the interval length that are typically observed in these tasks, we included in all of our experiments choice lists in which we asked subjects to complete tasks that have a subadditivity structure, where we varied (t_1, t_2, t_3) randomly to be $(0, 4, 8)$ or $(0, 6, 12)$.

Table 4 summarizes the evidence. Columns (1) and (2) show how implied annual impatience differs between the choice over interval (t_1, t_3) and the combined choices over

Table 4: Complexity and subadditivity

	Dependent variable: Implied annual impatience (in %)					
Phenomenon:	Subadditivity					
Treatment:	*Delay*	*Mirror*	*Delay-M*		*Voucher-M*	
	(1)	(2)	(3)	(4)	(5)	(6)
1 if one long interval	-7.57***	-9.93***	-8.58***	-3.55***	-9.39***	-1.14
	(1.38)	(1.16)	(0.63)	(1.34)	(0.60)	(1.60)
Cognitive uncertainty				0.47***		0.45***
				(0.06)		(0.08)
1 if one long interval × Cognitive uncertainty				-0.24***		-0.33***
				(0.06)		(0.06)
Payment amount FE	Yes	Yes	Yes	Yes	Yes	Yes
Task set FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	508	492	1948	1948	2000	2000
R^2	0.09	0.06	0.03	0.08	0.04	0.08

Notes. OLS estimates, robust standard errors (in parentheses) are clustered at the subject level. The data are restricted to problems that have a subadditivity structure. We combine the three choices that make up a subadditivity set into two observations according to fn. 5. Task set FE are fixed effects for each pair of tasks that have a subadditivity structure. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

(t_1, t_2) and (t_2, t_3), separately for treatments *Delay* and *Mirror*. We find strong evidence for subadditivity in both treatments: people are roughly 10 percentage points less "patient" when a composite interval is broken up into two sub-intervals. Most importantly, the effect is *similarly strong* in atemporal mirrors and true delays, suggesting that all of the insensitivity of subadditivity is attributable to complexity rather than non-standard intertemporal preferences or self-control problems.

Columns (3)–(6) present the results on cognitive uncertainty in treatments *Delay-M* and *Voucher-M*. In both treatments, we find that the magnitude of subadditivity is strongly correlated with CU. Indeed, we find that subjects with $CU = 0$ exhibit no subadditivity at all. Thus, again, complexity seems to entirely explain the insensitivity of decisions to the interval. Finally, consistent with these correlational results, Appendix Table 10 shows that subadditivity effects also become substantially more pronounced in our *Opaque* treatments that increase the difficulty of intertemporal decision making, creating a third link to complexity.

Thus, the atemporal mirrors, the measure of cognitive uncertainty and the experimental complexity manipulation all suggest that complexity causes interval insensitivity (as measured by subadditivity). Combining this with our other findings, complexity produces insensitivity to the length of time intervals, which, in turn, produces hyperbolicity of the empirical discount function and thereby structural mis-estimates of present bias.

This linkage between complexity and insensitivity in our data strongly comports with theoretical accounts of how complexity generates intertemporal choices anomalies. As reviewed in Section 2, most such explanations root anomalies in heuristic or noisy decision procedures that produce hyperbolicity by making valuations less sensitive (less elastic) to variation in time delays than perfectly rational valuations are (e.g., He et al., 2019; Lu and Saito, 2018; Gabaix and Laibson, 2022; Vieider, 2021).

Result 6. *Complexity causes hyperbolicity because it generates a general insensitivity to interval length.*

7 Discussion

Table 5 summarizes the results from our paper across all of our treatments. The main takeaway is that even though we deployed several very different types of research strategies, we find highly consistent results. Regardless of how we measure complexity and complexity responses (through atemporal mirrors, choice inconsistency, cognitive uncertainty and exogenous treatment interventions), we consistently find that complexity is strongly associated with short-run impatience, decreasing impatience, subadditivity and structural estimates of present bias, but not with front-end delay effects ("true" present bias). Indeed, across methods, we find strikingly similar *quantitative* evidence that each of these anomalies is *primarily* attributable to complexity.

We interpret this as evidence that intertemporal tradeoffs appear to generate behavioral distortions in large part because they require a great deal of costly cognitive information processing. Subjects respond to this complexity by substituting from costly rational procedures for evaluating tradeoffs to less costly noisy or heuristic alternatives that are relatively inelastic to time intervals, generating systematic departures from time consistency.

Understanding present bias. A perhaps surprising result from this paper is that the noisy / heuristic procedures that complexity triggers, produce systematic confounds in the structural estimation of one of the central objects of interest in behavioral economics: β. As we have shown, even though complexity has no impact on front-end delay effects (a *causal* estimate of present bias), model estimations that rely on cross-sectional variation in time intervals will generally wrongly attribute the hyperbolic shape of discounting to present-biased preferences. This is related to the argument in Chakraborty et al. (2017) that noise or confusion spuriously drives estimates of present bias in convex budget experiments, though we show that complexity responses also bias estimates of present bias in choice list experiments. Taken together, these results suggest that in some of

Table 5: Summary of results across experiments

	Short-run impatience	Decreasing impatience	Sub-additivity	Front-end delay effect	Estimated present bias
Present in atemporal mirrors?	✓	✓	✓	–	✓
More pronounced with cognitive uncertainty?	✓	✓	✓	x	✓
More pronounced with choice inconsistency?	✓	✓	n/a	n/a	✓
More pronounced in difficult problems?	✓	✓	✓	–	✓

Notes. "✓" means that an anomaly is present / more pronounced, "x" that it is not present / not more pronounced and "–" that the opposite is present / the anomaly is less pronounced. "n/a" means that data limitations do not allow us to assess a relationship.

the field contexts in which $\beta < 1$ has been estimated, there may be normative scope for "correcting" time-inconsistent behavior by reducing complexity and / or increasing comprehension.

Moreover, our results shed light on the widely-discussed puzzle that demand for intertemporal commitment is typically considerably weaker than models of present bias would imply. If a large part of estimated present bias reflects noise and heuristics, there is no reason to expect it to generate a demand for commitment (Gabaix and Laibson, 2022). These results jive well with those in Carrera et al. (2022) who document that demand for commitment itself is also largely driven by noise.

This being said, we emphasize that our data clearly suggest that "true" present bias exists and is conceptually distinct from complexity responses. This suggests that intertemporal choice models would benefit from modeling two types of distortions: those arising from non-stationarities and those arising from complexity and the heuristic and noisy behaviors it inspires.

What makes intertemporal choice complex? Our data provides some clarity on what it is about intertemporal choice that makes it complex. One possibility, ex ante, is that the complexity of intertemporal choice is a consequence of the fact that it is difficult to introspectively evaluate or calculate one's own time preferences. However, the fact that hyperbolic discounting arises with near-full strength in atemporal mirrors instead suggests that complexity is driven by the difficulty of iterations discounting (successive multiplication). Appendix C provides further evidence in support of this idea. There, we document that both cognitive uncertainty and the variance of subjects' decisions strongly increase in the interval length or the number of iterations required to discount. This evidence suggests that repeated discounting is cognitively costly, producing noise that increases in the number of iterations required to discount a reward.

Related literature. Our work connects to several literatures on intertemporal choice (Thaler, 1981; Frederick et al., 2002; Cohen et al., 2020; Ericson and Laibson, 2019).

Most directly, our work relates to the literature that studies "cognitive effects," including work showing that patience tends to be correlated with measures of cognitive ability (e.g., Dohmen et al., 2010; Benjamin et al., 2013; Falk et al., 2018) and is sensitive to cognitive load, time pressure and framing (e.g., Ebert and Prelec, 2007; Imas et al., 2021; Dertwinkel-Kalt et al., 2021). Regenwetter et al. (2018) provide an overview of the psychology literature, which argues through model-fitting exercises that noise contributes to the hyperbolicity of discounting (e.g., He et al., 2019). One of our contributions is to provide substantially more direct evidence on the role of complexity and resulting noise / heuristics in generating the famous anomalies than can be achieved through model-fitting exercises.[18] Related to us are also experimental literatures showing that people have difficulty with exponential reasoning, suffering an "exponential growth bias" (Stango and Zinman, 2009). We too find errors in exponential reasoning, though unlike this literature we show that these errors generate (and predict) classic intertemporal choice anomalies.

Broader takeaways and connection to decision making in other domains. An important takeaway from all of the experiments reported in this paper is that complexity (and the heuristic / noisy responses that it induces) produces a *particular type of behavioral response*: an insufficient elasticity of decisions to variation in the main parameter of the problem, the length of the time interval.[19] This observation may suggest deep connections between intertemporal choice anomalies and other anomalies that have similarly been identified as growing out of complexity. In two recent papers, Oprea (2022) and Enke and Graeber (2022), we use similar methods to show that some of the core anomalies behavioral economists have observed in the domain of risk (such as probability weighting) are similarly rooted in complexity and the heuristic and / or noisy procedures it induces. In particular, an overarching message that emerges from the three papers is that, when decisions are complex, observed behavior is insufficiently sensitive with respect to variation in objective problem parameters, including probabilities, deterministic frequencies, time delays, and atemporal discounting iterations. Together, these papers thus suggest the possibility that many apparently distinct phenomena in behavioral economics might in fact be outgrowths of closely related forces, and that they might be parsimoniously united by models built to describe the way humans manage and respond to complexity.

[18]In psychology, Bulley et al. (2022) elicit confidence judgments in intertemporal decisions but they do not look at the choice anomalies that we focus on.

[19]See Ebert and Prelec (2007) for a related discussion.

References

Abeler, Johannes and Simon Jäger, "Complex Tax Incentives," *American Economic Journal: Economic Policy*, 2015, *7* (3), 1–28.

Agranov, Marina and Pietro Ortoleva, "Stochastic choice and preferences for randomization," *Journal of Political Economy*, 2017, *125* (1), 40–68.

_ and _ , "Ranges of Preferences and Randomization," *Working Paper*, 2020.

_ , Paul J Healy, and Kirby Nielsen, "Non-Random Randomization," *Available at SSRN*, 2020.

Andreoni, James and Charles Sprenger, "Risk preferences are not time preferences," *American Economic Review*, 2012, *102* (7), 3357–76.

_ , Christina Gravert, Michael A Kuhn, Silvia Saccardo, and Yang Yang, "Arbitrage or narrow bracketing? On using money to measure intertemporal preferences," Technical Report, National Bureau of Economic Research 2018.

Arts, Sara, Qiyan Ong, and Jianying Qiu, "Measuring subjective decision confidence," *Working Paper*, 2020.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger, "Working over time: Dynamic inconsistency in real effort tasks," *The Quarterly Journal of Economics*, 2015, *130* (3), 1067–1115.

Benjamin, Daniel J, Sebastian A Brown, and Jesse M Shapiro, "Who is 'behavioral'? Cognitive ability and anomalous preferences," *Journal of the European Economic Association*, 2013, *11* (6), 1231–1255.

Brocas, Isabelle, Juan D Carrillo, and Jorge Tarrasó, "How long is a minute?," *Games and Economic Behavior*, 2018, *111*, 305–322.

Bulley, Adam, Karolina M Lempert, Colin Conwell, Muireann Irish, and Daniel L Schacter, "Intertemporal choice reflects value comparison rather than self-control: insights from confidence judgements," *Philosophical Transactions of the Royal Society B*, 2022, *377* (1866), 20210338.

Carrera, Mariana, Heather Royer, Mark Stehr, Justin Sydnor, and Dmitry Taubinsky, "Who chooses commitment? Evidence and welfare implications," *The Review of Economic Studies*, 2022, *89* (3), 1205–1244.

Chabris, Christopher F., David Laibson, Carrie L. Morris, Jonathon P. Schuldt, and Dmitry Taubinsky, "Individual Laboratory-Measured Discount Rates Predict Field Behavior," *Journal of Risk and Uncertainty*, 2008, *37* (2-3), 237.

Chakraborty, Anujit, "Present bias," *Econometrica*, 2021, *89* (4), 1921–1961.

_ , Evan M Calford, Guidon Fenig, and Yoram Halevy, "External and internal consistency of choices made in convex time budgets," *Experimental economics*, 2017, *20* (3), 687–706.

_ , Yoram Halevy, and Kota Saito, "The relation between behavior under risk and over time," *American Economic Review: Insights*, 2020, *2* (1), 1–16.

Cohen, Jonathan, Keith Marzilli Ericson, David Laibson, and John Myles White, "Measuring time preferences," *Journal of Economic Literature*, 2020, *58* (2), 299–347.

Curry, David, "Uber Eats Revenue and Usage Statistics (2021)," 2021.

Dertwinkel-Kalt, Markus, Holger Gerhardt, Gerhard Riener, Frederik Schwerter, and Louis Strang, "Concentration bias in intertemporal choice," *Review of Economic Studies*, 2021.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde, "Are risk aversion and impatience related to cognitive ability?," *American Economic Review*, 2010, *100* (3), 1238–60.

Ebert, Jane EJ and Drazen Prelec, "The fragility of time: Time-insensitivity and valuation of the near and far future," *Management science*, 2007, *53* (9), 1423–1438.

Enke, Benjamin and Thomas Graeber, "Cognitive uncertainty," Technical Report, National Bureau of Economic Research 2022.

Epper, Thomas, Ernst Fehr, Helga Fehr-Duda, Claus Thustrup Kreiner, David Dreyer Lassen, Søren Leth-Petersen, and Gregers Nytoft Rasmussen, "Time discounting and wealth inequality," *American Economic Review*, 2020, *110* (4), 1177–1205.

Ericson, Keith Marzilli and David Laibson, "Intertemporal choice," in "Handbook of Behavioral Economics: Applications and Foundations 1," Vol. 2, Elsevier, 2019, pp. 1–67.

_ , John Myles White, David Laibson, and Jonathan D Cohen, "Money earlier or later? Simple heuristics explain intertemporal choices better than delay discounting does," *Psychological science*, 2015, *26* (6), 826–833.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde, "Global evidence on economic preferences," *The Quarterly Journal of Economics*, 2018, *133* (4), 1645–1692.

Frederick, Shane, George Loewenstein, and Ted O'donoghue, "Time discounting and time preference: A critical review," *Journal of economic literature*, 2002, *40* (2), 351–401.

Fudenberg, Drew and David K Levine, "A dual-self model of impulse control," *American economic review*, 2006, *96* (5), 1449–1476.

Gabaix, Xavier and David Laibson, "Myopia and discounting," Technical Report, National bureau of economic research 2022.

Gershman, Samuel J and Rahul Bhui, "Rationally inattentive intertemporal choice," *Nature Communications*, 2020, *11*.

Gul, Faruk and Wolfgang Pesendorfer, "Temptation and self-control," *Econometrica*, 2001, *69* (6), 1403–1435.

Halevy, Yoram, "Strotz meets Allais: Diminishing impatience and the certainty effect," *American Economic Review*, 2008, *98* (3), 1145–62.

_ , "Some comments on the use of monetary and primary rewards in the measurement of time preferences," *Unpublished manuscript*, 2014.

_ , "Time consistency: Stationarity and time invariance," *Econometrica*, 2015, *83* (1), 335–352.

He, Lisheng, Russell Golman, and Sudeep Bhatia, "Variable time preference," *Cognitive psychology*, 2019, *111*, 53–79.

Imas, Alex, Michael Kuhn, and Vera Mironova, "Waiting to choose: The Role of Deliberation in Intertemporal Choice," *American Economic Journal: Microeconomics*, 2021.

Jackson, Matthew O and Leeat Yariv, "Present bias and collective dynamic choice in the lab," *American Economic Review*, 2014, *104* (12), 4184–4204.

Jang, Minkwang and Oleg Urminsky, "Cross-Period Impatience: Subjective Financial Periods Explain Time-Inconsistent Choices," *Available at SSRN*, 2022.

Kable, Joseph W and Paul W Glimcher, "An "as soon as possible" effect in human intertemporal decision making: behavioral evidence and neural mechanisms," *Journal of neurophysiology*, 2010, *103* (5), 2513–2531.

Khaw, Mel Win, Ziang Li, and Michael Woodford, "Cognitive imprecision and small-stakes risk aversion," *The review of economic studies*, 2021, *88* (4), 1979–2013.

Laibson, David, "Golden eggs and hyperbolic discounting," *The Quarterly Journal of Economics*, 1997, *112* (2), 443–478.

_ , "Why don't present-biased agents make commitments?," *American Economic Review*, 2015, *105* (5), 267–272.

Loewenstein, George and Drazen Prelec, "Anomalies in intertemporal choice: Evidence and an interpretation," *The Quarterly Journal of Economics*, 1992, *107* (2), 573–597.

Lu, Jay and Kota Saito, "Random intertemporal choice," *Journal of Economic Theory*, 2018, *177*, 780–815.

O'Donoghue, Ted and Matthew Rabin, "Doing it now or later," *American Economic Review*, 1999, *89* (1), 103–124.

Oprea, Ryan, "Simplicity Equivalents," *Working Paper*, 2022.

Read, Daniel, "Is time-discounting hyperbolic or subadditive?," *Journal of risk and uncertainty*, 2001, *23* (1), 5–32.

_ **, Shane Frederick, and Marc Scholten**, "DRIFT: an analysis of outcome framing in intertemporal choice.," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2013, *39* (2), 573.

Regenwetter, Michel, Daniel R Cavagnaro, Anna Popova, Ying Guo, Chris Zwilling, Shiau Hong Lim, and Jeffrey R Stevens, "Heterogeneity and parsimony in intertemporal choice.," *Decision*, 2018, *5* (2), 63.

Rubinstein, Ariel, ""Economics and psychology"? The case of hyperbolic discounting," *International Economic Review*, 2003, *44* (4), 1207–1216.

Stango, Victor and Jonathan Zinman, "Exponential growth bias and household finance," *The Journal of Finance*, 2009, *64* (6), 2807–2849.

Thaler, Richard, "Some empirical evidence on dynamic inconsistency," *Economics letters*, 1981, *8* (3), 201–207.

Vieider, Ferdinand M, "Noisy coding of time and reward discounting," Technical Report, Ghent University, Faculty of Economics and Business Administration 2021.

Weitzman, Martin L, "Gamma discounting," *American Economic Review*, 2001, *91* (1), 260–271.

Zauberman, Gal, B Kyu Kim, Selin A Malkoc, and James R Bettman, "Discounting time and time discounting: Subjective time perception and intertemporal preferences," *Journal of Marketing Research*, 2009, *46* (4), 543–556.

Online Appendix

A Additional Figures

	Option A	Option B
1	$42.00 in 12 months	$2.00 now
2	$42.00 in 12 months	$4.00 now
3	$42.00 in 12 months	$6.00 now
4	$42.00 in 12 months	$8.00 now
5	$42.00 in 12 months	$10.00 now
6	$42.00 in 12 months	$12.00 now
7	$42.00 in 12 months	$14.00 now
8	$42.00 in 12 months	$16.00 now
9	$42.00 in 12 months	$18.00 now
10	$42.00 in 12 months	$20.00 now
11	$42.00 in 12 months	$22.00 now
12	$42.00 in 12 months	$24.00 now
13	$42.00 in 12 months	$26.00 now
14	$42.00 in 12 months	$28.00 now
15	$42.00 in 12 months	$30.00 now
16	$42.00 in 12 months	$32.00 now
17	$42.00 in 12 months	$34.00 now
18	$42.00 in 12 months	$36.00 now
19	$42.00 in 12 months	$38.00 now
20	$42.00 in 12 months	$40.00 now
21	$42.00 in 12 months	$42.00 now

a) *Delay* treatment

	Option A	Option B
1	$42.00 shrunk 12 times	$2.00
2	$42.00 shrunk 12 times	$4.00
3	$42.00 shrunk 12 times	$6.00
4	$42.00 shrunk 12 times	$8.00
5	$42.00 shrunk 12 times	$10.00
6	$42.00 shrunk 12 times	$12.00
7	$42.00 shrunk 12 times	$14.00
8	$42.00 shrunk 12 times	$16.00
9	$42.00 shrunk 12 times	$18.00
10	$42.00 shrunk 12 times	$20.00
11	$42.00 shrunk 12 times	$22.00
12	$42.00 shrunk 12 times	$24.00
13	$42.00 shrunk 12 times	$26.00
14	$42.00 shrunk 12 times	$28.00
15	$42.00 shrunk 12 times	$30.00
16	$42.00 shrunk 12 times	$32.00
17	$42.00 shrunk 12 times	$34.00
18	$42.00 shrunk 12 times	$36.00
19	$42.00 shrunk 12 times	$38.00
20	$42.00 shrunk 12 times	$40.00
21	$42.00 shrunk 12 times	$42.00

b) *Mirror* treatment

Figure 8: Screenshots from the experimental software.

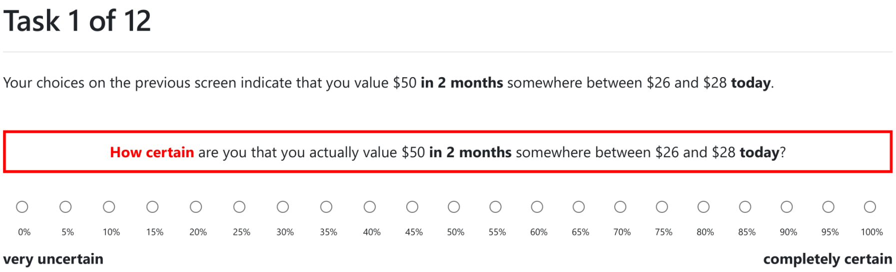

Figure 9: Screenshot of an example cognitive uncertainty elicitation screen in *Delay-M*

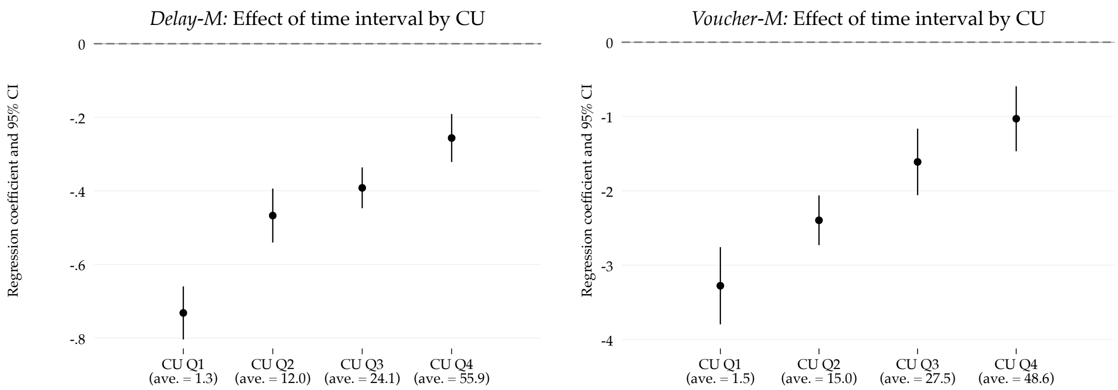

Figure 10: Coefficients from regressions of normalized indifference points on time interval, split by CU quartiles; left: *Delay-M*; right: *Voucher-M*.

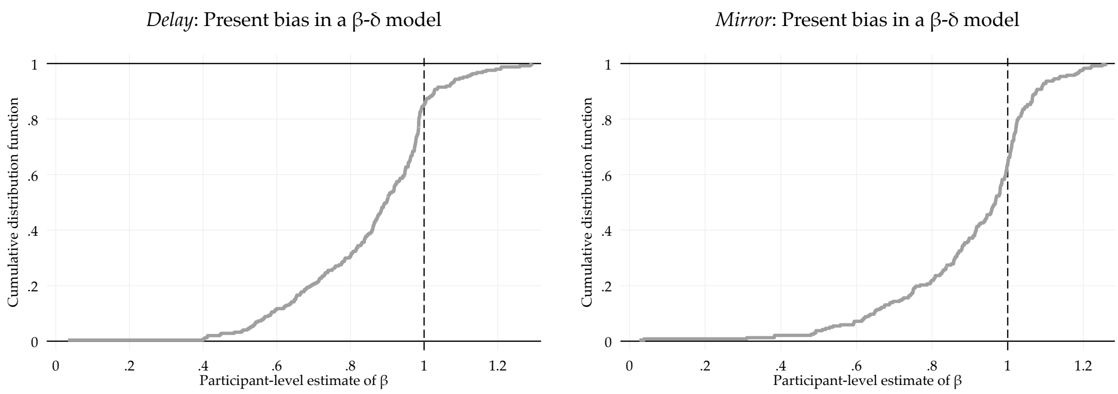

Figure 11: Empirical CDFs of individual-level estimates of a $\beta - \delta$ model (eq. (2)) in *Delay* ($N = 254$) and *Mirror* ($N = 246$), using first-assigned treatment only. Non-linear least squares estimation based on 18 decisions from each individual. For ease of readability we exclude subjects with $\hat{\beta} > 1.3$.

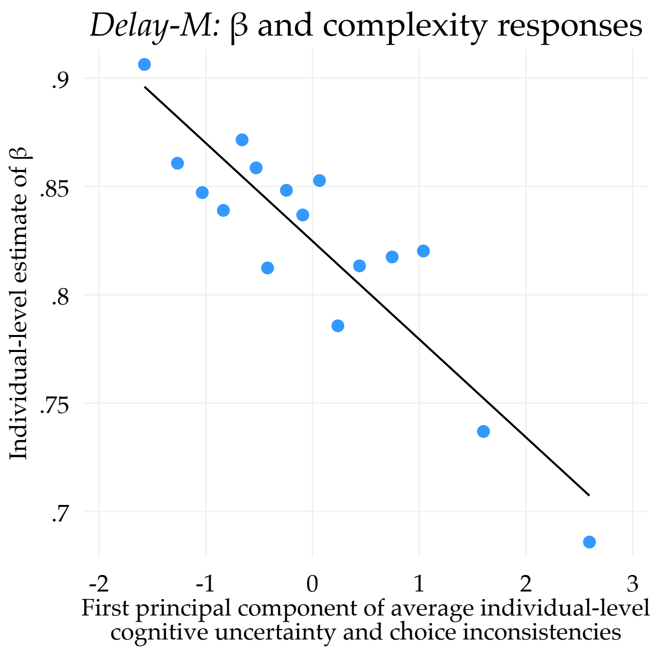

Figure 12: Binned scatter plot of individual-level $\hat{\beta}_{ST}$ in eq. (2) against first principal component of subject-level average CU and choice inconsistencies. In this figure the individual-level estimate of β is derived taking into account utility curvature, which is separately estimated at the population level based on lottery choice lists. Based on 643 subjects in *Delay-M*. The figure excludes subjects for whom we estimate $\hat{\beta} > 2$.

B Additional Tables

Table 6: Anomalies in *Delay* and *Mirror*

	Dependent variable: Implied annual impatience (in %)					
Phenomenon:	Decreasing impatience		Subadditivity		Front-end delay	
Treatment:	*Delay*	*Mirror*	*Delay*	*Mirror*	*Delay*	*Mirror*
	(1)	(2)	(3)	(4)	(5)	(6)
Time interval / number of discounting steps (in years)	-5.76***	-5.14***				
	(0.25)	(0.26)				
1 if one long interval			-7.57***	-9.93***		
			(1.38)	(1.16)		
1 if front end delay					-4.24**	3.79**
					(1.85)	(1.69)
Payment amount FE	Yes	Yes	Yes	Yes	Yes	Yes
Subadditivity set FE	No	No	Yes	Yes	Yes	Yes
Observations	4572	4428	508	492	508	492
R^2	0.17	0.19	0.09	0.06	0.07	0.02

Notes. OLS estimates, robust standard errors (in parentheses) are clustered at the subject level. In columns (1) and (2), the sample consists of all decisions in the respective treatment. In columns (3) and (4), the sample consists of two observations per subject: their implied annual discounting over the long (composite) interval and their implied discounting over the two shorter intervals that have a subadditivity structure. In columns (5) and (6), the sample includes those two decisions per subject that have a front-end delay structure. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Anomalies in *Delay* and *Mirror*, pooling first-assigned and second-assigned treamtents

	Dependent variable: Implied annual impatience (in %)					
Phenomenon:	Decreasing impatience		Subadditivity		Front-end delay	
Treatment:	*Delay*	*Mirror*	*Delay*	*Mirror*	*Delay*	*Mirror*
	(1)	(2)	(3)	(4)	(5)	(6)
Time interval / number of discounting steps (in years)	-6.11***	-4.68***				
	(0.18)	(0.18)				
1 if one long interval			-8.57***	-10.0***		
			(0.89)	(0.81)		
1 if front end delay					-4.59***	5.41***
					(1.28)	(1.21)
Payment amount FE	Yes	Yes	Yes	Yes	Yes	Yes
Subadditivity set FE	No	No	Yes	Yes	Yes	Yes
Observations	9000	8999	1000	1000	1000	1000
R^2	0.19	0.17	0.05	0.06	0.03	0.02

Notes. OLS estimates, robust standard errors (in parentheses) are clustered at the subject level. In columns (1) and (2), the sample consists of all decisions in the respective treatment. In columns (3) and (4), the sample consists of two observations per subject: their implied annual discounting over the long (composite) interval and their implied discounting over the two shorter intervals. In columns (5) and (6), the sample includes those two decisions per subject that have a front-end delay structure. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Decreasing impatience in *Delay-M* and within-subject variation of cognitive uncertainty

	Dependent variable: Implied annual impatience (in %)	
Dataset:	*Delay-M*	
Phenomenon:	Decreasing impatience	
	(1)	(2)
Time interval	-6.87***	-5.46***
	(0.18)	(0.26)
Cognitive uncertainty (standard. within subject)		0.28***
		(0.04)
Time interval × Cognitive uncertainty (standard. within subject)		-0.068***
		(0.01)
Payment amount FE	Yes	Yes
Observations	7740	7740
R^2	0.16	0.18

Notes. OLS estimates, robust standard errors (in parentheses) are clustered at the subject level. In these regressions, the measure of cognitive uncertainty was standardized at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Choice inconsistencies in *Mirror*

	Dependent variable: Implied annual impatience (in %)	
Phenomenon:	Short-run impatience	Decreasing impatience
	(1)	(2)
Inconsistent decision	17.8***	12.7***
	(3.86)	(3.90)
Number of discounting steps (in years)		-2.64***
		(0.51)
Number of discounting steps (in years) × Inconsistent decision		-3.07***
		(0.53)
Payment amount FE	Yes	Yes
Observations	417	3408
R^2	0.09	0.21

Notes. OLS estimates, robust standard errors (in parentheses) are clustered at the subject level. Regressions include sets of repeated decisions shown to a subject. Column (1) includes decisions with one discounting iteration only, column (2) includes decisions involving any number of iterations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Anomalies in the *Delay-M* vs. the *Opaque* treatments

Phenomenon:	Manipulation check		Decr. imp.	Subadd.	Front-end
	Dependent variable:				
	CU	Inconsistent	Implied annual impatience (in %)		
	(1)	(2)	(3)	(4)	(5)
Opaque treatments	13.6***	0.065**	0.99	3.07	-0.59
	(1.40)	(0.03)	(1.89)	(2.20)	(2.35)
Time interval			-6.88***		
			(0.18)		
Time interval × Opaque treatments			-1.92***		
			(0.34)		
1 if one long interval				-8.58***	
				(0.63)	
1 if one long interval × Opaque treatments				-7.99***	
				(1.30)	
1 if front-end delay					-3.06***
					(0.99)
1 if front-end delay × Opaque treatments					5.32***
					(1.86)
Payment amount FE	Yes	Yes	Yes	Yes	Yes
Subadditivity set FE	No	No	No	Yes	Yes
Observations	11364	3788	11364	2818	3465
R^2	0.06	0.01	0.18	0.04	0.01

Notes. OLS estimates, robust standard errors (in parentheses) are clustered at the subject level. In columns (1) and (3), the sample consists of all decisions in the *Delay-M* and *Opaque* treatments. In column (2), the sample includes all sets of repeated decisions shown to a subject. In column (4), the sample consists of two observations per subject: their implied annual discounting over the long (composite) interval and their implied discounting over the two shorter intervals. In column (5), the sample includes those two decisions per subject that have a front-end delay structure. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Population-level estimates of $\beta - \delta$ model

	Delay & Mirror		*Delay-M*					*Opaque*	*Voucher-M*				
	Delay (1)	Mirror (2)	All (3)	CU=0 (4)	CU>0 (5)	Incons.=0 (6)	Incons.>0 (7)	All (8)	All (9)	CU=0 (10)	CU>0 (11)	Incons.=0 (12)	Incons.>0 (13)
$\hat{\beta}$	.774	.846	.76	.872	.721	.822	.75	.72	.882	.953	.854	.957	.865
$\hat{\delta}$	.982	.96	.978	.973	.98	.983	.977	.989	.942	.955	.941	.968	.936

Notes. Population-level estimates of a $\beta - \delta$ model (eq. (2)). Columns (1) and (2) use the first-assigned treatment only, based on $N = 254$ subjects in *Delay* and $N = 246$ subjects in *Mirror*. Columns (3), (8) and (9) include all subjects in the respective treatments: $N = 645$ in *Delay-M*, $N = 302$ in *Opaque* and $N = 500$ in *Voucher-M*. All other columns are based on sample splits of the corresponding treatments. Non-linear least squares estimates.

C What Makes Intertemporal Choice Complex?

This Appendix tentatively investigates *what* it is about intertemporal choice that makes it complex, and therefore vulnerable to noisy or heuristic decision-making. One possibility, ex ante, is that complexity is a consequence of the fact that it is difficult to introspectively evaluate or calculate one's own time preferences (e.g., one's discount factor). Similarly, another ex ante possibility is that complexity is an outgrowth of the difficulty of integrating one's risk and time preferences to inform choice.

Results from our *Mirror* treatment (in which time preferences are clearly induced and risk and time preferences needn't be integrated), suggest an alternative possibility: that the complexity of intertemporal choice is instead a direct outgrowth of the costs and difficulties of iteratively discounting rewards, which requires an intensive type of recursive reasoning. If true, we would expect the number of required steps of discounting / a longer time delay to be associated with more pronounced complexity responses.[20]

To examine this, re-reconsider equation (2). Rearranging, taking logs and adding a mean-zero noise term yields that a subject's observed indifference point in our experiments can be expressed as

$$ln\left(\frac{x_1}{x_2}\right) = ln(\beta_{t_1=0}) + \Delta t \cdot ln(\delta) + \varepsilon. \tag{3}$$

where the first term on the right-hand side collapses to zero if $\beta = 1$. Importantly, our hypothesis that complexity responses (noisy or heuristic procedures) increase in the delay implies that $Var(\varepsilon)$ should not be constant but, instead, heteroscedastic and increasing in the delay. Because in equation (3) a subject's log normalized indifference point is a linear function of the delay, the equation can be estimated using simple OLS. We run this regression and then inspect the variance of the regression residuals.

The top left panel of Figure 13 shows the results for treatment *Delay*. We find that the variance of the regression residuals indeed strongly increases in the length of the delay. A different way of saying this is that the variance of subjects' normalized indifference points strongly increases in the delay.

The top right panel shows an analogous plot for treatment *Mirror*, where the x-axis now represents the required number of steps of discounting. Again, we see strong evidence of heteroscedasticity, in line with the hypothesis that complexity responses become more pronounced as the number of discounting steps increases.

In a standard exponential discounting model with preference heterogeneity, the regression residuals or the variance of log indifference points *should* increase in the delay[21]

[20]Some models of complexity and intertemporal discounting directly consider this possibility: Gabaix and Laibson (2022) model a decision maker whose degree of cognitive noisiness increases in the delay.

[21]With exponential discounting and linear utility, $Var[ln(x_1/x_2)] = (\Delta t)^2 Var[ln(\delta)] + Var(\epsilon)$.

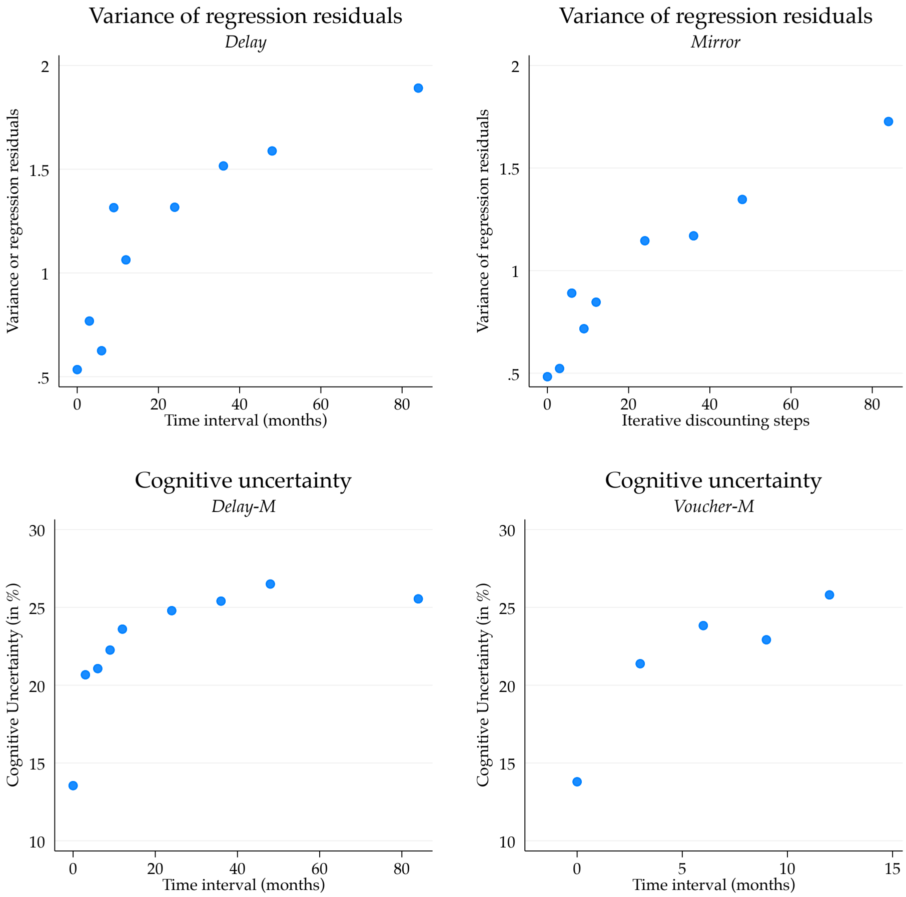

Figure 13: Noisiness as a function of the delay. Top panels show the variance of the regression residuals of eq. (3) in *Delay* and *Mirror*. Bottom panels show average cognitive uncertainty in *Delay Noise* and *Voucher Noise*. In all panels, delays are rounded to the nearest multiple of three.

However, in treatment *Mirror*, where the increase is almost equally strong, there is no preference heterogeneity available to rationalize the pattern because we experimentally induced the same discount factor for all subjects. In *Mirror*, this pattern must be driven by increasingly idiosyncratic responses to complexity as the number of steps of discounting increases. The fact that that the pattern (including magnitudes) is almost identical in *Delay* suggests the same complexity-based explanation likely applies there as well. Moreover, recall that decisions in *Delay* and *Mirror* are highly correlated within subject, providing further suggestive evidence that the increase in the variance of decisions has the same origin, which cannot be heterogeneity in discount factors.

The bottom panels of Figure 13 provide additional evidence in suport of this claim. We plot subjects' cognitive uncertainty as a function of the delay in treatments *Delay Noise* and *Voucher Noise*.[22] In both treatments, people report being much more uncertain

[22]Analyzing how choice inconsistencies vary with the delay is confounded by the relationship between choice inconsistency and the "extremity" of the intertemporal decision problem. In all treatments, we find that subjects exhibit less inconsistency when the delay is either very short or very long, in large part

about which decision to take as the delay gets longer. Going from very short delays of less than one month to delays of seven years, CU more than doubles. This increase is concave, with CU barely increasing for delays longer than 1–2 years (recall that in *Voucher Noise* the longest delay is one year).

Taken together, multiple streams of evidence suggest that the difficulty of decision-making increases in the length of the delay / the number of discounting steps required. This, when combined with the appearance of anomalies in the atemporal mirrors (where there is little to drive complexity except the difficulty of iterative discounting), is indicative that an important source of complexity in intertemporal decision-making is the cognitive act of iteratively discounting future rewards.

Of course, the insight that complexity increases in the number of cognitive steps required to discount does not imply that complexity is zero for very short delays. For instance, as Figure 13 shows, there is substantial CU even for delays of one month and less, consistent with people exhibiting noise-driven extreme short-run impatience.

because in these decision problems a large share of subjects make boundary choices that artifically make them look perfectly consistent.

D Experimental Instructions

D.1 Instructions for *Delay & Mirror* Experiment

D.1.1 First-assigned treatment: *Delay*

Delayed Choices

In this part of the study you will **choose between various hypothetical payments, which pay different amounts at different points in time**. An example decision is between the following two hypothetical payments.

Option A	Option B
$100.00 in 3 months	$90.00 now

In this example we are asking you (hypothetically) would you rather be paid $100 in three months (Option A) or $90 right now (Option B).

For all hypothetical payments in this study, please treat them as if you know you will receive them with certainty, even if they are delayed. That is, please assume there is no risk that you wouldn't actually get paid. Further, assume all payments were made by leaving a check in your mailbox which you can cash at the specified date.

For this part of the experiment, there are no right wrong answers, because how much you like an option depends on your personal taste. Just try your best to think hard about what you'd really prefer.

The Choice List

On your decision screen, you will be asked to choose which of two payment options you prefer. You will see choice lists such as the one below, where **each row is a separate choice**.

In every list, the left-hand option (Option A) is a delayed payment that is identical in all rows. The right-hand side option (Option B) is a payment with an *earlier payment date than Option A*. The earlier, right-hand side payment increases as you go down the list. To make a choice just click on the option you prefer for each choice (i.e. for each row), highlighting your choice yellow. **An effective way to complete these choice lists is to determine in which row you like to switch from preferring Option A to preferring Option B.**

	Option A	Option B
1	$40.00 in 3 months	$2.00 now
2	$40.00 in 3 months	$4.00 now
3	$40.00 in 3 months	$6.00 now
4	$40.00 in 3 months	$8.00 now
5	$40.00 in 3 months	$10.00 now
6	$40.00 in 3 months	$12.00 now
7	$40.00 in 3 months	$14.00 now
8	$40.00 in 3 months	$16.00 now
9	$40.00 in 3 months	$18.00 now
10	$40.00 in 3 months	$20.00 now
11	$40.00 in 3 months	$22.00 now
12	$40.00 in 3 months	$24.00 now
13	$40.00 in 3 months	$26.00 now
14	$40.00 in 3 months	$28.00 now
15	$40.00 in 3 months	$30.00 now
16	$40.00 in 3 months	$32.00 now
17	$40.00 in 3 months	$34.00 now
18	$40.00 in 3 months	$36.00 now
19	$40.00 in 3 months	$38.00 now
20	$40.00 in 3 months	$40.00 now

Auto-completion: The table auto-completes your choices so you don't have to click through all of the rows. You do not have to start at the top of the table. If you select Option A in any one row, we assume you will also prefer Option A in all rows *above* that row. If you select Option B in any one row, we assume that you will also prefer Option B in all rows *below* that row.

Next Part

In the next part of the experiment, we are going to have you make a **very different kind of decision**, also using choice lists.

Instead of making hypothetical decision about money paid out at various points in time (as in Part 1) **we will have you make real money decisions paid as a bonus today.** Specifically, we will ask you to choose between monetary amounts that are shrunk to varying degrees, using a choice list like the one you used in Part 1. The difference is, we will really pay some of you these amounts today!

D.1.2 First-assigned treatment: *Mirror*

Shrunk Choices

In this part of the study you will **choose between various payments (actually paid to you today), which will first be shrunk (reduced in value) some number of times.** An example decision is between the following two payments.

Option A	Option B
$100.00 shrunk 3 times	$90.00

Each time a payment is shrunk (as in Option A), its dollar value falls by 4% meaning it shrinks to only 96% of the dollar value from the previous step. For example

- If $100 is shrunk only 1 time, we would pay you 96% of $100 or $96.
- If $100 is shrunk in only 2 time, we would pay you 96% of 96% of $100 or $92.16
- If $100 is shrunk in only 3 time, we would pay you 96% of 96% of 96% of $100 or $88.47

And so on. So, in the example, if you chose Option A ($100 shrunk 3 times), you would earn $88.47. On the other hand, Option B isn't shrunk at all so it just pays the $90 shown (any time we don't mention shrinking for a payment, that means the payment is not shrunk at all).

At the end of the experiment we will randomly select 20% of participants to actually be paid their earnings as a bonus today from a randomly selected choice.

The Choice List

On your decision screen, you will be asked to choose which of two payment options you prefer. You will see choice lists such as the one below, where **each row is a separate choice**.

In every list, the left-hand option (Option A) is a delayed payment that is identical in all rows. The right-hand side option (Option B) is a payment with an *earlier payment date than Option A*. The earlier, right-hand side payment increases as you go down the list. To make a choice just click on the option you prefer for each choice (i.e. for each row), highlighting your choice yellow. **An effective way to complete these choice lists is to determine in which row you like to switch from preferring Option A to preferring Option B.**

	Option A	Option B
1	$40.00 in 3 months	$2.00 now
2	$40.00 in 3 months	$4.00 now
3	$40.00 in 3 months	$6.00 now
4	$40.00 in 3 months	$8.00 now
5	$40.00 in 3 months	$10.00 now
6	$40.00 in 3 months	$12.00 now
7	$40.00 in 3 months	$14.00 now
8	$40.00 in 3 months	$16.00 now
9	$40.00 in 3 months	$18.00 now
10	$40.00 in 3 months	$20.00 now
11	$40.00 in 3 months	$22.00 now
12	$40.00 in 3 months	$24.00 now
13	$40.00 in 3 months	$26.00 now
14	$40.00 in 3 months	$28.00 now
18	$40.00 in 3 months	$36.00 now
19	$40.00 in 3 months	$38.00 now
20	$40.00 in 3 months	$40.00 now

Auto-completion: The table auto-completes your choices so you don't have to click through all of the rows. You do not have to start at the top of the table. If you select Option A in any one row, we assume you will also prefer Option A in all rows *above* that row. If you select Option B in any one row, we assume that you will also prefer Option B in all rows *below* that row.

Next Part

In the next part of the experiment, we are going to have you make a **very different kind of decision**, also using choice lists.

Instead of making real money decisions about money shrunk to various degrees (as in Part 1) **we will have you make hypothetical money decisions about money amounts paid to you at various points in time.** Specifically, we will ask you to choose between monetary amounts paid sooner versus later, using a choice list like the one you used in Part 1. We won't actually pay you based on your choices in this part, but just want to understand when you'd hypothetically rather be paid various combinations of money.

D.2 Instructions for *Delay Noise*

Part 1 of this study: Instructions (1/3)

Please read these instructions carefully. There will be comprehension checks. If you fail these checks, you will immediately be excluded from the study and you will not receive the completion payment.

In this part of the study, you will **choose between various hypothetical payments, which pay different amounts of money at different points in time**. An example decision is between the following two hypothetical payments:

In 30 days: $ 40 OR **Today: $ 12**

For all hypothetical payments in this study, please treat them as if you knew that you would receive them with certainty, even if they are delayed. That is, please assume that there is no risk that you wouldn't actually get paid. Further assume that all payments were made by leaving a check in your mailbox.

Throughout the experiment, there are no right or wrong answers, because how much you like an option depends on your personal taste. There will be two types of decision screens.

Decision screen 1

On decision screen 1, you will be asked to choose which of two payment options you prefer. You will see choice lists such as the one below, where each row is a separate choice. In every list, the left-hand side option (Option A) is a delayed payment that is identical in all rows. The right-hand side option (Option B) is a payment *with an earlier payment date than Option A*. The earlier, right-hand side payment increases as you go down the list. **An effective way to complete these choice lists is to determine in which row you would like to switch from preferring Option A to preferring Option B.**

Based on where you switch from Option A to Option B in this list, we assess which amount at the early payment date (Option B) you value as much as the amount specified at the later payment date (Option A). For example, in the choice list below, you would value $40 in 30 days somewhere between $12 and $14 today, because this is where switching occurs.

Option A			Option B
	◉	○	Today: $2
	◉	○	Today: $4
	◉	○	Today: $6
	◉	○	Today: $8
	◉	○	Today: $10
	◉	○	Today: $12
	○	◉	Today: $14
	○	◉	Today: $16
	○	◉	Today: $18
In 30 days: $40	○	◉	Today: $20
	○	◉	Today: $22
	○	◉	Today: $24
	○	◉	Today: $26
	○	◉	Today: $28
	○	◉	Today: $30
	○	◉	Today: $32
	○	◉	Today: $34
	○	◉	Today: $36
	○	◉	Today: $38
	○	◉	Today: $40

On the next page, you will see an example choice list, and you can practice making your selections.

Click "Next" to proceed to the example page.

Part 1 of this study: Instructions (2/3)

Auto-completion: The table auto-completes your choices so you don't have to click through all of the rows. You do not have to start at the top of the table. If you select Option A in any one row, we assume that you will also prefer Option A in all rows *above* that row. If you select Option B in any one row, we assume that you will also prefer Option B in all rows *below* that row.

Option A			Option B
In 30 days: $40	○	○	Today: $2
	○	○	Today: $4
	○	○	Today: $6
	○	○	Today: $8
	○	○	Today: $10
	○	○	Today: $12
	○	○	Today: $14
	○	○	Today: $16
	○	○	Today: $18
	○	○	Today: $20
	○	○	Today: $22
	○	○	Today: $24
	○	○	Today: $26
	○	○	Today: $28
	○	○	Today: $30
	○	○	Today: $32
	○	○	Today: $34
	○	○	Today: $36
	○	○	Today: $38
	○	○	Today: $40

Part 1 of this study: Instructions (3/3)

Decision screen 2

When you fill out a choice list, you may feel **uncertain about whether you prefer the left or right payment option**. On decision screen 2, we will ask you to select a button to indicate **how certain** you are how much money the larger later payment is worth to you in terms of dollars at the earlier payment date.

In answering this question, we ask you to assume that you would receive both payment options with certainty. We are interested in **your uncertainty about your own preferences regarding these payments**, not in your potential uncertainty about whether you would actually receive the money.

Example

Suppose that on the first decision screen you indicated that you valued $40 in 30 days somewhere between $12 and $14 today. Your second decision screen would look like this.

How certain are you that you actually value $40 **in 30 days** somewhere between $12 and $14 **today**?

○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%

very uncertain **completely certain**

Comprehension questions

The questions below test your understanding of the instructions.

Important: If you fail to answer any one of these questions correctly, you will not be allowed to participate in the study.

1. Which of the following statements is true?
   - ○ In making my decisions, I am asked to assume that I will actually receive all payments as indicated, regardless of whether they take place now or in the future.
   - ○ In making my decisions, I am asked to assume that it is less likely that I will actually receive payments that are meant to take place in the future.
   - ○ In making my decisions, I am asked to assume that it is less likely that I will actually receive payments that are meant to take place now.

2. Suppose you are 80% certain that your decisions actually correspond to how much the different choice options are worth to you. Which button should you click in this case?

○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%

very uncertain **completely certain**

3. When we ask you how certain you are about how much different payments are worth to you at different points in time, then which type of uncertainty are we interested in?
   - ○ Uncertainty about whether I would actually receive the payments.
   - ○ Uncertainty about how much I value the payments, assuming that I know I would receive them with certainty.

D.3 Instructions for *Voucher Noise*

Part 1 of this study: Instructions (1/4)

Please read these instructions carefully. There will be comprehension checks. If you fail these checks, we will have to exclude you from the study and you will not receive the completion payment.

In this part of the study, you will **choose between different UberEats food delivery vouchers. These vouchers will vary along two dimensions**:

- **The vouchers will have different values**
- **The vouchers will be valid at different points in time**

How do the vouchers work?

Each voucher is valid for food delivery during a period of only seven days. A voucher can be used starting **from the indicated date**, and **it remains valid for exactly 7 days after** that date. Specifically, the vouchers work as follows:

- If you win a voucher, you will be informed about the voucher amount and the validity period on the last page of this study. You will then be asked to provide an email address associated with an UberEats account. The voucher will directly be credited to the corresponding UberEats account within the next 10 hours.
- However, the voucher amount **can only be spent during the validity period** of the voucher.
- Vouchers can be used to order from the entire range of restaurants, cafes, and bars that partner with UberEats in your area.
- You do not need to worry about forgetting the validity period: **UberEats will automatically send reminders** about your voucher 24 hours before the validity period starts and 24 hours before it ends.

What decisions will you be asked to make?

An example decision is between the following two vouchers:

Valid in 30 days: $40 Voucher OR **Valid today: $20 Voucher**

The left-hand side voucher carries an amount of $40 and can be spent in the 7-day period starting in 30 days from now. The right-hand side voucher is for an amount of only $20, but can be spent in the 7-day period starting immediately.

Throughout the experiment, there are no right or wrong answers because how much you like a voucher depends on your personal taste.

Part 1 of this study: Instructions (2/4)

Decision screen 1

On decision screen 1, you will be asked to choose which of two vouchers you prefer. You will see choice lists such as the one below, where each row is a separate choice. In every list, the left-hand side option (Voucher A) is a voucher that is identical in all rows. The right-hand side option (Voucher B) is a voucher *with an earlier validity period than Voucher A*. The amount associated with the earlier, right-hand side voucher increases as you go down the list. **An effective way to complete these choice lists is to determine in which row you would like to switch from preferring Voucher A to preferring Voucher B.**

Based on where you switch from Voucher A to Voucher B in this list, we assess which voucher amount in the early validity period (Voucher B) you value as much as the voucher amount specified in the later validity period (Voucher A). For example, in the choice list below, you would value a $40 voucher that is valid in 30 days somewhere between a $12 and a $14 voucher that is valid today, because this is where switching occurs.

Voucher A			Voucher B
	◉	○	Valid Today: $2 Voucher
	◉	○	Valid Today: $4 Voucher
	◉	○	Valid Today: $6 Voucher
	◉	○	Valid Today: $8 Voucher
	◉	○	Valid Today: $10 Voucher
	◉	○	Valid Today: $12 Voucher
	○	◉	Valid Today: $14 Voucher
	○	◉	Valid Today: $16 Voucher
	○	◉	Valid Today: $18 Voucher
Valid In 30 days: $40 Voucher	○	◉	Valid Today: $20 Voucher
	○	◉	Valid Today: $22 Voucher
	○	◉	Valid Today: $24 Voucher
	○	◉	Valid Today: $26 Voucher
	○	◉	Valid Today: $28 Voucher
	○	◉	Valid Today: $30 Voucher
	○	◉	Valid Today: $32 Voucher
	○	◉	Valid Today: $34 Voucher
	○	◉	Valid Today: $36 Voucher
	○	◉	Valid Today: $38 Voucher
	○	◉	Valid Today: $40 Voucher

If you are selected to receive an additional reward from part 1 of the study, your reward will be determined as follows:
Your choice in a randomly selected row of a randomly selected choice list determines the amount of your personal voucher. Each choice list and each row are equally likely to get selected.

Important:

- Your choices may matter for real money! If you are selected to receive a bonus, one of your choices will actually be implemented, and your decision will determine which type of voucher you receive.
- Since only one of your decisions will be randomly selected to count, you should consider each choice list independently of the others. There is no point in strategizing across decisions.

On the next page, you will see an example choice list, and you can practice making your selections.

Click "Next" to proceed to the example page.

Part 1 of this study: Instructions (3/4)

Auto-completion: The table auto-completes your choices so you don't have to click through all of the rows. You do not have to start at the top of the table. If you select Voucher A in any one row, we assume that you will also prefer Voucher A in all *above* that row. If you select Voucher B in any one row, we assume that you will also prefer Voucher B in all rows *below* that row.

Reminder: both vouchers are valid for 7 days starting on the day indicated for each voucher.

Voucher A			Voucher B
	○	○	**Valid today: $2 Voucher**
	○	○	**Valid today: $4 Voucher**
	○	○	**Valid today: $6 Voucher**
	○	○	**Valid today: $8 Voucher**
	○	○	**Valid today: $10 Voucher**
	○	○	**Valid today: $12 Voucher**
	○	○	**Valid today: $14 Voucher**
	○	○	**Valid today: $16 Voucher**
	○	○	**Valid today: $18 Voucher**
Valid in 30 days: $40 Voucher	○	○	**Valid today: $20 Voucher**
	○	○	**Valid today: $22 Voucher**
	○	○	**Valid today: $24 Voucher**
	○	○	**Valid today: $26 Voucher**
	○	○	**Valid today: $28 Voucher**
	○	○	**Valid today: $30 Voucher**
	○	○	**Valid today: $32 Voucher**
	○	○	**Valid today: $34 Voucher**
	○	○	**Valid today: $36 Voucher**
	○	○	**Valid today: $38 Voucher**
	○	○	**Valid today: $40 Voucher**

Part 1 of this study: Instructions (4/4)

Decision screen 2

When you fill out a choice list, you may feel **uncertain about whether you prefer the left or right voucher**. On decision screen 2, we will ask you to select a button to indicate **how certain** you are about how much the larger voucher amount with the later validity period is worth to you in terms of voucher credit that can be spent in the earlier validity period.

Example

Suppose that on the first decision screen you indicated that you value a $40 voucher that is valid in 30 days somewhere between a $12 and a $14 voucher that is valid today. Your second decision screen would look like this.

How certain are you that you actually value a $40 voucher that is valid **in 30 days** somewhere between a $12 and a $14 voucher that is valid **today**?

○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%

very uncertain **completely certain**

Comprehension questions

The questions below test your understanding of the instructions.

Important: If you fail to answer any one of these questions correctly, you will not be allowed to participate in the study, and you will not receive the completion payment.

1. Which of the following statements about the voucher below is true?

   Valid in 1 month: $30 Voucher

   ○ This voucher can be used to order food starting from today until no later than 1 month.
   ○ This voucher can be used to order food any time after 1 month. The validity period has no end date.
   ○ This voucher can be used to order food in the 7-day period starting in 1 month.

2. Suppose you are 80% certain that your decisions actually correspond to how much the different voucher options are worth to you.

   Which button should you click in this case?

○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%

   very uncertain **completely certain**

3. Which of the following statements is true?

   ○ Even if the validity period starts in the future, my voucher will be credited to my UberEats account shortly after the experiment. I do not have to remember the validity period because UberEats will send me reminders.
   ○ If the validity period of the voucher starts in the future, I should expect to get my voucher credited to my UberEats account only shortly before the validity period starts. I have to memorize the validity period, otherwise I may forget to use the voucher amount. There is also some risk that I will not actually receive the voucher.