On the dynamics of the responses in Frydman and Jin (2022): Nullius in verba

Hertel, Johanna and Igan, Deniz and Smith, John

30 June 2023

On the dynamics of the responses in Frydman and Jin (2022): Nullius in verba*

Johanna Hertel† Deniz Igan‡ John Smith§

June 30, 2023

Abstract

Frydman and Jin (2022) ["Efficient coding and risky choice," *Quarterly Journal of Economics*, 137, 161—213] present a model of *efficient coding* whereby decision makers are Bayesian learners of a stochastic distribution. The model predicts that decision makers will devote more cognitive resources to—and therefore be more sensitive to–values that appear more frequently. The authors conduct two experiments where subjects make binary choices between a certain amount and a lottery, where the trial-specific values are drawn from a stochastic distribution. While unknown to the subjects, the distribution can be learned over the course of the experiment. The authors conclude that the observations are consistent with efficient coding. However, we note that the authors do not examine observations across trials. When we examine the data from Experiment 1, we do not find evidence that the relationship between sensitivity and frequency increased across trials. When we include specifications that account for the parameters in the previous trial, the treatment interaction estimates are no longer significant. The effects identified by Frydman and Jin (2022) in Experiment 1 are simply a recency bias and not the result of Bayesian learning. We find that subjects in Experiment 2 are less—not more—sensitive to values they encounter more frequently. In summary, we do not find support for the central claims made by the authors. Finally, we describe some unreported details in the preregistration reports of Frydman and Jin (2022). We encourage economists to exercise more skepticism until convinced by the authors' arguments.

Keywords: data reanalysis, Bayesian learning, Bob Critique
JEL: C40, G02, G41

*We thank Carlos Alos-Ferrer, Roberto Barbera, Sean Duffy, Antonio Filippin, and Adam Sanjurjo for helpful comments. John Smith thanks Biblioteca de Catalunya. John Smith gratefully acknowledges financial support from the Rutgers University-Camden College of Arts and Sciences. The views expressed here are those of the authors and do not necessarily represent the views of the Bank for International Settlements or those who provided helpful comments.

†Avadel Pharmaceuticals

‡Bank for International Settlements and CEPR

§Corresponding Author; Rutgers University-Camden, Department of Economics, 311 North 5th Street, Camden, New Jersey, USA 08102; Email: smithj@camden.rutgers.edu; Phone: +1-856-225-6319; Fax: +1-856-225-6602

I. Introduction

Nullius in verba is Latin and can be translated as "take nobody's word for it."[1] In other words, regardless of the authority of a person making a scientific claim, one should only consider whether their evidence and their arguments are compelling.

Frydman and Jin (2022), hereafter FJ, develop and test a model of *efficient coding*. On page 162, FJ write, "...there is evidence that noise in perception of sensory stimuli—such as light or sound—changes optimally with the environment. Specifically, a core principle from neuroscience called efficient coding states that the brain should allocate resources so that perception is relatively more precise for those stimuli that are expected to occur relatively more frequently."

FJ also describe how their efficient coding decision maker comes to learn which stimuli occur more frequently. On page 173, FJ write, "Given the prior and likelihood functions defined above, the DM proceeds by using Bayesian inference to compute a posterior distribution of each payoff in the choice set." Clearly, at the heart of efficient coding, is Bayesian learning through experience.

FJ place subjects into stochastic environments and claim that subjects learn enough about the stochastic distributions so that behavior is affected by the likelihood of outcomes. Because subjects can only learn the distributions through experience, any learning must take place over the course of the experiment. However, the authors do not analyze choices across trials. Rather, FJ only analyze choices that have been averaged across trials and averaged across other important trial-specific details.

To see the problem with making claims of learning without analyzing choices across trials, consider Bob, who is engaged in a repeated guessing task. In the beginning of every trial, the experimenter makes an identical and independent draw from a symmetric distribution on a set of real numbers. Bob is asked to provide his best guess of that trial-specific draw. After Bob's guess, the trial-specific draw is revealed to Bob. While Bob is not explicitly given the

[1] This is the motto of the Royal Society (UK). From their website: The Royal Society's motto 'Nullius in verba' is taken to mean 'take nobody's word for it'. It is an expression of the determination of Fellows to withstand the domination of authority and to verify all statements by an appeal to facts determined by experiment (https://royalsociety.org/about-us/history/).

distribution, it can be learned from experience.

However, suppose that Bob simply guesses the revealed draw from the previous trial. Despite that the mean of Bob's guesses will converge to the mean of the distribution, a researcher should not claim that Bob has learned the distribution. We note that Bob's guesses do not become more accurate across trials. We also note that Bob's trial-specific guesses do not converge to the mean. In other words, measures of Bob's performance do not improve across trials.

Relatedly, when researchers place subjects into a stochastic environment, any claims that subjects have learned aspects of the environment should be supported by analyzing behavior across trials. Moreover, any standard by which learning is declared, should not conclude that Bob has learned the distribution, otherwise it is vulnerable to the *Bob Critique*.[2] Therefore, the reader should be skeptical of the FJ claims of learning, even before analyzing the data: the results of FJ are clearly vulnerable to the Bob Critique.

In support of the FJ claim that efficient coding is a "core principle" of neuroscience, the authors cite Barlow (1961), Laughlin (1981), Girshick, Landy, and Simoncelli (2011), Wei and Stocker (2015, 2017), Polanía, Woodford, and Ruff (2019), Heng, Woodford, and Polanía (2020), and Payzan-LeNestour and Woodford (2022).

Barlow (1961) does not analyze data and Laughlin (1981) analyzes the effect of light on the compound eyes of a blowfly. While the other more recent references analyze human behavior in stochastic environments, we note that none of them support their claims of learning by analyzing measures of performance across trials. Therefore, each of these references are also vulnerable to the Bob Critique.[3] It is surprising that current efforts would be vulnerable to the Bob Critique since the classic psychology literature devoted considerable attention to characterizing learning across trials.[4]

Further, Girshick, Landy, and Simoncelli (2011), Wei and Stocker (2015), and Polanía, Woodford, and Ruff (2019) cite Körding and Wolpert (2004) as evidence that subjects use

[2] A version of the Bob Critique first appeared in Duffy, Hertel, Igan, Pinheiro, and Smith (2022).

[3] Heng, Woodford, and Polanía (2020) estimate model parameter estimates across trials (Figure 5c,d). Further, "Figure 5-figure supplement 1" characterizes performance across trials and there does not appear to be learning.

[4] See Bush and Mosteller (1955) for a notable reference.

Bayes' rule to improve judgments in stochastic environments. Körding and Wolpert (2004) is a prominent contribution in the Bayesian judgments literature. However, it has recently emerged that the data in Körding and Wolpert (2004) do not support claims of learning. Körding and Wolpert (2004) also do not analyze data across trials and their claims of Bayesian learning seem to be a statistical artifact of analyzing averaged data. Duffy, Hertel, Igan, Pinheiro, and Smith (2022) find that the observations in Körding and Wolpert (2004) are explained by a recency bias–like Bob–and subjects do not appear to be learning.[5] However, Körding and Wolpert (2004) remains uncorrected in the pages of *Nature* and continues to receive citations.[6]

II. Overview

Below, we examine the datasets from Experiments 1 and 2 in FJ. In both cases, we proceed by describing their experimental design, making preliminary observations, reproducing the analysis reported by FJ, and conducting new analyses.

In both Experiments 1 and 2, subjects make binary choices between a certain payment amount and a simple lottery.[7] The trial-specific values are drawn from a stochastic distribution. The distribution is unknown to the subjects but it can be learned over the course of the experiment. Also in both experiments there are two within-subject stochastic distribution treatments.

In Experiment 1, the values are drawn from two uniform distributions, where one distribution has a larger variance than the other. From their analysis of the data, FJ report an increased sensitivity for parameter values that the subjects encounter more frequently. Specifically, for parameter values where the distributions overlap, subjects appear more sensitive to the values in treatment with low variance than the treatment with high variance. FJ claim that the subjects learned the distribution and this causes this difference in behavior. However,

[5] Duffy et al. (2022) also find that Körding and Wolpert (2004) mischaracterize key aspects of their experimental design.

[6] There is a small literature that examines claims of learning that do not analyze data across trials. For example, Duffy and Smith (2020a) reanalyze the replicated data from Huttenlocher, Hedges, and Vevea (2000). Further, Duffy and Smith (2018) reanalyze the data from Duffy, Huttenlocher, Hedges, Crawford (2010). Unfortunately, in the former case, the content also remains uncorrected in the pages of *Journal of Experimental Psychology: General*.

[7] Experiment 1 also contains a separate set of perceptual tasks. In this paper, we do not analyze these observations.

FJ does not support their claim of learning with an analysis across trials.

In our analysis of the data from Experiment 1, we do not find evidence that this effect increased across trials (Tables II, III, $A3$, and $A4$). It therefore seems as if the subjects did not learn the distribution. When we include specifications that account for the parameters in the previous trial, the treatment interaction estimates are no longer significant (Tables IV, $A5$, $A6$, $A7$, $A8$, and $A9$). We conclude that the effects identified by FJ in Experiment 1 are simply a recency bias and not the result of Bayesian learning.

In Experiment 2, the values are drawn from two piece-wise uniform distributions, where higher values are more likely in one distribution and lower values are more likely in the other. From their analysis of the data, FJ report a difference in the behavior of subjects based on the stochastic environment. However, although it seems to be an implication of their analysis in Experiment 1, FJ do not analyze whether subjects are more sensitive to parameter values that are encountered more frequently. We conduct such an analysis. We find the opposite: subjects are less sensitive to parameters that they encounter more frequently (Tables VI and $A11$). Further, consistent with the FJ analysis of Experiment 1, we restrict attention to trials before a regime shift to a new stochastic distribution treatment. Here we do not find differences in behavior reported by FJ (Tables VII, $VIII$, $A12$, and $A13$).

To summarize, in our analysis of the data from Experiments 1 and 2, we do not find support for the central claims made by FJ. Below we discuss these matters in more detail.

III. Experiment 1

III.A. Design

Subjects made 600 pairwise choices between a certain amount (C), and a lottery: amount X with probability 0.5 and 0 with probability 0.5. Even though the probability of 0.5 is fixed throughout the experiment, FJ often refers to this probability as p. FJ refer to the choice of the lottery as the *risky* choice. In the *high volatility* treatment, X was drawn from a uniform distribution on $\{8.00, 8.01, ..., 31.99, 32.00\}$ and C was drawn from a uniform distribution on $\{4.00, 4.01, ..., 15.99, 16.00\}$. In the *low volatility* treatment, X was drawn from a uniform

distribution on $\{16.00, 16.01, ..., 23.99, 24.00\}$ and C was drawn from a uniform distribution on $\{8.00, 8.01, ..., 11.99, 12.00\}$.[8]

Subjects were presented with 300 consecutive trials in the high volatility treatment and 300 consecutive trials in the low volatility treatment. A total of 150 subjects participated in the experiment: 75 were first given the high volatility treatment and 75 were first given the low volatility treatment. Since there were 150 subjects each making 600 binary choices, the dataset contains $90,000$ observations.

Within both blocks of 300 trials, the first 30 are referred to as *adapt* trials. The remaining 270 trials within the block of 300 are referred to as *test* trials. Within the test trials, 30 specific and predetermined pairs of X and C were presented to every subject in both treatments.[9] FJ refer to the trials containing these 30 pairs as *common trials*. There are $4,500$ common trial observations in both treatments for a total of $9,000$ common trial observations.

One of the 600 trials was randomly selected and the subject was paid based on the decision in that trial. The average amount earned was \$10.14. The reader is directed to FJ for further details on the design.

III.B. Preliminary observations

FJ assert that subjects learn the stochastic distributions and this differentially affects choice. FJ assume that subjects learn the distribution in the adapt trials and this affects choice in the test trials. On page 202, FJ write, "...we have assumed that subjects in our experiments are fully adapted to the population distribution after completing an initial set of preregistered 'adaptation trials.'"[10] However, we note that these adapt trials are neither analyzed nor compared to the test trials. We address this matter in our analysis below.

[8] In the appendix, we discuss some unexpected differences between the treatments.

[9] See the appendix or the FJ supplemental material for a list of these 30 pairs.

[10] The full quote is: Most empirical tests of efficient coding in sensory perception assume full adaptation to the prior distribution (Laughlin 1981; Wei and Stocker 2015), and this assumption has also been recently invoked in papers on efficient coding in value-based decisions (Rustichini et al. 2017; Polanía, Woodford, and Ruff 2019). Following this literature, we have assumed that subjects in our experiments are fully adapted to the population distribution after completing an initial set of preregistered "adaptation trials."

III.C. Partial reproduction of Table I from FJ

Here we reproduce the first three columns from Table I in FJ. These regressions have a dependent variable, which has a value of 1 if the risky option was selected in that trial and 0 otherwise. As an independent variable, FJ include the *high* dummy variable, which has a value of 1 in the high volatility treatment, and 0 otherwise. Independent variables also include, X, C, the interaction of X and high, and the interaction of C and high.

FJ run linear regressions for Table I.[11] Correspondingly, we run linear regressions. However, we include the analogous logistic specifications in the appendix.

We attempt to replicate the technique employed by FJ to account for the panel data nature of the observations. Our attempts are somewhat hampered by the fact that FJ did not report fit statistics. In the end, we estimate a variance component model where the intercept, X, and C are random.

Recall that the dataset contains $90,000$ choices between the certain option and the risky option. FJ focus attention on the $9,000$ common trials. In specifications (1) and (2), FJ further restrict the sample to the first half of the experiment. In specification (1), consistent with their preregistration, FJ exclude the observations from subjects who selected the same option within these trials. A single subject satisfied this condition in the first half of the trials.[12] In specification (2), FJ impose a further restriction, which was not preregistered: they further exclude trials where the decision was faster than 0.5 seconds. We note that this exclusion only affects fast decisions, and not the slow decisions, despite that response time has a maximum of 104 seconds. We also note that this restriction excludes 6.7% of the observations in specification (1).

In specification (3), FJ analyze the data from common trials in both halves of the experiment.[13] FJ Table I also includes specifications that analyze nonadjacent sets of trials from both halves of the experiment. Specifically, FJ presents an analysis that excludes trials

[11] We note that the online appendix of FJ contains Table $D.2$, which is a logistic version of Table I in FJ.

[12] Our regressions have one more observation than those reported by FJ because one subject had two common trial observations within a treatment.

[13] It is not clear to us why our analysis of this specification has the same number of observations as reported by FJ.

$301 - 450$. In other words, the analysis contains observations only from trials $1 - 300$ and $451 - 600$. We admit that the change in the stochastic distribution poses challenges to the analysis. Here we focus on the more straightforward analysis of the first 300 trials of the experiment, rather than the ad-hoc analysis of non-adjacent trials.

Motivated by the nature of the response time exclusions in specification (2), in specification (4), we only exclude decisions faster than 0.25 seconds. This restriction excludes only 0.87% of the observations analyzed in specification (1). In specification (5), we include every first half trial where both X and C fall within the domain of the low volatility treatment. In other words, we only include trials where both $16.00 \leq X \leq 24.00$ and $8.00 \leq C \leq 12.00$, which allows us to use more of the $90,000$ observations in the dataset. We refer to these observations as *common domain trials*. In specification (6), we exclude first half common domain trials with a response time less than 0.25 seconds. This restriction excludes 0.86% of trials analyzed in specification (5).

These linear random-effects regressions are summarized in Table I.

Table I: Linear random-effects regressions of risky choice

	(1)	(2)	(3)	(4)	(5)	(6)
high	0.026	0.005	-0.025	0.013	0.159	0.134
	(0.190)	(0.199)	(0.098)	(0.189)	(0.149)	(0.146)
X	0.066^{***}	0.074^{***}	0.062^{***}	0.067^{***}	0.072^{***}	0.072^{***}
	(0.006)	(0.006)	(0.004)	(0.006)	(0.005)	(0.005)
C	-0.167^{***}	-0.186^{***}	-0.164^{***}	-0.168^{***}	-0.165^{***}	-0.166^{***}
	(0.013)	(0.012)	(0.009)	(0.013)	(0.011)	(0.011)
X*high	-0.017^{*}	-0.023^{**}	$-0.006^{\dagger}$	-0.017^{*}	-0.023^{**}	-0.022^{**}
	(0.008)	(0.008)	(0.003)	(0.008)	(0.007)	(0.007)
C*high	$0.033^{\dagger}$	0.049^{**}	$0.014^{\dagger}$	0.034^{*}	0.030^{*}	0.031^{*}
	(0.017)	(0.017)	(0.008)	(0.017)	(0.015)	(0.015)
Interc.	0.772^{***}	0.783^{***}	0.826^{***}	0.758^{***}	0.640^{***}	0.640^{***}
	(0.166)	(0.179)	(0.106)	(0.167)	(0.120)	(0.120)
AIC	3,928.0	3,552.3	6,624.4	3,832.2	18,421.2	17,941.2
Obs.	4,471	4,171	8,257	4,432	24,679	24,465

We provide the coefficient estimates and the standard errors in parentheses. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

We note the close agreement between our results and the first three columns of Table I

in FJ. We also note that FJ argue that the significant interaction terms present evidence of learning. Further, we note that the interaction terms in specification (4) are both significant at 0.05. We also note that, in specifications (5) and (6), the interaction terms involving X and C are significant (respectively) at 0.01 and 0.05. The results are largely unchanged when we run logistic regressions, rather than the linear probability specifications presented by FJ.[14] We also find analogous effects when we employ $0.5 * X - C$ as an independent variable, rather than X and C.[15]

III.D. The search for evidence of learning

Here we investigate whether there is evidence of learning in the FJ dataset. In the body of the paper, FJ do not appear to offer any evidence of learning by examining performance across trials. In the FJ appendix, Table $D.4$ characterizes an analysis of test trials that fall among the first 300 trials. The analysis includes a dummy variable indicating whether the trial is between 166 and 300, or not. Their analysis does not find differences in sensitivity between these two sets of trials.

It can seem surprising, if there is actually learning across trials, that there would not be evidence of learning between trials 31 and 300. Below, we conduct a more detailed analysis of common trials across observations. We offer specifications where we include trials between 31 and 300. However, it is possible that learning–if it exists–converged before the end of the first half of trials. We therefore also conduct analyses of common trials that fall between 31 and 210, between 31 and 120, and between 31 and 90. In each of these regressions, we restrict attention to observations with a response time greater than 0.25 seconds.

Because there might be differences in learning between the high and low volatility treatments and additionally it is not straightforward to interpret a triple interaction, we restrict attention to a volatility treatment. We analyze the model:

[14] We summarize this analysis in Table $A1$. In this and every logistic analysis, we conduct repeated measures specifications. We report QIC (Pan, 2001) rather than AIC.

[15] We summarize this analysis in Table $A2$.

$$Risky = \alpha + \beta_X * X + \beta_C * C + \beta_f * f(trial) + \gamma_X * X * f(trial) + \gamma_C * C * f(trial) + \varepsilon,$$

for various specifications of $f(trial)$. We analyze a linear specification, a square root specification, and a dummy variable indicating the second half of the set of trials. Similar to Table I, we estimate a variance component model where the intercept, X, and C are random.

The estimates of γ_X and γ_C for the 24 regressions are summarized in Table II. Each estimate of β_X is positive and each estimate of β_C is negative. Therefore evidence of increased sensitivity across trials would be identified by a positive and significant estimate of γ_X and a negative and significant estimate of γ_C.

Table II: Estimates of γ_X and γ_C from 24 linear random-effects regressions on common trials

	Trial		Sqrt. Trial		Second Half		Obs.
	γ_X	γ_C	γ_X	γ_C	γ_X	γ_C	
High	−0.00004	0.00005	−0.0009	0.0011	−0.005	0.0018	2,239
31-300	(0.00004)	(0.0001)	(0.0010)	(0.002)	(0.006)	(0.011)	
Low	−0.00006	0.00005	−0.0014	0.0015	−0.008	0.012	2,193
31-300	(0.00005)	(0.0001)	(0.0012)	(0.0021)	(0.007)	(0.014)	
High	0.0001	0.0001	0.0012	0.0022	0.003	0.014	1,508
31-210	(0.0001)	(0.0001)	(0.0014)	(0.0027)	(0.007)	(0.015)	
Low	−0.0001	−0.00013	−0.0025	−0.0017	−0.019*	−0.027†	1,471
31-210	(0.0001)	(0.00015)	(0.0018)	(0.0031)	(0.008)	(0.015)	
High	0.0002	0.0003	0.0037	0.0046	0.015	0.021	771
31-120	(0.0002)	(0.0005)	(0.0033)	(0.0077)	(0.010)	(0.023)	
Low	0.0002	0.0007	0.0035	0.014†	0.0084	0.0076	734
31-120	(0.0002)	(0.0004)	(0.0041)	(0.007)	(0.013)	(0.022)	
High	0.0003	−0.0005	0.0035	−0.009	0.014	−0.033	512
31-90	(0.0003)	(0.0008)	(0.0052)	(0.012)	(0.014)	(0.027)	
Low	−0.0001	0.0021*	−0.0008	0.034*	−0.015	0.060†	481
31-90	(0.0004)	(0.0009)	(0.0057)	(0.013)	(0.014)	(0.030)	

We provide the coefficient estimates and the standard errors in parentheses. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Among the 48 coefficient estimates, only three are significant at 0.05, however each of these are the opposite sign as predicted by learning across trials. There are two estimates

significant at only 0.1, however one is opposite sign as predicted by learning. In fact, among the 48 estimates, 30 have the opposite sign from that predicted by learning across trials. Our results are similar when we perform logistic regressions rather than the linear probability model employed by FJ.[16] In other words, we do not find evidence of learning when we examine only common trials.

Despite that our analysis of common trials in Table II fails to find evidence of learning, we note that the vast majority of the trials are not common trials. Next, we examine whether we can find evidence of learning in common domain trials.

Recall that common domain trials have X and C that fall in the domain of both volatility treatments. Specifically, we include all trials where both $16.00 \leq X \leq 24.00$ and $8.00 \leq C \leq 12.00$. We can therefore include the first 30 trials (the adapt trials) in this analysis. We can also include even more granular partitions of the data. We offer specifications where we analyze data on trials $1-300$, $1-210$, $1-120$, $1-90$, and $1-60$. We exclude excessively fast decisions in these common domain trials: our dataset is identical to that analyzed in specification (6) in our Table I.

The techniques of analysis are similar to those summarized in Table II, but we also include a specification that indicates whether the trial was after the first 30 trials. In other words, our *After First* 30 dummy variable indicates that the observation was from a test–and not adapt–trial. These specifications allow us to learn how decisions in the adapt trials are different than the test trials.[17]

The estimates of γ_X and γ_C for the 38 regressions are summarized in Table III. Again, each estimate of β_X is positive and each estimate of β_C is negative. Likewise, evidence of increased sensitivity across trials would be identified by a positive and significant γ_X and a negative and significant γ_C.

Table III presents 76 coefficient estimates, where 20 are significant at 0.1 or greater: 7

[16] We summarize this analysis in Table $A3$. Our results are largely not changed.

[17] Note that the Second Half and After First 30 specifications are identical on trials $1-60$. Therefore, we only report the former.

Table *III*: Estimates of γ_X and γ_C from 38 linear random-effects regressions on common domain trials

	Trial		Sqrt. Trial		Second Half		After First 30		Obs.
	γ_X	γ_C	γ_X	γ_C	γ_X	γ_C	γ_X	γ_C	
High	-0.00003	0.00003	-0.0004	0.001	-0.005	0.005	0.005	-0.005	$2,468$
1-300	(0.00003)	(0.00007)	(0.0007)	(0.001)	(0.006)	(0.010)	(0.012)	(0.019)	
Low	-0.00005^{**}	0.00004	-0.0012^{**}	0.0009	-0.009^{**}	0.007	-0.010^{*}	$0.016^{\dagger}$	$21,997$
1-300	(0.00002)	(0.00005)	(0.0004)	(0.0010)	(0.003)	(0.008)	(0.004)	(0.010)	
High	0.00007	0.00005	0.001	0.001	0.006	0.013	0.012	-0.006	$1,737$
1-210	(0.00006)	(0.0001)	(0.001)	(0.002)	(0.007)	(0.012)	(0.013)	(0.020)	
Low	-0.00007^{**}	0.00005	-0.0012^{**}	0.001	-0.007^{*}	0.003	-0.009^{*}	$0.014^{\dagger}$	$15,486$
1-210	(0.00003)	(0.0001)	(0.0005)	(0.001)	(0.003)	(0.008)	(0.004)	(0.009)	
High	0.0002	0.00005	0.002	0.0002	0.014	-0.012	0.011	-0.016	$1,000$
1-120	(0.0002)	(0.0003)	(0.002)	(0.004)	(0.010)	(0.018)	(0.013)	(0.022)	
Low	-0.00008	0.00007	$-0.0013^{\dagger}$	0.0011	-0.004	0.008	$-0.007^{\dagger}$	$0.014^{\dagger}$	$8,873$
1-120	(0.00005)	(0.00013)	(0.0007)	(0.0017)	(0.004)	(0.008)	(0.004)	(0.008)	
High	0.0001	-0.0004	0.002	-0.004	-0.007	-0.027	0.004	-0.026	741
1-90	(0.0002)	(0.0004)	(0.003)	(0.005)	(0.012)	(0.023)	(0.014)	(0.025)	
Low	-0.0001	0.0004^{*}	-0.001	0.004^{*}	-0.002	0.022^{**}	$-0.006^{\dagger}$	0.019^{**}	$6,656$
1-90	(0.0001)	(0.0002)	(0.001)	(0.002)	(0.004)	(0.008)	(0.003)	(0.007)	
High	-0.0004	0.0001	-0.003	0.001	-0.007	-0.012	–	–	476
1-60	(0.0005)	(0.0010)	(0.005)	(0.009)	(0.017)	(0.031)			
Low	$-0.0002^{\dagger}$	0.0002	-0.0023^{*}	0.0015	$-0.007^{\dagger}$	0.010	–	–	$4,440$
1-60	(0.0001)	(0.0003)	(0.0011)	(0.0026)	(0.004)	(0.009)			

We provide the coefficient estimates and the standard errors in parentheses. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

are significant at only 0.01 and each has a sign opposite to that predicted by learning; 6 are significant at only 0.05 and each has a sign opposite to that predicted by learning; and 7 are significant at only 0.1 and each has a sign opposite to that predicted by learning. Moreover, among the 76 coefficient estimates, 55 have a sign opposite to that predicted by learning. Perhaps most surprisingly, low volatility adapt trials appear to be significantly more sensitive to X and C then low volatility test trials. Our results are largely unchanged when we perform logistic regressions rather than the linear probability model employed by FJ.[18]

In Tables II, III, $A3$, and $A4$, we report 248 coefficient estimates from 124 regressions. A total of 44 of these coefficient estimates are significant at 0.05, however each have the opposite sign as predicted by learning across trials. We also find that 19 are significant at only 0.1, however 18 have the opposite sign as predicted by learning. Further, 178 of the 248 estimates have the opposite sign as predicted by learning across trials. In light of our analysis, we fail to find any evidence of learning across trials.

III.E. The evidence for recency effects

As we cannot find any evidence that subjects learned the distribution, here we seek other mechanisms to explain the apparent differential sensitivity to parameter values. Recall that Bob displayed a recency bias: simply guessing the number from the previous trial. When Bob's responses are averaged across trials, a researcher could mistake the behavior for evidence of learning.

Here we ask whether recency effects could produce a candidate for explaining the apparent differential sensitivity to X and C. Perhaps when X and C from the previous trial are different than X and C in the current trial, subjects are less sensitive to the values in the current trial. It is also the case that in high volatility trials, the distance between the current X (C) and the X (C) in the previous trial will be larger than that in low volatility trials. This suggests a natural measure of the difference between the parameters in the previous trial $(t-1)$ and those in the current trial (t):[19]

[18] We summarize this analysis in Table $A4$. Our results are largely not changed.

[19] In the dataset, this appears as PrevX2CDist.

$$\text{Prev}_1 = \sqrt{(X_t - X_{t-1})^2 + 2 * (C_t - C_{t-1})^2}.$$

We investigate whether our Prev_1 variable can explain the differential sensitivity better than the volatility treatment dummy employed by FJ. We perform the analysis on the data analyzed in specifications (1), (4), (5), and (6) of our Table I. Within each of these four specifications, we include a specification (a) without the high volatility dummy, and a specification (b) with the high volatility dummy. These regressions are summarized in Table IV. Note that the observations of specifications (5) and (6) in Table I and Table IV are not identical, because 83 common domain trials do not have a previous trial.

Table IV: Linear random-effects regressions of risky choice

	(1a)	(1b)	(4a)	(4b)	(5a)	(5b)	(6a)	(6b)
high	–	0.02	–	0.01	–	0.33*	–	0.32*
		(0.22)		(0.22)		(0.16)		(0.16)
X	0.07***	0.07***	0.07***	0.07***	0.08***	0.08***	0.08***	0.08***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
C	−0.18***	−0.18***	−0.18***	−0.18***	−0.19***	−0.20***	−0.20***	−0.20***
	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
X*high	–	−0.009	–	−0.009	–	−0.013†	–	−0.013†
		(0.009)		(0.009)		(0.008)		(0.008)
C*high	–	0.0178	–	0.018	–	−0.006	–	−0.005
		(0.0185)		(0.019)		(0.015)		(0.015)
Prev_1	0.005	0.004	0.003	0.002	−0.026†	−0.037*	−0.030*	−0.039**
	(0.021)	(0.023)	(0.021)	(0.023)	(0.014)	(0.015)	(0.014)	(0.015)
X*Prev_1	−0.002**	−0.002*	−0.002**	−0.002*	−0.002***	−0.002***	−0.002***	−0.002***
	(0.0007)	(0.0008)	(0.0008)	(0.0008)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
C*Prev_1	0.005*	0.003†	0.005*	0.003†	0.008***	0.008***	0.008***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Interc.	0.75***	0.75***	0.74***	0.74***	0.83***	0.77***	0.84***	0.79***
	(0.20)	(0.19)	(0.20)	(0.19)	(0.13)	(0.14)	(0.13)	(0.14)
AIC	3,939.0	3,954.0	3,843.0	3,857.9	18,293.2	18,303.8	17,812.8	17,823.8
Obs.	4,471	4,471	4,432	4,432	24,596	24,596	24,382	24,382

We provide the coefficient estimates and the standard errors in parentheses. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Of the 16 estimated interaction terms involving Prev_1, 14 are significant at 0.05. However

none of the 8 estimated interaction terms with high are significant at 0.05. These results clearly demonstrate that the recency effects do a better job explaining the data than the volatility treatment dummy. Our results are largely unchanged when we perform logistic regressions rather than the linear probability model employed by FJ.[20] We also find analogous effects when we employ $0.5 * X - C$ as an independent variable, rather than X and C.[21]

Additionally, we note that our results are not sensitive to the specification of the recency effects. We consider alternate specifications of the recency effects:

$$\text{Prev}_2 = \sqrt{(X_t - X_{t-1})^2 + (C_t - C_{t-1})^2},$$

$$\text{Prev}_3 = |X_t - X_{t-1}| + 2 * |C_t - C_{t-1}|, \text{ and}$$

$$\text{Prev}_4 = |X_t - X_{t-1}| + |C_t - C_{t-1}|.$$

Our results are largely unchanged when we use Prev_2, Prev_3, or Prev_4.[22]

In Tables IV, $A5$, $A6$, $A7$, $A8$, and $A9$, we conduct 24 "(b)" specifications where we account for both recency effects and the volatility treatment. These analyses produce 44 estimates of sensitivity for both recency effects and the volatility treatment. We find that 24 of the estimates involving recency effects are significant at 0.05. By contrast, we find that only 2 of the 44 estimates involving the volatility treatment are significant at 0.05.

Further, consider that in every $(5b)$ and $(6b)$ specification (employing more that $24,000$ observations) the 22 estimates involving recency effects for both X and C are each significant at 0.001. However, for these corresponding 22 estimates for the volatility treatment dummy, only 2 are significant at 0.05. Our analysis strongly suggests that the differential sensitivity to parameter values is driven by recency effects, rather than Bayesian learning.

[20] This analysis is summarized in Table $A5$.

[21] This analysis is summarized in Table $A6$.

[22] These analyses for Prev_2, Prev_3, and Prev_4 are summarized respectively in Tables $A7$, $A8$, and $A9$. These variables appear in the dataset respectively as PrevXCDist, Sum2PrevAbs, and SubPrevAbs.

III.F. Summary of the results

After carefully analyzing the FJ data, we do not find evidence of Bayesian learning. Rather, we find that the effects are driven by a bias toward the parameters in the previous trial.

IV. Experiment 2

IV.A. Design

Similar to Experiment 1, subjects made 600 pairwise choices between a certain amount (C), and a lottery: amount X with probability 0.5 and 0 with probability 0.5. In this experiment, X and C were drawn from piece-wise uniform distributions. In the *increasing* treatment, X was drawn from a distribution with density:

$$f(X) = \begin{cases} \frac{1}{125} \text{ if } 2 \leq X \leq 4.5 \\ \frac{7}{25} \text{ if } 4.5 < X \leq 8 \end{cases}$$

and C was drawn from a distribution with density:

$$f(C) = \begin{cases} \frac{1}{125} \text{ if } 1 \leq X \leq 2.25 \\ \frac{7}{25} \text{ if } 2.25 < X \leq 4 \end{cases}.$$

In the *decreasing* treatment, X was drawn from a distribution with density:

$$f(X) = \begin{cases} \frac{7}{25} \text{ if } 2 \leq X \leq 5.5 \\ \frac{1}{125} \text{ if } 5.5 < X \leq 8 \end{cases}$$

and C was drawn from a distribution with density:

$$f(C) = \begin{cases} \frac{7}{25} \text{ if } 1 \leq X \leq 2.75 \\ \frac{1}{125} \text{ if } 2.75 < X \leq 4 \end{cases}.$$

A total of 200 subjects participated in the experiment, which was administered online. Also similar to Experiment 1, subjects were presented with 300 consecutive trials in the increasing

treatment and 300 consecutive trials in the decreasing treatment. There were 102 subjects who were first given the increasing treatment and 98 were first given the decreasing treatment. Since 200 subjects each made 600 choices, the dataset has 120,000 observations.

Within both blocks of 300 trials, the first 60 are considered adapt trials, and the remaining 240 trials are considered test trials. Within the test trials, 8 specific and predetermined pairs of X and C were presented to every subject in both treatments. Each *common* pair has $C = 2.70$ and $X \in \{7.13, 7.26, 7.37, 7.49, 7.62, 7.76, 7.87, 7.99\}$. A randomly selected common pair was presented to subjects on trial 90 then every 30 trials until trial 300. There were 1,600 common trials in both treatments for a total of 3,200 common trials.

Decisions that were not completed within 10 seconds were not recorded. There were 13 common trials that had a response time that exceeded this limit, so there are 3,187 common trial observations.

With the exception of the common trials, the dataset lists X and C to 5 decimals. We assume that subjects were presented with values rounded to the nearest \$0.01. A randomly selected trial was selected for payment, in addition to a \$6.50 show up fee. The average amount earned in the experiment was \$9.27. The reader is directed to FJ for further details on the design.

IV.B. Preliminary observations

Recall that the online experiment involved 120,000 pairwise choices, however there are 3,187 common trial observations. Based on preregistered exclusion criteria, FJ exclude an additional 909 observations, which account for 28.5% of common trial observations. Therefore, FJ present an analysis of 2,278 observations.

IV.C. Reproduction of Table III from FJ

The analysis in FJ Table *III* is similar to that in FJ Table *I*, with the following exceptions. Because C is constant at 2.70 in the common trials, the authors do not include C as an independent variable. The analysis contains a dummy variable indicating whether the observation

was made in the increasing treatment. Note, we do not repeat the use of the term "prior" as the variable name because this seems to assume the conclusions of the efficient coding model. Rather, we refer to the treatment dummy variable as *increasing*. Further, we note that the analysis summarized in FJ Table III also does not include interaction terms.

Specifications (1) and (2) analyze the full set of $2,287$ observations. Specification (3) only analyzes the first half of observations within each block of 300, and specification (4) only analyzes the second half of each block. These linear random-effects regressions are summarized in Table V.

Table V: Linear random-effects regressions of risky choice

	(1)	(2)	(3)	(4)
increasing	0.075^{***}	0.075^{***}	0.088^{**}	0.070^{**}
	(0.022)	(0.022)	(0.029)	(0.024)
X	–	$0.045^{\dagger}$	–	–
		(0.024)		
Interc.	0.690^{***}	$0.351^{\dagger}$	0.684^{***}	0.695^{***}
	(0.029)	(0.188)	(0.032)	(0.030)
AIC	$1,717.5$	$1,719.7$	827.4	$1,091.9$
Obs.	$2,278$	$2,278$	862	$1,416$

We provide the coefficient estimates and the standard errors in parentheses. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

We note that our results closely match those in FJ Table III. Specifically, we observe that the increasing dummy estimate is positive and significant in every specification.[23]

IV.D. Differences between the FJ analyses of Experiments 1 *and* 2

There appears to be two important differences between the FJ analyses of Experiments 1 and 2. First, the analysis of Experiment 1 examined the relationship between the sensitivity to a value and the frequency with which it was encountered. However, the analogous sensitivity relationship was not examined in the analysis of Experiment 2. Second, the analysis of Experiment 1 focused on the first 300 trials within a treatment because it is difficult to interpret

[23] We include the logistic specification in Table $A10$.

learning after a change in regime across treatments. However, the analysis of Experiment 2 examines data from both treatments in all 600 trials and does not contain an analysis restricted to a treatment in the first 300 trials. In this section, we address these two differences.

According to Figure II, Panel B in FJ, the apparent predictions of efficient coding for the increasing treatment in this setting are *both* a larger propensity to risk *and* a greater sensitivity toward X than the decreasing treatment. Specifically, for values of X between 7.13 and 7.99, the difference between the increasing and decreasing treatments are both positive and increasing in X. On page 199, FJ write, "Figure II, Panel B shows that under efficient coding, the decreasing distribution of X generates perception that is insensitive and biased downward for large values of X."[24]

The relationship between sensitivity and frequency seems to be a fundamental prediction of efficient coding. In the abstract, FJ write, "In our first experiment, we find that risk taking is more sensitive to payoffs that are presented more frequently." However, the analysis summarized in FJ Table III does not permit an analysis of the sensitivity to X and it is not clear to us why this prediction was not tested in the analysis of Experiment 2. Below, we supplement the analysis in our Table V, with an interaction between the increasing treatment dummy and X. In an effort to aid the interpretation of the increasing dummy variable, we normalize X by subtracting the mean (7.5).

In specifications (1) and (2), we conduct our analysis on the $2,278$ common trials that were not excluded by FJ. In specifications (3) and (4), we analyze the $3,187$ common trials in the dataset. We use the random-effects specifications from Table V. This analysis is summarized in Table VI.

[24] On page 179, also FJ write, "...for the decreasing distribution, the DM allocates little coding resources toward large values of X. As a result, the DM is insensitive to high values of X."

Table VI: Linear random-effects regressions of risky choice

	(1)	(2)	(3)	(4)
increasing	0.075^{***}	0.080^{***}	0.054^{**}	0.059^{***}
	(0.022)	(0.023)	(0.017)	(0.017)
X-7.5	$0.045^{\dagger}$	0.091^{**}	0.039^{*}	0.080^{**}
	(0.024)	(0.034)	(0.018)	(0.027)
increasing*(X-7.5)	–	-0.094^{*}	–	-0.082^{*}
		(0.042)		(0.033)
Interc.	0.687^{***}	0.684^{***}	0.627^{***}	0.625^{***}
	(0.029)	(0.029)	(0.028)	(0.028)
AIC	1,719.7	1,720.0	2,065.9	2,065.9
Obs.	2,278	2,278	3,187	3,187

We provide the coefficient estimates and the standard errors in parentheses. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

We find that on both datasets of common trials, the increasing treatment is less–not more–sensitive to X than the decreasing treatment. This is the opposite of the prediction of efficient coding in this setting.[25]

As stated before Table V, there can be confounding aspects of interpreting claims of learning across a regime change. It was for this reason that the analysis of Experiment 1 largely focused on the first half of trials. Here we make a similar restriction. We analyze the 1,143 first half common trial observations that were not excluded by FJ and the 1,594 first half common trial observations in the dataset. In the dataset with exclusions, we find that the fraction of risky choice in the first half common trials in the increasing treatment ($mean = 0.769$, $SE = 0.0179$) is greater than the fraction of common trials in the decreasing treatment ($mean = 0.750$, $SE = 0.0178$), although the difference is not significant ($t(1139.6) = -0.73$, $p = 0.46$).[26] In the dataset without exclusions, we find that the fraction of risky choice in common trials in the increasing treatment ($mean = 0.647$, $SE = 0.0168$) is less than–not more than–the fraction of common trials in the decreasing treatment ($mean = 0.698$, $SE = 0.0164$) and the difference is significant ($t(1592) = 2.20$, $p = 0.028$).[27] However, below we conduct an

[25] We include the logistic specification in Table $A11$. Our results are not changed.
[26] The analogous Wilcoxon test also has a p-value of 0.46.
[27] The analogous Wilcoxon test also has a p-value of 0.028.

analysis similar to that in Table VI, but restricted to first half trials. This is summarized in Table VII.

Table VII: Linear random-effects regressions of risky choice

	(1)	(2)	(3)	(4)
increasing	0.011	0.018	−0.051	−0.044
	(0.054)	(0.054)	(0.057)	(0.055)
X-7.5	0.065*	0.123*	0.052*	0.106*
	(0.033)	(0.052)	(0.026)	(0.041)
increasing*(X-7.5)	–	−0.119†	–	−0.107*
		(0.065)		(0.051)
Interc.	0.738***	0.734***	0.695***	0.692***
	(0.036)	(0.037)	(0.037)	(0.037)
AIC	830.8	830.9	1,042.2	1,041.8
Obs.	1,143	1,143	1,594	1,594

We provide the coefficient estimates and the standard errors in parentheses. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Although the evidence is weaker than in Table VI, here we also find that the increasing treatment is less sensitive to X than the decreasing treatment. In addition, in these first half common trials, we do not find a positive and significant estimate of the increasing dummy. In other words, in the first half trials, we do not find evidence of either of the two predictions of efficient coding.[28]

We admit that the number of first half common trials is not large. In an effort to further test the robustness of our results, we consider every first half trial such that $X > 5.5$ and $2.25 < C < 2.75$. On this set, there are no differences in the likelihood of C, yet these values of X are more likely in the increasing treatment than the decreasing treatment. We note that these are the attributes of the designated common trials. We conduct an analysis on the 5,235 first half trials in this region that were not excluded by FJ and on the 7,542 first half trials in the dataset. This analysis is summarized in Table $VIII$.

[28] We include the logistic specification in Table $A12$. Again, our results are largely not changed.

Table $VIII$: Linear random-effects regressions of risky choice

	(1)	(2)	(3)	(4)	(5)	(6)
increasing	-0.126	-0.009	-0.036	$-0.135^{\dagger}$	-0.063	-0.068
	(0.083)	(0.052)	(0.056)	(0.074)	(0.053)	(0.057)
X-7.5	0.128^{***}	0.074^{*}	0.089^{**}	0.094^{***}	0.064^{**}	0.067^{*}
	(0.017)	(0.031)	(0.033)	(0.014)	(0.024)	(0.028)
C-2.5	-0.068^{*}	$-0.063^{\dagger}$	-0.215	-0.030	-0.027	-0.056
	(0.033)	(0.033)	(0.136)	(0.024)	(0.024)	(0.127)
increasing*(X-7.5)	–	$0.065^{\dagger}$	0.050	–	0.037	0.034
		(0.037)	(0.039)		(0.029)	(0.032)
increasing*(C-2.5)	–	–	0.158	–	–	0.030
			(0.140)			(0.129)
Interc.	0.812^{***}	0.754^{***}	0.781^{***}	0.740^{***}	0.703^{***}	0.708^{***}
	(0.050)	(0.035)	(0.040)	(0.048)	(0.036)	(0.041)
AIC	3,890.6	3,891.2	3,892.1	4,510.2	4,513.6	4,515.8
Obs.	5,235	5,235	5,235	7,542	7,542	7,542

We provide the coefficient estimates and the standard errors in parentheses. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

We do not find any positive and significant estimates of the increasing dummy coefficient. Further, we find that 1 of 4 estimates of the sensitivity of X are significant, but it is only significant at 0.1.[29] In conclusion, we do not find evidence in support of the predictions of efficient coding.

IV.E. Summary of the results

Despite being a key prediction of efficient coding, we do not find that subjects are more sensitive to parameters they are more likely to encounter. In fact, in some specifications, we find that subjects are *less* sensitive to parameters they are more likely to encounter. When we restrict attention to trials before the stochastic distribution regime shift, we also do not find differences in choices between the treatments.

V. On the preregistration of FJ

[29] We include the logistic specification in Table $A13$. Our results are similar, but we do not find significance, even at 0.1.

The appendix of the final draft of FJ contains two preregistration reports. This fact is mentioned 12 times throughout FJ–including the abstract–in attempt to convey the integrity and transparency of the paper.

The appended report for Experiment 1 (both risky and perceptual choice) was created on February 9, 2020 and shared on July 17, 2020. The appended report for Experiment 2 was created on February 21, 2021 and was shared on March 1, 2021. Both of these reports respond to the item, "1. Have any data been collected for this study yet?" with, "No, no data have been collected for this study yet." The ReadMe file, posted on August 12, 2020, indicates that the data for Experiment 1 was collected in February 2020 and that the data for Experiment 2 was collected in February 2021.

The publicly circulated drafts of FJ—dated 2018 and 2019—report on the data from two experiments, neither of which were preregistered. We also note that these two experiments are similar to—but nonetheless different than—the two components which form Experiment 1 in the version published in 2022. Neither the Experiment 1 preregistration report nor the body of the paper notes that data from designs similar to Experiment 1 had previously been analyzed and apparently discarded.

Moreover, the publicly circulated draft of FJ—dated August 7, 2020—contains a preregistration report of Experiment 2, which was created on February 27, 2020 and shared on July 17, 2020. We note that the design in this preregistration report is similar to—but again different than—the design in the preregistration report appended in the version published in 2022. Neither the reported Experiment 2 preregistration report nor the body of the paper notes that data from a design similar to Experiment 2 has previously been analyzed and apparently discarded.[30]

In other words, neither these appended preregistration reports nor the body of FJ mention previous versions of the paper that analyzed data from different designs. If transparency is the goal, then it is our view that the final preregistration reports and the final version of the paper should have disclosed these previous designs and that data were collected, analyzed,

[30] We note that the unreported preregistration report of Experiment 2 and the reported preregistration report of Experiment 1, were both shared on July 17, 2020.

and apparently discarded.

VI. Conclusions

We have analyzed the datasets from Experiments 1 and 2 in FJ. In both cases, we do not find support for their main claims. Efficient coding predicts that subjects use Bayesian learning over the course of the experiment to better allocate cognitive resources. This implies that subjects will be more sensitive to parameter values that appear more frequently.

In the data from Experiment 1, we do not find evidence that the predictions of efficient coding are increasing across trials. It therefore seems as if the subjects did not learn the stochastic distribution. When we include specifications that account for the parameters in the previous trial, the treatment interaction estimates are no longer significant. The effects reported by FJ in Experiment 1 are apparently just a recency bias and not the result of Bayesian learning.

In the data from Experiment 2, we do not find evidence that subjects are more sensitive to parameter values that are encountered more frequently. Also, consistent with the FJ analysis of Experiment 1, we restrict attention to trials before a regime shift to a new stochastic distribution treatment. Using these observations, we do not find treatment differences in behavior reported by FJ.

To summarize, in our analyses of the data from Experiments 1 and 2, we do not find support for the central claims made by FJ.

A key reason that claims of learning without an analysis across trials was not deemed suspicious, was that it appears in prestigious journals, such as *Nature*. Our efforts testify to the corrosive effects of incorrect content in the scientific literature. Körding and Wolpert (2004)–published in *Nature*–is a prominent contribution in the Bayesian judgments literature. Duffy, Hertel, Igan, Pinheiro, and Smith (2022) reported that their claims of Bayesian learning were a statistical artifact of analyzing averaged data. We note that an earlier version of Duffy et al. (2022) was submitted to *Nature*. We regret that *Nature*–despite being confronted with incorrect content in its pages–elected to leave the content unaddressed. It is not clear how science can proceed under these circumstances. Further, discussions on the improvements

of science—appearing in *Nature* partner journals—can appear insincere when the journal is apparently indifferent to incorrect content in its pages.[31]

In addition to these problems it will likely surprise the economist reader that the vast majority—and perhaps all—of the Bayesian judgments literature seems unaware of the implications of Bayesian learning found in Savage (1954) and Blackwell and Dubins (1962): Bayesians have beliefs that converge to the truth. Specifically, Savage (1954) shows that, as long as the truth is considered possible, when a Bayesian observes draws from a stochastic distribution posterior beliefs will converge to the true distribution. Blackwell and Dubins (1962) show that two Bayesians with different priors will eventually agree about the distribution after observing enough information. We cannot locate a single example in the Bayesian judgments literature that cites these works. This omission can be difficult to understand because these insights have been in the brain science literature for 60 years (Edwards, Lindman, & Savage, 1963).[32]

Nullius in verba: our advice is for the reader to maintain a skeptical posture until sufficiently convinced by the arguments of the authors.

References

Aczel, Balazs, et al., "A consensus-based transparency checklist." *Nature Human Behaviour*, 4 (2020), 4–6.

Barlow, Horace, "Possible Principles Underlying the Transformations of Sensory Messages," in *Sensory Communication*, Walter A. Rosenblith, ed. (Cambridge,MA: MIT Press, 1961), 217–234.

Blackwell, David, and Lester Dubins, "Merging of opinions with increasing information," *Annals of Mathematical Statistics*, 33 (1962), 882–886.

[31] For example, see Aczel et al. (2020), Camerer et al. (2018), Munafò and Smith (2018), Munafò et al. (2017), and Rahnev et al. (2020).

[32] At points in the paper, FJ can appear unaware of these insights. For example, on page 200, FJ write, "...because the DM's learning problem is more complex than in standard settings, where Bayesian inference would typically generate convergence."

Bush, Robert R., and Frederick Mosteller, *Stochastic Models for Learning*, (1955), Wiley, New York.

Camerer, Colin F., et al., "Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015," *Nature Human Behaviour*, 2 (2018), 637–644.

Duffy, Sean, Janellen Huttenlocher, Larry V. Hedges, and L. Elizabeth Crawford, "Category effects on stimulus estimation: Shifting and skewed frequency distributions," *Psychonomic Bulletin & Review*, 17 (2010), 224–230.

Duffy, Sean, Johanna Hertel, Deniz Igan, Marcelo Pinheiro, and John Smith, "On Bayesian integration in sensorimotor learning: Another look at Kording and Wolpert (2004)," *Cortex*, 153 (2022), 87-96.

Duffy, Sean, and John Smith, "Category effects on stimulus estimation: Shifting and skewed frequency distributions—A reexamination," *Psychonomic Bulletin & Review*, 25 (2018), 1740–1750.

Duffy, Sean, and John Smith, "On the Category Adjustment Model: Another look at Huttenlocher, Hedges, and Vevea (2000)," *Mind and Society*, 19 (2020), 163–193.

Edwards, Ward, Harold Lindman, and Leonard J. Savage, "Bayesian statistical inference for psychological research," *Psychological Review*, 70 (1963), 193–242.

Frydman, Cary, and Lawrence J. Jin, "Efficient coding and risky choice," *Quarterly Journal of Economics*, 137 (2022),161–213.

———,"Efficient coding and risky choice," Working paper, dated August 7, 2020, https://tinbergen.nl/media/event/papers/frydman_jin_820.pdf

———,"Efficient coding and risky choice," Working paper, dated July 8, 2019, https://mendoza.nd.edu/wp-content/uploads/2020/01/Frydman-Paper.pdf

———,"Efficient coding and risky choice," Working paper, dated June 6, 2019,

http://www.its.caltech.edu/~jin/Frydman_Jin_06062019.pdf

———,"Efficient coding and risky choice," Working paper, dated March 2019, https://gattonweb.uky.edu/faculty/hankins/conf2019/ecrc.pdf

———,"Efficient coding and risky choice," Working paper, dated October 2018, https://rady.ucsd.edu/docs/seminars/Jin - Frydman - Efficient Coding and Risky Choice.pdf

Girshick, Ahna R., Michael S. Landy, and Eero P. Simoncelli, "Cardinal Rules: Visual Orientation Perception Reflects Knowledge of Environmental Statistics," *Nature Neuroscience*, 14 (2011), 926–932.

Hayes, K. J. (1953). The backward curve: A method for the study of learning. *Psychological Review, 60*, 269–275.

Heng, Joseph, Michael Woodford, and Rafael Polanía, "Efficient Sampling and Noisy Decisions," *eLife*, 9 (2020), e54962.

Huttenlocher, Janellen, Larry V. Hedges, and Jack L. Vevea, "Why do categories affect stimulus judgment?" *Journal of Experimental Psychology: General*, 129 (2000), 220–241.

Körding, Konrad P., and Daniel M. Wolpert, "Bayesian integration in sensorimotor learning," *Nature*, 427 (2004), 244–247.

Laughlin, Simon, "A Simple Coding Procedure Enhances a Neuron's Information Capacity," *Zeitschrift fur Naturforschung*, 36 (1981), 910–912.

Munafò, Marcus R., and George Davey Smith, "Robust research needs many lines of evidence," *Nature*, 553 (2018), 399-401.

Munafò, Marcus R., et al., "A manifesto for reproducible science," *Nature Human Behaviour*, 1 (2017), 0021.

Pan, Wei , "Akaike's information criterion in generalized estimating equations," *Biometrics*, 57 (2001), 120–125.

Payzan-LeNestour, Elise, and Michael Woodford, "Outlier Blindness: A Neurobiological Foundation for Neglect of Financial Risk," *Journal of Financial Economics*, 143 (2022), 1316–1343.

Polanía, Rafael, Michael Woodford, and Christian Ruff, "Efficient Coding of Subjective Value," *Nature Neuroscience*, 22 (2019), 134–142

Rahnev, Dobromir, et al., "The Confidence Database," *Nature Human Behaviour*, 4 (2020), 317–325.

Rahnev, Dobromir, and Rachel N. Denison, "Suboptimality in perceptual decision making," *Behavioral and Brain Sciences*, 41 (2018), E223.

Rustichini, Aldo, Katherine E. Conen, Xinying Cai, and Camillo Padoa-Schioppa, "Optimal Coding and Neuronal Adaptation in Economic Decisions," *Nature Communications*, 8 (2017), 1208.

Savage, Leonard J., *The Foundations of Statistics*, (1954), Wiley, New York.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116*, 250–264.

Wei, Xue-Xin, and Alan A. Stocker, "A Bayesian Observer Model Constrained by Efficient Coding Can Explain 'Anti-Bayesian' Percepts," *Nature Neuroscience*,18 (2015), 1509–1517.

———, "Lawful Relation between Perceptual Bias and Discriminability," *Proceedings of the National Academy of Sciences*, 114 (2017), 10244–10249.

Supplemental appendix for online publication only

Experiment 1 common trial values of (X, C)

$(17.30, 11.20)$	$(18.64, 11.16)$	$(20.04, 11.22)$	$(21.34, 11.23)$	$(22.69, 11.21)$	$(23.98, 10.37)$
$(17.38, 10.38)$	$(18.72, 10.44)$	$(20.03, 10.40)$	$(21.30, 10.39)$	$(22.71, 10.42)$	$(23.97, 11.18)$
$(17.31, 9.62)$	$(18.68, 9.61)$	$(19.96, 9.62)$	$(21.34, 9.58)$	$(22.62, 9.62)$	$(23.96, 9.55)$
$(17.34, 8.75)$	$(18.63, 8.79)$	$(19.99, 8.81)$	$(21.29, 8.82)$	$(22.66, 8.79)$	$(23.99, 8.84)$
$(17.32, 8.04)$	$(18.68, 8.03)$	$(19.98, 8.01)$	$(21.38, 8.00)$	$(22.67, 8.04)$	$(23.98, 8.03)$

Experiment 1 treatment differences

The specification of the stochastic environment warrants comment. On page 180, FJ write, "...we keep the mean of each payoff distribution constant across conditions. The mean of X is fixed at 20, and the mean of C is fixed at 10." Later on the same page, FJ write, "...the distributions of X and C are independent, and pX and C are identically and uniformly distributed." However, FJ do not offer tests of these claims.

We find significant differences in the means of these variables by treatment. When examining trials before the common trials, we find that the mean of X in the high volatility treatment ($mean = 20.3563$, $SE = 0.1036$) and the mean in the low volatility treatment ($mean = 20.1190$, $SE = 0.0339$) are significantly different ($t(5, 450.3) = -2.18$, $p = 0.0295$). On the same set of trials, the mean of C in the high volatility treatment ($mean = 9.8293$, $SE = 0.0519$) and the mean in the low volatility treatment ($mean = 9.9467$, $SE = 0.0172$) are significantly different ($t(5, 475.5) = 2.15$, $p = 0.0317$). Finally, on the same set of trials, the mean of $0.5 * X - C$ in the high volatility treatment ($mean = 0.3489$, $SE = 0.0732$) and the mean in the low volatility treatment ($mean = 0.1128$, $SE = 0.0245$) are significantly different ($t(5, 494.1) = -3.06$, $p = 0.0022$).

We note that these tests were conducted on $9,005$ trials, because there were 5 instances where the common pair was given twice to the same subject within the same volatility treatment. We also note that the Wilcoxon sum rank tests for X, C, and $0.5 * X - C$ (respectively) have p-values of 0.003, less than 0.001, and less than 0.001.

Therefore, we conclude that there appear to be differences between the treatments beyond

simply the volatility.

Robustness

We conduct a robustness check on the analysis summarized in Table I. Because of the discrete dependent variable, we conduct a logistic specification. We estimate an exchangeable log odds ratio, clustered by subject. In other words, we assume a unique relationship between any two observations involving a particular subject. However, we assume that observations involving two different subjects are statistically independent. The regressions are estimated using Generalized Estimating Equations (GEE). Since GEE is not a likelihood-based method, Akaike's Information Criterion is not available. Therefore, we provide the Quasilikelihood information criterion (QIC), Pan (2001). We summarize these regressions in Table $A1$.

Table $A1$: Logistic random-effects regressions of risky choice

	(1)	(2)	(3)	(4)	(5)	(6)
high	0.101	0.022	−0.082	0.045	0.796	0.692
	(0.848)	(0.917)	(0.428)	(0.846)	(0.661)	(0.655)
X	0.325***	0.379***	0.299***	0.331***	0.318***	0.320***
	(0.038)	(0.040)	(0.025)	(0.039)	(0.032)	(0.032)
C	−0.800***	−0.913***	−0.755***	−0.806***	−0.714***	−0.719***
	(0.086)	(0.090)	(0.055)	(0.087)	(0.067)	(0.067)
X*high	−0.104*	−0.149**	−0.045*	−0.105*	−0.093*	−0.090*
	(0.048)	(0.050)	(0.019)	(0.048)	(0.043)	(0.043)
C*high	0.209*	0.315**	0.098*	0.217*	0.112	0.116
	(0.105)	(0.110)	(0.044)	(0.107)	(0.090)	(0.091)
Interc.	1.092	1.074	1.209*	1.032	0.435	0.436
	(0.746)	(0.829)	(0.478)	(0.754)	(0.538)	(0.543)
QIC	5,350.2	4,824.3	9,609.3	5,283.8	27,402.2	27,026.9
Obs.	4,471	4,171	8,257	4,432	24,679	24,465

We provide the coefficient estimates and the standard errors in parentheses. QIC refers to the Quasi-likelihood information criterion (Pan, 2001). *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

With the exception of the estimates of the intercept, our results from Table I are largely unchanged.

Here we conduct another robustness check of the analysis summarized in Table I. Rather than independent variables of X and C, we use the expected value (EV) of the risky minus sure choice $(0.5 * X - C)$. We summarize this in Table $A2$.

Table $A2$: Linear random-effects regressions of risky choice

	(1)	(2)	(3)	(4)	(5)	(6)
high	0.020	0.032	-0.001	0.021	0.011	0.011
	(0.046)	(0.045)	(0.013)	(0.046)	(0.043)	(0.043)
EV	0.149^{***}	0.161^{***}	0.142^{***}	0.150^{***}	0.154^{***}	0.155^{***}
	(0.010)	(0.009)	(0.007)	(0.010)	(0.009)	(0.009)
EV*high	-0.033^{*}	-0.043^{**}	-0.012^{*}	-0.033^{*}	-0.038^{**}	-0.037^{**}
	(0.014)	(0.013)	(0.005)	(0.014)	(0.013)	(0.013)
Interc.	0.417^{***}	0.406^{***}	0.426^{***}	0.416^{***}	0.424^{***}	0.424^{***}
	(0.027)	(0.027)	(0.023)	(0.028)	(0.025)	(0.025)
AIC	3,698.1	3,385.8	6,450.9	3,593.3	19,039.3	18,563.1
Obs.	4,471	4,171	8,257	4,432	24,679	24,465

We provide the coefficient estimates and the standard errors in parentheses. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

We find that subjects in low volatility trials are more sensitive to the parameters than those in the high volatility trials. As such, our results from Table I are largely unchanged.

Here we conduct a robustness check of Table II. However we employ a logistic specification rather than a linear probability model.

Table $A3$: Estimates of γ_X and γ_C from 24 logistic random-effects regressions on common trials

	Trial		Sqrt. Trial		Second Half		Obs.
	γ_X	γ_C	γ_X	γ_C	γ_X	γ_C	
High	-0.0003	0.0003	-0.0056	0.0079	-0.0274	0.0248	$2,239$
31-300	(0.0002)	(0.0004)	(0.0050)	(0.0093)	(0.0299)	(0.0563)	
Low	-0.0003	0.0004	-0.0076	0.0113	-0.0434	0.0768	$2,193$
31-300	(0.0003)	(0.0005)	(0.0068)	(0.0128)	(0.0387)	(0.0786)	
High	-0.0001	0.0007	-0.0015	0.0147	-0.0214	0.1109	$1,508$
31-210	(0.0003)	(0.0006)	(0.0064)	(0.0132)	(0.0305)	(0.0724)	
Low	-0.0007	-0.0002	-0.0140	0.0022	$-0.0866^{\dagger}$	-0.0974	$1,471$
31-210	(0.0005)	(0.0009)	(0.0099)	(0.0186)	(0.0505)	(0.0887)	
High	0.0008	0.0005	0.0138	0.0046	0.0036	0.1188	771
31-120	(0.0010)	(0.0022)	(0.0162)	(0.0362)	(0.0483)	(0.1003)	
Low	0.0006	$0.0044^{\dagger}$	0.0071	0.0876^{*}	0.0332	0.0492	734
31-120	(0.0015)	(0.0025)	(0.0251)	(0.0430)	(0.0793)	(0.1382)	
High	0.0015	-0.0024	0.0216	-0.0401	0.0768	-0.1722	512
31-90	(0.0016)	(0.0038)	(0.0236)	(0.0574)	(0.0681)	(0.1272)	
Low	-0.0030	0.0140^{**}	-0.0480	0.2324^{**}	$-0.1643^{\dagger}$	0.5098^{*}	481
31-90	(0.0024)	(0.0050)	(0.0367)	(0.0787)	(0.0963)	(0.2017)	

We provide the coefficient estimates and the standard errors in parentheses. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Among our 48 estimates, here we find 4 significant at 0.05. However, each are the wrong sign as predicted by learning across trials. Further, 34 of our 48 estimates have the opposite sign as predicted by learning.

In Table $A4$, we conduct an analysis as summarized in Table III, but with logistic regressions.

In Table $A4$, we find that 25 of our 76 estimates are significant at 0.05, however each are the opposite sign as predicted by learning. Additionally, 59 of these 76 estimates have the opposite sign as predicted.

Here we conduct an analysis similar to that summarized in Table IV, however with logistic regressions.

Table A4: Estimates of γ_X and γ_C from 38 logistic random-effects regressions on common domain trials

	Trial		Sqrt. Trial		Second Half		After First 30		Obs.
	γ_X	γ_C	γ_X	γ_C	γ_X	γ_C	γ_X	γ_C	
High	-0.0002	0.0003	-0.0037	0.0062	-0.0347	0.0454	-0.0056	0.0151	$2,468$
1-300	(0.0002)	(0.0003)	(0.0036)	(0.0071)	(0.0275)	(0.0523)	(0.0562)	(0.0993)	
Low	-0.0004^{***}	$0.0005^{\dagger}$	-0.0093^{***}	$0.0120^{\dagger}$	-0.0641^{**}	0.0732	-0.0973^{***}	0.1739^{*}	$21,997$
1-300	(0.0001)	(0.0003)	(0.0026)	(0.0066)	(0.0199)	(0.0495)	(0.0258)	(0.0641)	
High	0.0000	0.0004	0.0010	0.0076	-0.0017	0.0765	0.0208	0.0048	$1,737$
1-210	(0.0003)	(0.0005)	(0.0054)	(0.0086)	(0.0320)	(0.0590)	(0.0610)	(0.0987)	
Low	-0.0006^{***}	0.0007	-0.0105^{***}	$0.0141^{\dagger}$	-0.0559^{***}	0.0645	-0.0865^{***}	0.1601^{**}	$15,486$
1-210	(0.0002)	(0.0005)	(0.0030)	(0.0082)	(0.0176)	(0.0486)	(0.0253)	(0.0610)	
High	0.0007	-0.0004	0.0101	-0.0055	0.0668	-0.1156	0.0350	-0.0408	$1,000$
1-120	(0.0007)	(0.0013)	(0.0099)	(0.0182)	(0.0483)	(0.0882)	(0.0628)	(0.1134)	
Low	-0.0009^{*}	0.0013	-0.0128^{**}	$0.0195^{\dagger}$	-0.0499^{*}	$0.1055^{\dagger}$	-0.0723^{**}	0.1472^{**}	$8,873$
1-120	(0.0003)	(0.0008)	(0.0048)	(0.0112)	(0.0254)	(0.0555)	(0.0254)	(0.0546)	
High	0.0004	-0.0015	0.0059	-0.0166	-0.0314	-0.0994	-0.0018	-0.0538	741
1-90	(0.0011)	(0.0021)	(0.0136)	(0.0257)	(0.0584)	(0.1164)	(0.0679)	(0.1313)	
Low	-0.0011^{*}	0.0031^{**}	-0.0147^{*}	0.0365^{**}	$-0.0470^{\dagger}$	0.1775^{**}	-0.0654^{**}	0.1690^{***}	$6,656$
1-90	(0.0005)	(0.0010)	(0.0060)	(0.0125)	(0.0284)	(0.0544)	(0.0245)	(0.0508)	
High	-0.0024	0.0021	-0.0190	0.0205	-0.0786	0.0476	-	-	476
1-60	(0.0024)	(0.0044)	(0.0240)	(0.0441)	(0.0826)	(0.1582)			
Low	-0.0018^{*}	$0.0029^{\dagger}$	-0.0191^{**}	0.0250	-0.0594^{*}	0.1130^{*}	-	-	$4,440$
1-60	(0.0007)	(0.0018)	(0.0069)	(0.0174)	(0.0242)	(0.0552)			

We provide the coefficient estimates and the standard errors in parentheses. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Table $A5$: Logistic random-effects regressions of risky choice

	$(1a)$	$(1b)$	$(4a)$	$(4b)$	$(5a)$	$(5b)$	$(6a)$	$(6b)$
high	–	−0.036	–	−0.059	–	1.257*	–	1.120*
		(0.960)		(0.959)		(0.708)		(0.71)
X	0.34***	0.35***	0.34***	0.36***	0.36***	0.352***	0.36***	0.35***
	(0.037)	(0.040)	(0.037)	(0.041)	(0.032)	(0.034)	(0.032)	(0.03)
C	−0.79***	−0.84***	−0.80***	−0.85***	−0.81***	−0.82***	−0.82***	−0.83***
	(0.080)	(0.091)	(0.082)	(0.093)	(0.072)	(0.076)	(0.07)	(0.09)
X*high	–	−0.067	–	−0.067	–	−0.031	–	−0.031
		(0.054)		(0.054)		(0.043)		(0.043)
C*high	–	0.147	–	0.149	–	−0.064	–	−0.059
		(0.114)		(0.115)		(0.087)		(0.087)
Prev_1	0.046	0.042	0.035	0.035	−0.026	−0.101	−0.039	−0.110
	(0.090)	(0.094)	(0.091)	(0.096)	(0.064)	(0.064)	(0.064)	(0.064)
X*Prev_1	−0.011**	−0.008*	−0.010**	−0.008*	−0.01***	−0.01***	−0.01***	−0.01***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)
C*Prev_1	0.017*	0.011	0.018*	0.012	0.028***	0.031***	0.028***	0.031***
	(0.008)	(0.009)	(0.008)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)
Interc.	0.833	0.877	0.815	0.852	0.587	0.8107	0.638	0.852
	(0.850)	(0.843)	(0.862)	(0.853)	(0.693)	(0.6241)	(0.698)	(0.628)
QIC	5,325.2	5,348.1	5,259.5	5,282.1	27,241.9	27,327.8	26,868.5	26,953.0
Obs.	4,471	4,471	4,432	4,432	24,596	24,596	24,382	24,382

We provide the coefficient estimates and the standard errors in parentheses. QIC refers to the Quasi-likelihood information criterion (Pan, 2001). *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

The results involving the parameters of interest are not changed.

Below we conduct an analysis similar to that summarized in Table IV, but with EV as an independent variable, rather than X and C.

Table $A6$: Linear random-effects regressions of risky choice

	$(1a)$	$(1b)$	$(4a)$	$(4b)$	$(5a)$	$(5b)$	$(6a)$	$(6b)$
high	–	0.018	–	0.018	–	−0.0003	–	−0.0003
		(0.047)		(0.047)		(0.0430)		(0.043)
EV	0.16***	0.16***	0.16***	0.16***	0.17***	0.18***	0.18***	0.18***
	(0.010)	(0.011)	(0.010)	(0.011)	(0.009)	(0.010)	(0.010)	(0.01)
EV*high	–	−0.020	–	−0.019	–	−0.010	–	−0.009
		(0.015)		(0.015)		(0.013)		(0.013)
Prev_1	0.001	0.0001	0.001	0.0004	0.002†	0.002†	0.002†	0.002†
	(0.002)	(0.0022)	(0.002)	(0.0022)	(0.001)	(0.001)	(0.001)	(0.001)
EV*Prev_1	−0.004***	−0.003**	−0.004***	−0.003**	−0.01***	−0.01***	−0.01***	−0.01***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Interc.	0.42***	0.42***	0.42***	0.42***	0.41***	0.42***	0.42***	0.42***
	(0.025)	(0.029)	(0.026)	(0.029)	(0.024)	(0.026)	(0.024)	(0.027)
AIC	3,703.4	3,712.4	3,598.2	3,607.4	18,904.6	18,915.2	18,429.8	18,440.5
Obs.	4,471	4,471	4,432	4,432	24,596	24,596	24,382	24,382

We provide the coefficient estimates and the standard errors in parentheses. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Here we find that there is a decreasing sensitive to parameter values that are different than the parameter values in the previous trial. In specifications that include the high volatility dummy variable, we find that these estimates are not significant. We conclude that the previous trial parameters better explain the differential sensitivity than the high volatility dummy variable.

Here we conduct a version of Table IV, but with Prev_2 not Prev_1.

Table $A7$: Linear random-effects regressions of risky choice

	$(1a)$	$(1b)$	$(4a)$	$(4b)$	$(5a)$	$(5b)$	$(6a)$	$(6b)$
high	–	0.015	–	0.012	–	0.361*	–	0.351*
		(0.215)		(0.215)		(0.159)		(0.157)
X	0.07***	0.07***	0.07***	0.07***	0.07***	0.08***	0.07***	0.08***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
C	−0.18***	−0.18***	−0.18***	−0.18***	−0.19***	−0.19***	−0.19***	−0.19***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.012)	(0.013)	(0.012)	(0.013)
X*high	–	−0.010	–	−0.010	–	−0.016*	–	−0.016*
		(0.009)		(0.009)		(0.008)		(0.008)
C*high	–	0.020	–	0.020	–	−0.003	–	−0.003
		(0.018)		(0.018)		(0.015)		(0.015)
Prev_2	0.005	0.005	0.002	0.002	−0.037**	−0.048**	−0.041**	−0.052***
	(0.022)	(0.023)	(0.022)	(0.023)	(0.014)	(0.015)	(0.014)	(0.015)
X*Prev_2	−0.002**	−0.002*	−0.002**	−0.002*	−0.002***	−0.002***	−0.002***	−0.001***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
C*Prev_2	0.005*	0.003	0.005*	0.003†	0.008***	0.008***	0.008***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Interc.	0.75***	0.74***	0.74***	0.74***	0.86***	0.80***	0.87***	0.81***
	(0.183)	(0.184)	(0.185)	(0.186)	(0.123)	(0.136)	(0.123)	(0.136)
AIC	3,940.3	3,954.4	3,844.4	3,858.4	18,309.1	18,317.7	17,827.1	17,836.1
Obs.	4,471	4,471	4,432	4,432	24,596	24,596	24,382	24,382

We provide the coefficient estimates and the standard errors in parentheses. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Two high volatility estimates involving X are significant at 0.05 and have the sign predicted by efficient coding. However, in these specifications, the estimates involving C are not significant. Also in these specifications, both of the Prev_2 estimates involving both X and C are significant at 0.001.

We conduct an analysis similar to that in Table IV, but with Prev_3 not Prev_1.

Table $A8$: Linear random-effects regressions of risky choice

	$(1a)$	$(1b)$	$(4a)$	$(4b)$	$(5a)$	$(5b)$	$(6a)$	$(6b)$
high	–	0.026	–	0.018	–	0.213	–	0.198
		(0.222)		(0.220)		(0.160)		(0.158)
X	0.07***	0.07***	0.07***	0.07***	0.08***	0.08***	0.08***	0.08***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.005)	(0.006)	(0.005)	(0.006)
C	−0.18***	−0.18***	−0.18***	−0.18***	−0.19***	−0.19***	−0.19***	−0.19***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.011)	(0.013)	(0.012)	(0.013)
X*high	–	−0.011	–	−0.011	–	−0.011	–	−0.011
		(0.009)		(0.009)		(0.008)		(0.008)
C*high	–	0.021	–	0.022	–	0.001	–	0.0015
		(0.019)		(0.019)		(0.015)		(0.014)
Prev_3	0.002	0.0004	0.0009	−0.0002	−0.004	−0.009	−0.006	−0.010
	(0.014)	(0.0153)	(0.0141)	(0.0155)	(0.009)	(0.009)	(0.009)	(0.009)
X*Prev_3	−0.001**	−0.0008	−0.001**	−0.0008	−0.002***	−0.002***	−0.002***	−0.002***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
C*Prev_3	0.003*	0.002	0.003*	0.0018	0.004***	0.004***	0.004***	0.004***
	(0.0012)	(0.001)	(0.0012)	(0.0011)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Interc.	0.77***	0.77***	0.76***	0.76***	0.73***	0.69***	0.73***	0.70***
	(0.190)	(0.187)	(0.192)	(0.189)	(0.124)	(0.135)	(0.124)	(0.135)
AIC	3,945.7	3,959.2	3,850.1	3,863.3	18,300.3	18,313.4	17,820.3	17,833.6
Obs.	4,471	4,471	4,432	4,432	24,596	24,596	24,382	24,382

We provide the coefficient estimates and the standard errors in parentheses. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

We do not find the differential sensitivity is related to the high volatility treatment. In several specifications, we find that evidence that Prev3 is related to the differential sensitivity.

We conduct an analysis similar to Table IV, but with Prev_4 not Prev_1.

Table $A9$: Linear random-effects regressions of risky choice

	$(1a)$	$(1b)$	$(4a)$	$(4b)$	$(5a)$	$(5b)$	$(6a)$	$(6b)$
high	–	0.019	–	0.016	–	$0.307^{\dagger}$	–	$0.298^{\dagger}$
		(0.216)		(0.216)		(0.159)		(0.157)
X	0.07^{***}	0.07^{***}	0.07^{***}	0.07^{***}	0.07^{***}	0.08^{***}	0.08^{***}	0.08^{***}
	(0.006)	(0.007)	(0.006)	(0.007)	(0.005)	(0.006)	(0.005)	(0.006)
C	-0.18^{***}	-0.18^{***}	-0.18^{***}	-0.18^{***}	-0.19^{***}	-0.19^{***}	-0.19^{***}	-0.19^{***}
	(0.015)	(0.015)	(0.015)	(0.015)	(0.011)	(0.013)	(0.012)	(0.013)
X*high	–	-0.011	–	-0.011	–	$-0.015^{\dagger}$	–	$-0.014^{\dagger}$
		(0.009)		(0.009)		(0.008)		(0.008)
C*high	–	0.021	–	0.021	–	-0.0004	–	-0.0004
		(0.018)		(0.018)		(0.0147)		(0.0148)
Prev_4	0.004	0.003	0.001	0.001	$-0.020^{\dagger}$	-0.028^{*}	-0.023^{*}	-0.031^{**}
	(0.018)	(0.019)	(0.018)	(0.019)	(0.011)	(0.012)	(0.011)	(0.012)
X*Prev_4	-0.002^{**}	$-0.001^{\dagger}$	-0.002^{**}	$-0.001^{\dagger}$	-0.002^{***}	-0.002^{***}	-0.002^{***}	-0.001^{***}
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
C*Prev_4	0.004^{*}	0.0021	0.004^{*}	0.0023	0.006^{***}	0.006^{***}	0.006^{***}	0.006^{***}
	(0.002)	(0.0014)	(0.0016)	(0.0014)	(0.001)	(0.001)	(0.001)	(0.001)
Interc.	0.75^{***}	0.75^{***}	0.75^{***}	0.75^{***}	0.81^{***}	0.75^{***}	0.82^{***}	0.76^{***}
	(0.184)	(0.184)	(0.186)	(0.186)	(0.122)	(0.135)	(0.122)	(0.134)
AIC	3,943.7	3,957.1	3,848.1	3,861.3	18,313.3	18,323.5	17,831.3	17,841.8
Obs.	4,471	4,471	4,432	4,432	24,596	24,596	24,382	24,382

We provide the coefficient estimates and the standard errors in parentheses. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Again, our results are similar to those summarized in Table IV.

We conduct a version of the analysis summarized in Table V, but with logistic regressions.

Table $A10$: Logistic random-effects regressions of risky choice

	(1)	(2)	(3)	(4)
increasing	0.387^{***}	0.388^{***}	0.448^{**}	0.366^{**}
	(0.114)	(0.114)	(0.149)	(0.123)
X	–	0.239^{*}	–	–
		(0.121)		
Interc.	0.795^{***}	-1.011	0.771^{***}	0.820^{***}
	(0.136)	(0.927)	(0.146)	(0.142)
QIC	2,613.8	2,613.2	1,001.7	1,613.5
Obs.	2,278	2,278	862	1,416

We provide the coefficient estimates and the standard errors in parentheses.

QIC refers to the Quasi-likelihood information criterion (Pan, 2001). *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Our results are largely unchanged from those in Table V.

Here we conduct a version of the analysis summarized in Table VI but with logistic regressions.

Table $A11$: Logistic random-effects regressions of risky choice

	(1)	(2)	(3)	(4)
increasing	0.388***	0.411***	0.240**	0.260***
	(0.114)	(0.115)	(0.076)	(0.0770)
X-7.5	0.239*	0.432**	0.173*	0.342**
	(0.121)	(0.159)	(0.082)	(0.115)
increasing*(X-7.5)	–	−0.449*	–	−0.354*
		(0.212)		(0.149)
Interc.	0.779***	0.771***	0.518***	0.511***
	(0.137)	(0.137)	(0.120)	(0.120)
QIC	2,613.2	2,611.8	4,112.9	4,111.4
Obs.	2,278	2,278	3,187	3,187

We provide the coefficient estimates and the standard errors in parentheses. QIC refers to the Quasi-likelihood information criterion (Pan, 2001). *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Our results are largely unchanged from those in Table VI.

We conduct an analysis similar to that in Table VII but with logistic regressions.

Table $A12$: Logistic random-effects regressions of risky choice

	(1)	(2)	(3)	(4)
increasing	0.086	0.091	−0.211	−0.199
	(0.286)	(0.287)	(0.248)	(0.249)
X-7.5	0.363*	0.675*	0.235*	0.513**
	(0.176)	(0.273)	(0.116)	(0.197)
increasing*(X-7.5)	–	−0.653†	–	−0.514*
		(0.344)		(0.236)
Interc.	1.022***	1.020***	0.813***	0.808***
	(0.188)	(0.189)	(0.173)	(0.174)
QIC	1,277.8	1,277.2	2,033.5	2,032.2
Obs.	1,143	1,143	1,594	1,594

We provide the coefficient estimates and the standard errors in parentheses. QIC refers to the Quasi-likelihood information criterion (Pan, 2001). *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

The results seem to be unchanged from those in Table VII.

We conduct a version of Table $VIII$ but with logistic regressions.

Table $A13$: Logistic random-effects regressions of risky choice

	(1)	(2)	(3)	(4)	(5)	(6)
increasing	−0.182	−0.074	−0.243	−0.371	−0.329	−0.362
	(0.260)	(0.271)	(0.305)	(0.235)	(0.240)	(0.257)
X-7.5	0.587***	0.377**	0.470**	0.395***	0.295**	0.314*
	(0.080)	(0.142)	(0.164)	(0.058)	(0.105)	(0.124)
C-2.5	−0.279†	−0.274†	−1.194†	−0.116	−0.112	−0.299
	(0.150)	(0.151)	(0.715)	(0.099)	(0.100)	(0.563)
increasing*(X-7.5)	–	0.252	0.156	–	0.118	0.099
		(0.169)	(0.188)		(0.124)	(0.141)
increasing*(C-2.5)	–	–	0.954	–	–	0.193
			(0.732)			(0.572)
Interc.	1.180***	1.119***	1.284***	0.886***	0.859***	0.892***
	(0.192)	(0.187)	(0.232)	(0.174)	(0.172)	(0.195)
QIC	6,444.2	6,441.3	6,441.1	10112.6	10113.5	10,114.0
Obs.	5,235	5,235	5,235	7,542	7,542	7,542

We provide the coefficient estimates and the standard errors in parentheses. QIC refers to the Quasi-likelihood information criterion (Pan, 2001). *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$.

Similar to Table $VIII$, we do not find the increasing dummy variable to be significant in any of our specifications. Also similar to Table $VIII$, we do not find any interaction terms to be significant at 0.05.