

**Attention constraints and learning in categories**

Rahul Bhui<sup>1</sup> and Peiran Jiao<sup>2</sup>

<sup>1</sup>Sloan School of Management, MIT

<sup>2</sup>Department of Finance, Maastricht University

September 22, 2022

Author note: We thank Florian Fröhlich for helpful research assistance. Support from the Pershing Square Fund for Research on the Foundations of Human Behavior and the MIT Sloan School of Management is gratefully acknowledged.

## Abstract

Many decision makers are thought to economize on attention by processing information at the simpler level of a category. We directly test whether such category focus reflects an adaptive response to attention constraints, in five preregistered experiments using an information sampling paradigm with mouse tracking. Consistent with rational principles, participants focus more on category-level information when individual differences are small, when the category contains more members, and when time constraints are more severe, though cognitive load has no effect. Participants are sensitive to the statistical structure of the category even when it must be learned from experience, and they respond to a latent shift in this structure. Beliefs about category members tend to cluster together more when category focus is high—a key element of rational inattention. However, this is counteracted by greater weight placed on salient and idiosyncratic information when the category is large. Our results broadly substantiate influential theories of categorical thinking, giving us a clearer view on the drivers and consequences of inattention.

*Word count:* 6,000

*Keywords:* categorical thinking; rational inattention; information choice

## **Attention constraints and learning in categories**

### **Introduction**

Our world is vast, but our attention is finite (Simon, 1971; Kahneman, 1973; Caplin, 2016). We thus have to split our attention across the immense array of information available to us. In many situations, decision makers cope with this complexity by processing information at the simpler level of a category. For instance, investors have limited time and effort and cannot learn about all of the countless stocks in a market. They might choose to study the value of an index rather than appraise each individual stock contained in that index (Peng & Xiong, 2006). Similarly, managers may pay more attention to macroeconomic data than firm-level signals (Kacperczyk, Van Nieuwerburgh, & Veldkamp, 2016), analysts may compile aggregated rather than segmented information about branches of a company (Bens, Monahan, & Steele, 2018), and multiproduct firms may be more responsive to aggregate demand shocks than good-specific shocks (Pasten & Schoenle, 2016).

Focusing on information at the category level leads agents to neglect heterogeneity among category members, resulting in economic anomalies. These anomalies are generally characterized by excess correlation in beliefs or outcomes of members in the same category and exaggerated differences across categories. For example, a range of inefficiencies have been tied to categorical investment patterns (aka “style investing”) in behavioral finance (Barberis & Shleifer, 2003), such as excess comovement of assets in the same class (Barberis, Shleifer, & Wurgler, 2005) and abnormal gains of companies that merely changed to dotcom names during the Internet bubble (Cooper, Dimitrov, & Rau, 2001). Phenomena like these (e.g., Gilbert, Kogan, Lochstoer, & Ozyildirim, 2012) can persist because attention is naturally taxed more as scale increases, in contrast to the traditional view that distortions will be minimized in thick markets with many participants.

Nevertheless, seminal theoretical work has shown how such category focus could be an individually efficient response when attention is scarce, because category-level signals

are informative about all category members, whereas idiosyncratic information pertains only to each member separately (Peng & Xiong, 2006; Maćkowiak & Wiederholt, 2009; Kacperczyk et al., 2016). This entails that attention to the category level versus the individual level should vary based on the costs and benefits induced by the information environment (as described below more specifically). These boundedly rational theories have provided influential explanations for the anomalous behavior of managers, firms, and households with large-scale economic consequences (reviewed briefly in the next section). However, despite the wide-ranging impact of these theories, we still lack direct evidence for the crucial assumption that people rationally balance attention to the category level and the individual level.

We conduct the most direct empirical test to date of rational inattention<sup>1</sup> applied to learning in categories, by developing a new laboratory paradigm. Our task was designed to transparently measure attention using mouse tracking, while precisely controlling the statistical structure of information via an abstract sampling paradigm. Across 5 pre-registered experiments, we test whether selective attention to category-level information adapts to the environment in line with rational principles.

People playing our “stock prediction game” had to accurately estimate the values of various hypothetical stocks based on a stream of incoming information.<sup>2</sup> These values were generated by a known categorical structure (following Peng & Xiong, 2006). Participants were told that the stocks were all in the same industry and so the value of each was equal to the arithmetic sum of two latent components, a common industry-level factor (reflecting the category average) and a unique stock-specific factor (reflecting individual deviation

---

<sup>1</sup> Throughout the paper, we use a broad definition of rational inattention which retains the core idea of selective attention that is adaptively deployed, but does not restrict it to information-theoretic formalisms. Our task similarly places some limits on the shape of attainable signals, so that we can directly and transparently measure attention. This allows us to focus on—and is consistent with—our motivating applications to categorical information choice (e.g., Kacperczyk et al., 2016) rather than the most subtle internal properties of information processing.

<sup>2</sup> Note that minimizing the variance of beliefs can emerge naturally from a broader optimization problem, such as in the setting of Peng and Xiong (2006) where agents invest in the stock market under budget and attention constraints.

from the average). These factors varied randomly and independently across periods.

In each period, participants could reveal noisy signals every half-second about any component (either the common industry factor or any one of the stock-specific factors) by hovering their mouse over the corresponding factor, until time ran out. They could only acquire signals for one factor at a time, and so might have to alternate between factors depending on their strategy. Longer time spent on a factor meant more signals were acquired. Therefore, time was a proxy for attention, consistent with both theoretical tradition (Sims, 2003) and high empirical correlations between the two (Caplin, Csaba, Leahy, & Nov, 2020). This link is commonly made in process-tracing studies (Willemsen & Johnson, 2011; Schulte-Mecklenbeck et al., 2017; Gabaix, 2019) and enshrined in popular sequential sampling models of information processing (Krajbich, 2019), as processing time is crucial in the brain's functioning (Pashler & Johnston, 1998; Nobre & Coull, 2010). In the field, viewing time of free, publicly available information is associated with reduced analyst error in earnings forecasts (Gibbons, Iliev, & Kalodimos, 2021). We thus investigated theoretical predictions by measuring the amount of time spent mousing over each factor, and evaluating how this changed when we manipulated properties of the environment.

Existing theory generates sharp predictions in our paradigm. If attention is rationally deployed, one's category focus should adjust flexibly based on the value of information attainable at each level. Several implications follow from this central idea (Peng & Xiong, 2006). First, when members of a category are similar, there is little to be gained by learning each one's unique qualities; thus, people should focus more on information at the category level when idiosyncratic variation is low relative to shared variation. Second, category-level information reduces uncertainty about every member, and so its value scales with category size; thus, people should focus more on information at the category level when the category contains many members. Finally, continuing to accumulate information about a given variable yields diminishing returns; thus, people should focus more on

information at the category level when they face severe attention constraints.

We found that people preferentially attended to information at the industry level when stock-specific variation was relatively lower (Experiments 1 and 5), when the industry contained more stocks (Experiment 2), and when time constraints were more severe (Experiment 3). There was no apparent effect of cognitive load caused by forcing signals to be kept in working memory (Experiment 4), which was consistent with the theory under the experimental design. People were sensitive to the prior variation at each level even when they were not given explicit information on these statistics or feedback on the accuracy of their predictions but had to learn them purely from the acquired signals; they were also able to adjust their attention allocation following a latent shift in the statistical structure of the category (Experiment 5).

We found further signatures of rational inattention in our data. When a person's category focus is higher, their predictions of stock values in a given period should be more similar to each other, since almost no differentiating information is being processed. Consistent with this, predictions tended to be less dispersed in periods with higher category focus—a key behavioral implication of models based on inattention (Peng & Xiong, 2006; Kacperczyk et al., 2016). Moreover, we found that category focus had a U-shaped relationship with prediction error that broadly matched the theoretical predictions in each experiment. Too much attention to the category leads to a detrimental neglect of individual differences, while too little attention means the category information is not being efficiently exploited. The intermediate allocation which balanced these opposing forces varied depending on the structure of the environment as described above.

Categorical attention, behavior, and performance were thus linked to each other and the environment in accordance with rational inattention—with one exception. In conditions where attention is more strained, value predictions should be less dispersed (controlling for category focus) because less individuating information is acquired. However, when there were many stocks in Experiment 3, predictions were more dispersed

rather than less. Model fitting revealed that in this condition, participants placed even more weight on the stock-specific signals and on the most recent signals, which were salient because their values were highlighted and displayed numerically (Bordalo, Gennaioli, & Shleifer, 2022). This finding suggests that as the task becomes more challenging, people may be more inclined to fixate on salient information, which can counteract the clustering of values that stems from categorical thinking.

Overall, we found that people adapted their degree of categorical focus broadly in line with rational principles. Our results substantiate core elements of influential theories of categorical information processing, while revealing how judgments might deviate from this benchmark. This work sharpens the link between categorical attention, behavior, and performance, giving us a clearer view on the drivers and consequences of inattention, and offering a reproducible platform for further investigations.

### **Rational Inattention and Learning in Categories**

Our results speak to many prominent applications of rational inattention built around the hypothesized cognitive mechanism. Peng and Xiong (2006) theoretically demonstrate how the optimal allocation of attention can lead investors to focus on category-level information. Combining attention allocation with portfolio allocation, their model reproduces several elements of style investing and empirical features of asset returns, such as excess comovement of assets within a category (Barberis & Shleifer, 2003). Kacperczyk et al. (2016) theoretically and empirically analyze the performance of mutual fund managers, and propose that an important part of manager skill involves properly balancing attention to macroeconomic aggregates and idiosyncratic firm-level data. They argue that in recessions (characterized by high aggregate volatility and price of risk), attention to aggregates should increase and fund outperformance should rise. Maćkowiak and Wiederholt (2009) posit that firms rationally attend more to idiosyncratic conditions when they vary more than aggregate conditions, which could explain why prices respond

rapidly to sector-specific shocks and slowly to monetary policy shocks. Maćkowiak and Wiederholt (2015) extend this analysis to include households, with analogous implications for consumption patterns. These applications all rest on the assumption that decision makers rationally adapt their category focus.

Although past empirical work has aimed to tease out implications of rational inattention in related market contexts (Peng, Xiong, & Bollerslev, 2007; Drake, Jennings, Roulstone, & Thornock, 2017; Huang, Huang, & Lin, 2019; Ehrmann & Jansen, 2020; Choi & Gupta-Mukherjee, 2022; Liu, Peng, & Tang, in press), field settings pose many challenges to researchers. Naturalistic information structures are often opaque and high-dimensional, categorization schemes and information processing capacity may vary widely among individuals and circumstances, the lack of controlled experimental variation makes it hard to establish what causes attention to shift, and attention itself is difficult to clearly measure. Analyses of field data thus require various indirect and assumption-laden methods to infer critically important variables like attention (Gabaix, 2019). Our experimental approach enables more straightforward tests.

Our research also relates to influential perspectives in cognitive science which maintain that attention is spent where it has maximum benefit (Gottlieb, 2018).<sup>3</sup> Past work has shown that when tasked with forming guesses of unknown feature values (i.e., inference; Markman & Ross, 2003), people are thought to concentrate on prototypical category information, and some have argued this focus depends on how the information is to be used (Rehder, Colner, & Hoffman, 2009; Braunlich & Love, 2022). Similar to the present setting, young children—another set of highly capacity-constrained agents—preferentially seek out information about kinds of things rather than about concrete individuals, such as facts applicable to all dogs rather than one specific dog (Cimpian & Park, 2014; Cimpian & Petro, 2014). This tendency seems sensitive to the

---

<sup>3</sup> This work fits into a wider body of research asserting that cognitive resources broadly are spent where they are most valuable (Bhui, Lai, & Gershman, 2021).

potential for information gain, as it only emerges when talking to knowledgeable adults. The rational predictions we test provide precise insight into conditions which stimulate attention to prototypical information, and have escaped experimental scrutiny so far.

### Experiments 1–4

**Participants.** Five hundred and eighty-four participants from the United States were recruited on Amazon Mechanical Turk, split across four experiments (Experiment 1,  $n = 147$ ; Experiment 2,  $n = 145$ ; Experiment 3,  $n = 146$ ; Experiment 4,  $n = 146$ ). They were paid a base of \$2 plus a bonus of up to \$6 that depended on performance as described below. Sample sizes were documented in the preregistrations, and based on monetary constraints, informed partly by pilot studies which suggested that this number should be sufficient to detect the hypothesized effects if present. To help ensure data quality, participation was restricted to workers who had completed 100+ HITs with at least a 98% approval rate. Participants provided informed consent, and the study was approved by the Harvard Committee on the Use of Human Subjects (Experiments 1–3) or the MIT Committee on the Use of Humans as Experimental Subjects (Experiment 4). The preregistrations can be found at <https://osf.io/6mcqy> (Experiments 1–3) and <https://osf.io/245ks> (Experiment 4).

**Procedure.** All experiments used the same “stock prediction game” paradigm pictured in Figure 1, with some variations. In each period, participants had to estimate the values of several hypothetical stocks after selectively acquiring a stream of information about the components of value (a common industry-level factor and idiosyncratic stock-specific factors). The stocks were abstractly labeled  $A$ ,  $B$ ,  $C$ , and so on. Values were given by the arithmetic sum of the two factors, meaning each stock’s value was equal to the common industry-level factor plus an idiosyncratic stock-specific factor. These factors were generated independently in every period from zero-mean Gaussian distributions portrayed on screen by sideways bell curves (so participants were shown these priors). Hence, the

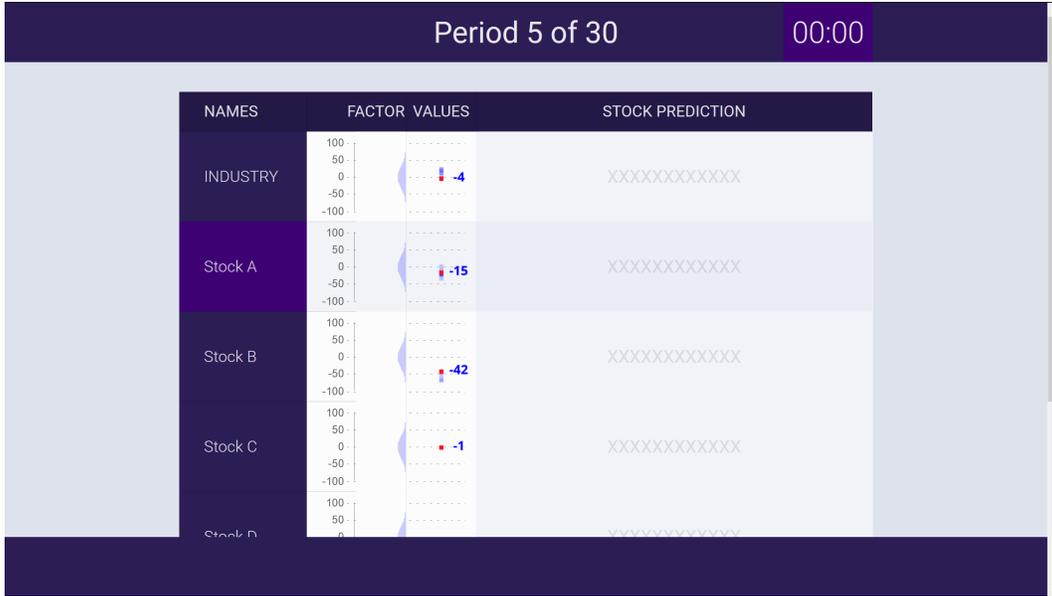


Figure 1. Screenshot of Experiment. Participants had to predict the values of hypothetical stocks, which were given by the sum of a common industry factor and idiosyncratic stock-specific factors. Noisy signals could be acquired moment-to-moment by mousing over any factors until time ran out.

periods were effectively repetitions of the same estimation problem. For simplicity, all stock-specific factors had the same prior distribution. In short, letting  $c \sim \mathcal{N}(0, \sigma_{industry}^2)$  be the industry factor and  $m_A$ ,  $m_B$ , and  $m_C \sim \mathcal{N}(0, \sigma_{stock}^2)$  be stock-specific factors in a given period, participants had to make their best guess as to the stock values  $v_A = c + m_A$ ,  $v_B = c + m_B$ , and  $v_C = c + m_C$ .

However, the exact factor values were not explicitly provided, and participants instead had to learn about them by mousing over the corresponding factor. While their mouse cursor was positioned over a given factor, a noisy Gaussian signal of its true value would be revealed every 500 milliseconds, drawn accordingly from  $\mathcal{N}(c, \sigma_{signal}^2)$  or  $\mathcal{N}(m_i, \sigma_{signal}^2)$ . Participants could mouse over any factor they wanted at any moment before a limited budget of time ran out, which was represented by an on-screen timer.

After time expired in a period, participants recorded their point prediction of every stock's total value (the sum of its two relevant factors) using a set of sliders. They did not provide predictions for the industry factor or any stock-specific factors by themselves. This phase had no time limit. Upon submitting these predictions, they were shown the true

stock values and the magnitudes of their errors (only for feedback, as these did not affect any random variables in subsequent periods). At the end of the experiment, they were paid a bonus based on the mean squared error of their predictions in each period according to a quadratic loss function of which they were informed.<sup>4</sup>

We used a within-subjects design in each of the four experiments. All experiments consisted of two blocks of 10 periods each. Within each block the design parameters were fixed, and the treatments occurred across blocks. We implemented the following treatments, which were expected to increase the relative attention paid to the industry factor (for reasons laid out in the Introduction):

- Experiment 1’s treatment increased the relative prior variance of the industry factor (simultaneously increasing  $\sigma_{industry}$  from 5 to 30 and decreasing  $\sigma_{stock}$  from 30 to 5, to keep the total variance constant).
- Experiment 2’s treatment increased the number of stocks (raising  $n_{stocks}$  from 2 to 8).
- Experiment 3’s treatment decreased the time budget (reducing available time from 20 seconds to 8 seconds).
- Experiment 4’s treatment increased the degree of cognitive load (rather than the signals remaining onscreen during the prediction stage, they vanished right as the next signal appeared).

The design parameters for each experiment are documented in Table 1. The order of blocks was counterbalanced. The serial position of the industry factor was also counterbalanced across subjects, but kept the same across blocks for any given subject. In Experiments 1–3, the last signal for each factor was highlighted and its number was displayed as pictured in Figure 1.

---

<sup>4</sup> The total payoff function was given by  $\$6 - \sum_{t=1}^{20} MSE_t/200$  where  $t$  denotes the period. The penalty term was capped at  $\$6/20 = \$0.30$  in each period so that the total payment would not drop below \$0.

	$\sigma_{industry}$	$\sigma_{stock}$	$\sigma_{signal}$	$n_{stock}$	<i>time</i>	<i>vanish</i>
<i>Exp. 1</i> – low category variance	5	30	10	5	12 s	N
high category variance	30	5	10	5	12 s	N
<i>Exp. 2</i> – few category members	30	5	10	2	12 s	N
many category members	30	5	10	8	12 s	N
<i>Exp. 3</i> – long time limit	30	5	10	5	20 s	N
short time limit	30	5	10	5	8 s	N
<i>Exp. 4</i> – signals remain	20	20	10	5	12 s	N
signals vanish	20	20	10	5	12 s	Y
<i>Exp. 5</i> – low category variance	1	30	10	5	12.75 s	Y
high category variance	30	1	10	5	12.75 s	Y
equal category variance	15	15	10	5	12.75 s	Y

Table 1

*Experimental Design Parameters.* In Experiment 5, the low and high category variance conditions occurred in the first half of periods, and the equal category variance condition occurred in the second half.

After reading the instructions, participants were provided with two self-paced practice periods in which they were told the true values of each factor and allowed unlimited time to sample information. This was intended to clearly explicate the task structure. They were subsequently asked two basic comprehension check questions to verify their understanding of the task (see Appendix B).

**Exclusions.** Based on our preregistered criteria, we exclude participants who failed to correctly answer the two comprehension check questions, participants who spent more than 20% of their available viewing time on average not mousing over any factor, and then remaining periods in which more than 20% of the available viewing time was spent not mousing over any factor. As these were relatively fine-grained measures of attention, this left 100 (Experiment 1), 91 (Experiment 2), 100 (Experiment 3), and 114 (Experiment 4) participants remaining in the analyses. We also exclude a small number of periods across the experiments where responses were made outside of the permissible slider bounds due to technical glitches (3 periods in Experiment 1, 3 in Experiment 2, and 1 in Experiment 3).

## Results

First, we note that participants were able to perform reasonably well in the task. Predictions were moderately to highly correlated with true values, with median correlations ranging from 0.520–0.928 across experiments (see Appendix D).

**Category focus.** Our primary variable of interest is the category focus, which we define as the fraction of time spent mousing over the industry factor compared to the average stock-specific factor. For example, if out of the 12 second time limit, a participant spent 7 seconds attending to the industry factor and a total of 5 seconds attending to the 5 stock-specific factors (meaning an average of 1 second per stock-specific factor), the category focus would be  $7/12 - 1/12 = 0.5$ . This metric was used because it scales appropriately with the time limit and number of stocks, and is motivated by the theoretical model we draw upon (see Appendix A for our streamlined Bayesian variant of the model in Peng & Xiong, 2006).

The category focus is displayed for each experimental condition in Figure 2. The treatment effects appear to be in line with the first three predictions. These conclusions are formally supported by Bayesian random effects models reported in Table 2, which predict category focus based on the treatment condition, with subject-specific coefficients for both intercept and treatment effect.<sup>5</sup> The regressions indicate positive effects of higher category-level variance ( $P(\beta_{variance} > 0) > 0.999$ ), category size ( $P(\beta_{size} > 0) > 0.999$ ), and time pressure ( $P(\beta_{time} > 0) = 0.956$ ), but not of vanishing signals ( $P(\beta_{load} > 0) = 0.679$ ). Our preregistered test criterion of  $P(\beta > 0) > 0.95$  is met in the first three cases. Although the null effect from Experiment 4 ran counter to our initial expectations, as we will see later, it turns out to be consistent with the model due to the experiment’s different design parameters. Thus, participants appear to alter their patterns of attention as predicted by

---

<sup>5</sup> The regressions were computed using the *brms* package in R version 4.0.5 with default weakly informative priors (Bürkner, 2017).

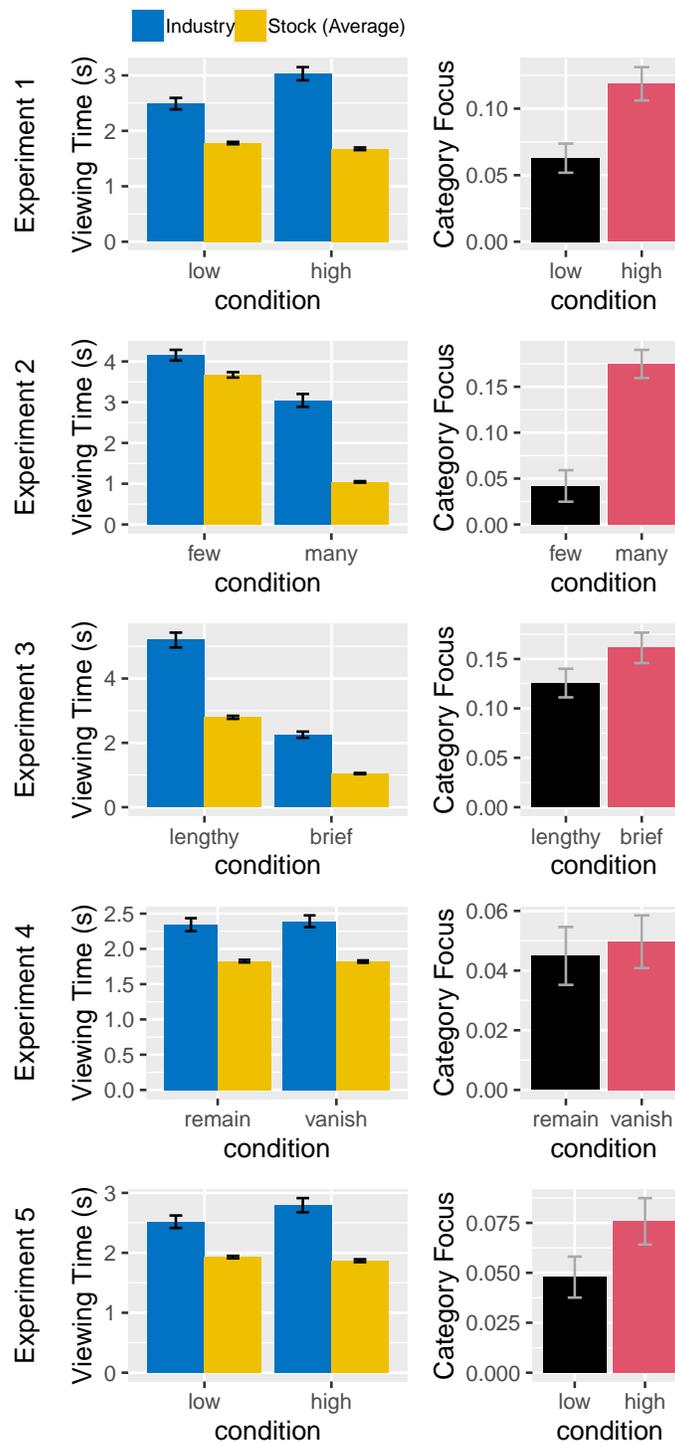


Figure 2. *Attention Allocation Patterns*. (Left) Mean time spent attending to each factor; (right) category focus (i.e., difference between proportions of time spent on category and average member). Low vs high relative category variance (Experiment 1 and first half of Experiment 5), few vs many category members (Experiment 2), lengthy vs brief time limit (Experiment 3), and signals remaining onscreen or vanishing (Experiment 4). Error bars depict 95% confidence intervals.

	Coefficient	Category focus		
		Posterior mean	Posterior 95% CI	$P(\beta > 0)$
<i>Exp. 1</i>	Intercept	0.069	[0.042, 0.098]	>0.999
	Higher category variance	0.053	[0.026, 0.079]	>0.999
<i>Exp. 2</i>	Intercept	0.038	[0.000, 0.076]	0.975
	Larger category size	0.131	[0.091, 0.174]	>0.999
<i>Exp. 3</i>	Intercept	0.129	[0.091, 0.169]	>0.999
	Shorter time limit	0.030	[-0.006, 0.066]	0.956
<i>Exp. 4</i>	Intercept	0.046	[0.026, 0.064]	>0.999
	Data points vanish	0.004	[-0.013, 0.021]	0.679
<i>Exp. 5</i>	Intercept	0.048	[0.036, 0.059]	>0.999
	Higher category variance	0.028	[0.012, 0.043]	>0.999

Table 2

*Treatment Effects on Category Focus. Posterior estimates from Bayesian random effects models predicting category focus from experimental condition.  $P(\beta > 0)$  denotes the posterior probability that the coefficient is positive. First half of periods included for Experiment 5. CI = credible interval.*

the theory.<sup>6</sup>

**Prediction dispersion.** We investigate another key behavioral signature of category thinking under rational inattention: when category focus is higher, predictions of stock values in a given period should be more similar to each other, because less individuating information is obtained.

The standard deviation<sup>7</sup> of participants' stock predictions is plotted conditional on the category focus in Figure 3. We also derive and plot the theoretical relationship between the two for comparison (see Appendix A for details), parameterized based on the experimental design.<sup>8</sup> The model implies that as category focus increases, the variance

<sup>6</sup> Due to the i.i.d. Gaussian nature of the signals, the theory does not constrain the dynamic sequence of information processing. Because we accordingly focus on the overall amount of attention paid to different information sources rather than the sequence, we report analyses of the attention trajectory in Appendix E, along with analyses of learning across periods. We also explore the effect of the factors' serial positions on attention in Appendix F.

<sup>7</sup> Due to high skewness, we log transform the standard deviation and, later, the error of the stock predictions.

<sup>8</sup> The plotted theoretical predictions for Experiment 4 assume that when signals vanish, it translates into cutting signal precision in half. Qualitative implications are not appreciably different with other fractions.

across stock predictions in a period should decline from the true prior variance of the stock-specific factor down to 0 (plus any baseline response noise due, for instance, to the slider interface<sup>9</sup>). Consistent with this implication, predictions of stocks are generally more similar (i.e., their standard deviation is lower) when category focus is higher. Table 3 contains the results of Bayesian random effects models revealing this relationship.

Variable	$\log(1 + \text{Prediction SD})$	
<i>Exp. 1</i>	Higher category variance	Lower category variance
Intercept	2.317 [2.255, 2.377] (<.001)	3.287 [3.244, 3.330] (<.001)
Category focus	0.367 [-0.981, 1.697] (.291)	-1.303 [-2.932, 0.222] (.954)
Category focus <sup>2</sup>	0.551 [-0.833, 2.062] (.227)	0.017 [-1.466, 1.404] (.486)
<i>Exp. 2</i>	Larger category size	Smaller category size
Intercept	2.387 [2.302, 2.475] (<.001)	1.912 [1.821, 2.001] (<.001)
Category focus	-2.165 [-3.957, -0.150] (.984)	-1.457 [-3.648, 0.790] (.897)
Category focus <sup>2</sup>	-0.144 [-1.618, 1.403] (.577)	-0.274 [-2.602, 2.203] (.600)
<i>Exp. 3</i>	Shorter time limit	Longer time limit
Intercept	2.362 [2.298, 2.424] (<.001)	2.341 [2.273, 2.414] (<.001)
Category focus	-1.972 [-3.364, -0.526] (.996)	-1.171 [-2.803, 0.662] (.911)
Category focus <sup>2</sup>	-0.732 [-2.011, 0.612] (.862)	-1.221 [-2.735, 0.529] (.919)
<i>Exp. 4</i>	Data points remain	Data points vanish
Intercept	2.846 [2.798, 2.895] (<.001)	2.989 [2.956, 3.023] (<.001)
Category focus	-2.609 [-3.993, -1.203] (>.999)	-1.331 [-2.506, -0.160] (.984)
Category focus <sup>2</sup>	-0.227 [-1.658, 1.297] (.624)	-0.991 [-2.212, 0.228] (.949)
<i>Exp. 5</i>	Higher category variance	Lower category variance
Intercept	2.362 [2.286, 2.440] (<.001)	3.152 [3.099, 3.204] (<.001)
Category focus	0.220 [-1.261, 1.769] (.393)	-2.132 [-3.334, -0.993] (>.999)
Category focus <sup>2</sup>	-0.975 [-2.628, 0.709] (.393)	-1.905 [-3.131, -0.772] (.999)

Table 3

*Regression Results: Prediction Dispersion and Category Focus. Posterior mean estimates from Bayesian random effects models with prediction dispersion regressed on category focus (using orthogonal polynomials). 95% credible intervals in brackets and  $P(\beta < 0)$  in parentheses.*

<sup>9</sup> When plotting the theoretical predictions, we roughly calibrate the level of response noise to the data. Specifically, we set the standard deviation of response noise to be the 10th percentile of prediction standard deviations across all included periods in each experiment. This provides a balance between finding the minimal level of variation (corresponding to irreducible noise) while avoiding undue influence from outliers (either participants or periods with unusually low noise).

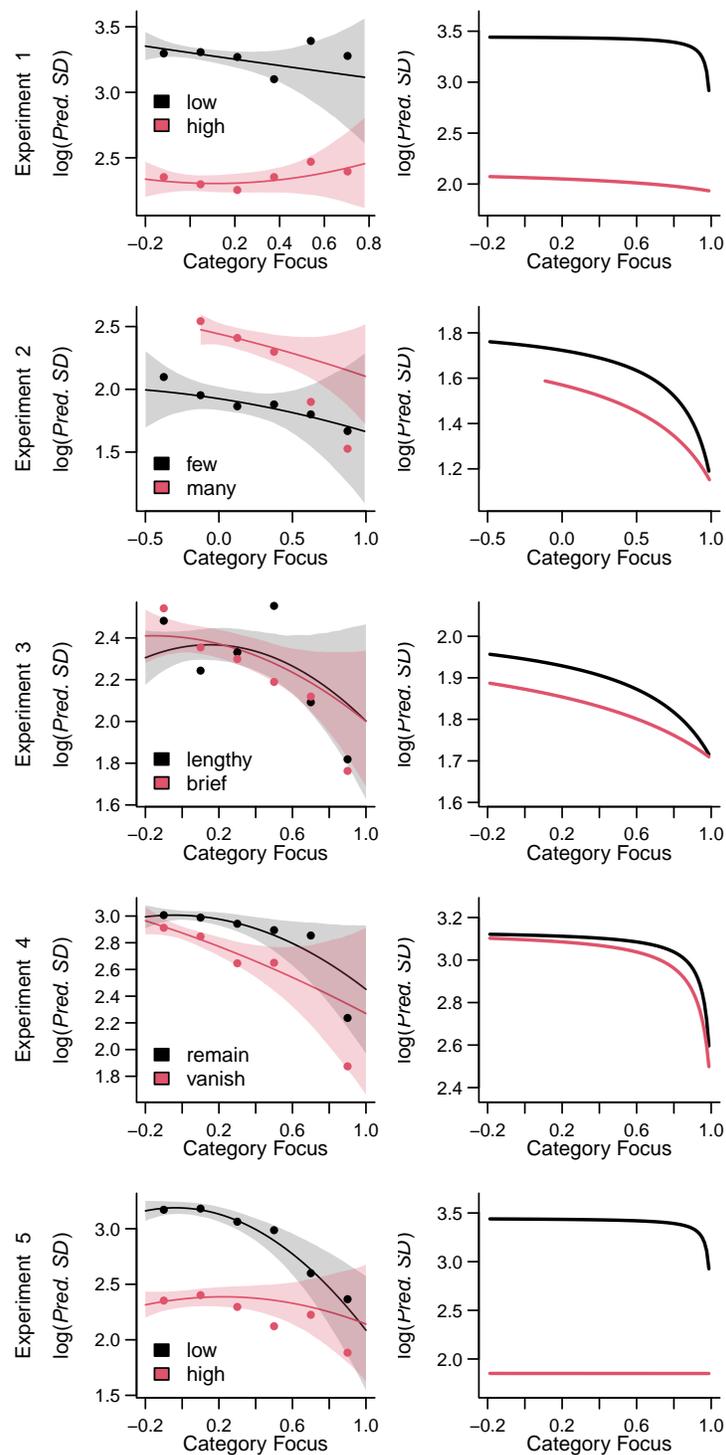


Figure 3. *Prediction Dispersion and Category Focus.* The dependent variable  $\log(\text{Pred. SD})$  is the logarithm of 1 plus the standard deviation across stock value predictions in a given period. (Left) Data, binned averages with quadratic regression lines and 95% credible intervals from Bayesian random effects models. (Right) Theory, plus response noise.

The theory also implies that prediction dispersion should be lower in conditions where attention is more strained, holding category focus constant. A violation of this is apparent in Experiment 2, though: predictions are more variable, rather than less, when there are many stocks. We investigate this finding further using model fitting to capture how individual participants form predictions based on the signals they acquire (see Appendix C). The model includes parameters to reflect the weight placed on industry signals, the weight placed on stock signals, the extra weight placed on the most recent signal from each factor, and response noise.<sup>10</sup> This analysis reveals that when there are many stocks (compared to when there are few), participants place even more weight on the stock-specific signals and on the most recent signals ( $P(\beta > 0) = .990$  and  $.996$  respectively, Bayesian signed rank test; see Table S1). The latter are salient because their values are displayed numerically and highlighted in red (Bazley, Cronqvist, & Mormann, 2021). Such overweighting of a few noisy signals for each factor makes the predictions more dispersed. Regression analysis confirms that these weight parameters are indeed positively associated with dispersion in the stock predictions ( $P(\beta > 0) = .999$  and  $.994$ ; see Tables S2 and S3). This result suggests that when the task is complex due to scale, people may be inclined to fixate on salient information, which can counteract the correlation in predictions that stems from categorical thinking.

**Prediction accuracy.** While category focus is monotonically related to prediction dispersion, it should have a curvilinear relationship with prediction error. Attending too much to the category leads the agent to neglect heterogeneity, while attending too little prevents the agent from efficiently drawing upon the category information. An intermediate level of category focus balances these considerations (with corner solutions obtaining in more extreme cases). The exact location of this optimum depends on the environmental structure as described earlier.

---

<sup>10</sup> The model captured individual judgments well, as model predictions were moderately to very highly correlated with participant responses, with median correlations ranging from 0.590–0.934 across experiments (see Appendix D).

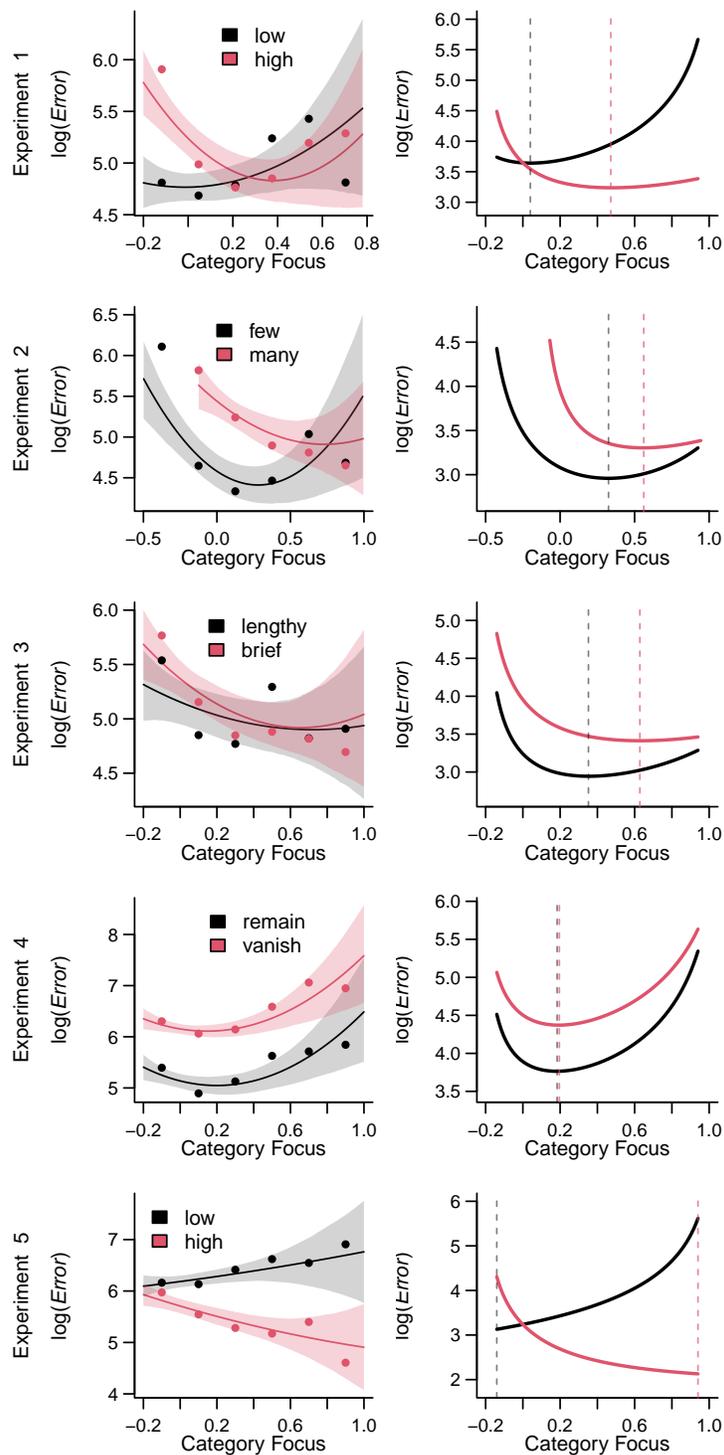


Figure 4. *Prediction Error and Category Focus.* The dependent variable  $\log(\text{Error})$  is the logarithm of 1 plus the mean squared error in stock value predictions in a given period. (Left) Data, binned averages with quadratic regression lines and 95% credible intervals from Bayesian random effects models. (Right) Theory, plus response noise; dashed lines indicate error-minimizing levels of category focus.

Variable	$\log(1 + \text{Prediction Error})$	
<i>Exp. 1</i>	Higher category variance	Lower category variance
Intercept	5.131 [4.968, 5.292] (<.001)	4.809 [4.678, 4.935] (<.001)
Category focus	-6.015 [-9.057, -2.777] (>.999)	2.429 [-0.593, 5.345] (.056)
Category focus <sup>2</sup>	4.572 [1.430, 8.142] (.004)	1.733 [-1.036, 4.648] (.112)
<i>Exp. 2</i>	Larger category size	Smaller category size
Intercept	5.269 [5.110, 5.427] (<.001)	4.676 [4.488, 4.876] (<.001)
Category focus	-4.785 [-8.112, -1.391] (.997)	-6.812 [-10.434, -2.957] (>.999)
Category focus <sup>2</sup>	2.365 [-0.806, 5.591] (.078)	7.207 [3.231, 11.112] (<.001)
<i>Exp. 3</i>	Shorter time limit	Longer time limit
Intercept	5.236 [5.094, 5.382] (<.001)	5.099 [4.917, 5.277] (<.001)
Category focus	-5.715 [-8.904, -2.469] (>.999)	-2.705 [-5.955, 0.757] (.942)
Category focus <sup>2</sup>	2.494 [-0.534, 5.938] (.061)	1.179 [-1.812, 4.208] (.223)
<i>Exp. 4</i>	Data points remain	Data points vanish
Intercept	6.174 [6.081, 6.268] (<.001)	5.160 [5.021, 5.295] (<.001)
Category focus	0.748 [-1.593, 3.036] (.254)	0.149 [-2.421, 2.696] (.458)
Category focus <sup>2</sup>	3.781 [1.334, 6.382] (.002)	4.361 [1.440, 7.404] (.003)
<i>Exp. 5</i>	Higher category variance	Lower category variance
Intercept	5.640 [5.506, 5.774] (<.001)	6.210 [6.104, 6.320] (<.001)
Category focus	-6.554 [-9.580, -3.505] (>.999)	2.710 [0.119, 5.245] (.020)
Category focus <sup>2</sup>	0.615 [-2.314, 3.635] (.338)	0.194 [-2.220, 2.684] (.439)

Table 4

*Regression Results: Prediction Error and Category Focus. Posterior mean estimates from Bayesian random effects models with prediction error regressed on category focus (using orthogonal polynomials). 95% credible intervals in brackets and  $P(\beta < 0)$  in parentheses.*

Error in the stock predictions is plotted against the category focus in Figure 3. Table 4 contains the results of Bayesian random effects models capturing this relationship. We also derive and plot the theoretical relationship between the two for comparison (see Appendix A for details), parameterized based on the experimental design as before. Participants exhibit more error than the theoretical bound because perceptual limitations prevent them from extracting the entire information content of the signals. However, the shapes of the theoretical curves are broadly recapitulated in the data. The empirical incentive structure of the task thus seems commensurate with the theory.

These plots reveal that in Experiment 4, the optimal level of category focus is scarcely affected by the change in signal precision. This null effect occurs largely because the industry- and stock-level prior variances were equal in that experiment, negating the benefit of focusing on the category. Hence, the theory entails that under the actual task conditions, no effect should be expected. This observation underscores the importance of formal models that predict how category focus emerges from a complex interaction between environmental variables.

### Experiment 5

In the previous experiments, participants were shown the prior variances of the industry and stock-specific factors. However, in many natural settings, this internal category structure might be unknown and would have to be learned from experience. Moreover, it could even change over time without being explicitly signposted. Experiment 5 explores whether people can cope under these more challenging conditions in two ways. First, can people adjust their attention allocation when the category's internal statistics must be learned from minimal information? Second, does the degree of category focus adapt when these statistics surreptitiously change? This investigation probes the boundaries of the claim that rational inattention applies well in repeated situations, where agents might discover the optimal strategy through experience (Maćkowiak, Matějka, & Wiederholt, in press).

**Participants.** Two hundred and ninety-nine participants from the United States were recruited on Amazon Mechanical Turk according to the same criteria as the previous experiments. They were paid a base of \$3 plus a bonus of up to \$6 based on performance in the same way as the previous tasks. The sample size was increased because the treatment (described below) was between-subjects rather than within-subjects. Participants provided informed consent, and the study was approved by the MIT Committee on the Use of Humans as Experimental Subjects. The preregistration can be

found at <https://osf.io/y8jcq>.

**Procedure.** The basic elements of the task were similar to those of Experiment 1. However, in the first block of 10 periods, half of the participants experienced high relative category variance ( $\sigma_{industry} = 30$ ,  $\sigma_{stock} = 1$ ) and the other half experienced low relative category variance ( $\sigma_{industry} = 1$ ,  $\sigma_{stock} = 30$ ). In the second block of 10 periods, all participants encountered equal variances at both levels ( $\sigma_{industry} = 15$ ,  $\sigma_{stock} = 15$ ). The blocks were not explicitly demarcated, meaning there was no overt sign of this transition. The design parameters are documented in Table 1.

**Exclusions.** We use the same preregistered exclusion criteria as the previous experiments, leaving 228 participants remaining in the analysis.

## Results

For consistency with the previous experiments, we conduct similar analyses and construct the same plots using the data from the first half of the periods in Experiment 5. The results are displayed in Figures 2, 3, and 4. They replicate our results from Experiment 1, with small differences due to variation in the design parameters.

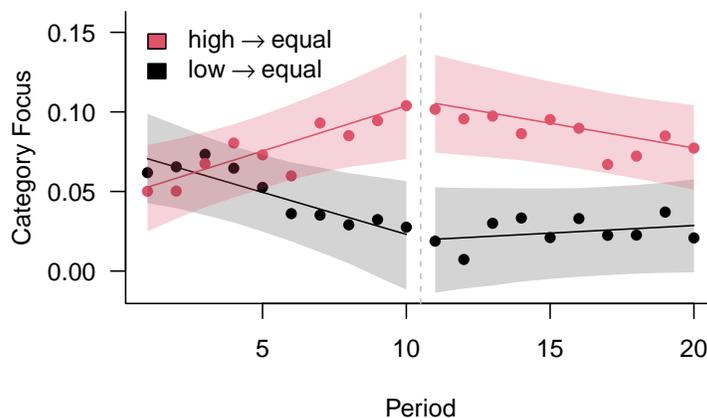


Figure 5. Dynamics of Category Focus in Experiment 5. Dashed gray line indicates the change point in the variance structure. Regression lines shown from Bayesian random effects model with 95% credible intervals.

**Learning dynamics.** The dynamics of category focus are displayed in Figure 5. Category focus begins at the same level in each condition, as there is hardly any way to

Coefficient	Category focus	
	First half	Second half
Intercept	0.071 [0.043, 0.099] (>.999)	0.020 [-0.014, 0.053] (.878)
High cond. (first)	-0.018 [-0.057, 0.022] (.189)	0.085 [0.040, 0.130] (>.999)
Period	-0.005 [-0.010, -0.001] (.009)	0.001 [-0.002, 0.004] (.735)
High cond. (first) $\times$ Period	0.011 [0.005, 0.017] (>.999)	-0.004 [-0.008, 0.000] (.024)

Table 5

*Regression Results: Dynamics of Category Focus in Experiment 5. Posterior estimates from Bayesian random effects models with category focus regressed on experimental condition and period. 95% credible intervals in brackets and  $P(\beta > 0)$  in parentheses.*

determine the statistics with such little data. However, patterns of attention diverge across the first half of the task as environmental statistics are learned, consistent with rational principles. Category focus increases in the condition with high relative category variance where stocks have nearly identical values, while it decreases in the condition with low relative category variance where stocks have nearly uncorrelated values (interaction  $P(\beta > 0) > .999$ ; see Bayesian random effects regressions reported in Table 5).

After the latent shift occurs halfway through the task, these trends change. Category focus starts to converge again (interaction  $P(\beta < 0) = .976$ ), which makes sense as both conditions then have the same environmental statistics. Persistent changes do remain throughout the task. Category focus remains higher in the condition starting with higher category focus ( $P(\beta > 0)$  greater than 0.95 for periods 7 onward, Bayesian  $t$ -tests).

### Concluding Remarks

Influential theories rely on the pivotal assumption that learning at the category level emerges from the rational allocation of limited attention (Peng & Xiong, 2006; Maćkowiak & Wiederholt, 2009; Kacperczyk et al., 2016). We conduct the most direct empirical test of this assumption to date. We develop an abstract sampling paradigm which lets us tightly control the structure of the information environment and reveal attention transparently using mouse tracking. These design features enable us to directly weigh the data against

implications of rational inattention. The results indicate that people flexibly adjust their patterns of attention to category information broadly in line with rational principles.

Future research can build on our work in several ways. First, although our study assumed a fixed category structure, inattention to individual differences could increase the stability of category representations by leading people to neglect evidence that certain members should be recategorized (Kruschke, 2011). For example, in finance, this would be consistent with lengthy persistence of asset classes until long stretches of poor performance trigger a reclassification (Barberis & Shleifer, 2003). Meaningful, testable interactions can hence occur between inference within a category and classification into a category.

Second, more complex interactions between attention and decision making should be experimentally characterized. The objective function we used was meant to reflect a generic goal involving the accurate evaluation of each individual item. This allowed us to cleanly isolate the formation of beliefs, increasing the general applicability of our findings. However, it may not fully capture the ultimate goal of the decision maker in any given scenario, such as investing in a portfolio, although the two are related in some cases (Peng & Xiong, 2006). It would be valuable to investigate further how the downstream uses of information affect the manner in which information is processed upstream (Gottlieb, 2018). Our task serves as a useful jumping-off point for such studies.

Third, other tests will be needed to determine whether our findings extend to more covert forms of attention (Carrasco, 2011). In contrast to some (but not all) models of rational inattention (e.g., Mondria, 2010), participants could not shape information with full flexibility. This may be appropriate for settings where information is available in restricted forms, and might serve as a sensible approximation in other cases, but testing subtler aspects of attention would require another paradigm.

Finally, our results indicate the need for theories that integrate both “top down” task relevance and “bottom up” stimulus salience into a unified framework. For example, Heinke (2019) extends a model of rational inattention to incorporate the salience or

visibility of signals, producing a rich interplay between active and passive information processing. Our findings suggest that the scale of the decision problem may influence attention via both of these pathways. A unified formal perspective on attention is essential for a truly complete framework.

## References

- Barberis, N., & Shleifer, A. (2003). Style investing. *Journal of Financial Economics*, *68*(2), 161–199.
- Barberis, N., Shleifer, A., & Wurgler, J. (2005). Comovement. *Journal of Financial Economics*, *75*(2), 283–317.
- Bazley, W. J., Cronqvist, H., & Mormann, M. (2021). Visual finance: the pervasive effects of red on investor behavior. *Management Science*, *67*(9), 5616–5641.
- Bens, D. A., Monahan, S. J., & Steele, L. B. (2018). The effect of aggregation of accounting information via segment reporting on accounting conservatism. *European Accounting Review*, *27*(2), 237–262.
- Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, *41*, 15–21.
- Bordalo, P., Gennaioli, N., & Shleifer, A. (2022). Saliency. *Annual Review of Economics*, *14*(1), 521–544.
- Braunlich, K., & Love, B. C. (2022). Bidirectional influences of information sampling and concept learning. *Psychological Review*, *129*(2), 213–234.
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.
- Caplin, A. (2016). Measuring and modeling attention. *Annual Review of Economics*, *8*, 379–403.
- Caplin, A., Csaba, D., Leahy, J., & Nov, O. (2020). Rational inattention, competitive supply, and psychometrics. *Quarterly Journal of Economics*, *135*(3), 1681–1724.
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525.
- Choi, H. M., & Gupta-Mukherjee, S. (2022). Analysts’ reliance on industry-level versus firm-specific information: Implications for information production. *Journal of Banking & Finance*, 106555.

- Cimpian, A., & Park, J. J. (2014). Tell me about Pangolins! Evidence that children are motivated to learn about kinds. *Journal of Experimental Psychology: General*, *143*(1), 46.
- Cimpian, A., & Petro, G. (2014). Building theory-based concepts: Four-year-olds preferentially seek explanations for features of kinds. *Cognition*, *131*(2), 300–310.
- Cooper, M. J., Dimitrov, O., & Rau, P. R. (2001). A rose.com by any other name. *Journal of Finance*, *56*(6), 2371–2388.
- Drake, M. S., Jennings, J., Roulstone, D. T., & Thornock, J. R. (2017). The comovement of investor attention. *Management Science*, *63*(9), 2847–2867.
- Ehrmann, M., & Jansen, D.-J. (2020). Stock return comovement when investors are distracted: More, and more homogeneous. *mimeo*.
- Gabaix, X. (2019). Behavioral inattention. In *Handbook of behavioral economics: Applications and foundations 1* (Vol. 2, pp. 261–343). Elsevier.
- Gibbons, B., Iliev, P., & Kalodimos, J. (2021). Analyst information acquisition via EDGAR. *Management Science*, *67*(2), 769–793.
- Gilbert, T., Kogan, S., Lochstoer, L., & Ozyildirim, A. (2012). Investor inattention and the market impact of summary statistics. *Management Science*, *58*(2), 336–350.
- Gottlieb, J. (2018). Understanding active sampling strategies: Empirical approaches and implications for attention and decision research. *Cortex*, *102*, 150–160.
- Heinke, S. (2019). Active and passive information acquisition: Application to financial markets. *Available at SSRN 3391221*.
- Huang, S., Huang, Y., & Lin, T.-C. (2019). Attention allocation and return co-movement: Evidence from repeated natural experiments. *Journal of Financial Economics*, *132*(2), 369–383.
- Kacperczyk, M., Van Nieuwerburgh, S., & Veldkamp, L. (2016). A rational theory of mutual funds' attention allocation. *Econometrica*, *84*(2), 571–626.
- Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ:

Prentice-Hall.

- Krajbich, I. (2019). Accounting for attention in sequential sampling models of decision making. *Current Opinion in Psychology*, *29*, 6–11.
- Kruschke, J. K. (2011). Models of attentional learning. In E. M. Pothos & A. J. Wills (Eds.), *Formal Approaches in Categorization* (Vol. 120, pp. 120–152). Cambridge University Press.
- Liu, H., Peng, L., & Tang, Y. (in press). Retail attention, institutional attention. *Journal of Financial and Quantitative Analysis*.
- Maćkowiak, B., Matějka, F., & Wiederholt, M. (in press). Rational inattention: A disciplined behavioral model. *Journal of Economic Literature*.
- Maćkowiak, B., & Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *American Economic Review*, *99*(3), 769–803.
- Maćkowiak, B., & Wiederholt, M. (2015). Business cycle dynamics under rational inattention. *Review of Economic Studies*, *82*(4), 1502–1532.
- Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, *129*(4), 592.
- Mondria, J. (2010). Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory*, *145*(5), 1837–1864.
- Nobre, A. C., & Coull, J. T. (2010). *Attention and time*. Oxford University Press, USA.
- Pashler, H., & Johnston, J. C. (1998). Attentional limitations in dual-task performance. In H. Pashler (Ed.), *Attention* (pp. 155–189). Psychology Press/Erlbaum (UK) Taylor & Francis.
- Pasten, E., & Schoenle, R. (2016). Rational inattention, multi-product firms and the neutrality of money. *Journal of Monetary Economics*, *80*, 1–16.
- Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, *80*(3), 563–602.
- Peng, L., Xiong, W., & Bollerslev, T. (2007). Investor attention and time-varying

- comovements. *European Financial Management*, 13(3), 394–422.
- Rehder, B., Colner, R. M., & Hoffman, A. B. (2009). Feature inference learning and eyetracking. *Journal of Memory and Language*, 60(3), 393–419.
- Schulte-Mecklenbeck, M., Johnson, J. G., Böckenholt, U., Goldstein, D. G., Russo, J. E., Sullivan, N. J., & Willemsen, M. C. (2017). Process-tracing methods in decision making: On growing up in the 70s. *Current Directions in Psychological Science*, 26(5), 442–450.
- Simon, H. A. (1971). Designing organizations for an information-rich world. *Computers, communications, and the public interest*, 72, 37.
- Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3), 665–690.
- van Doorn, J., Ly, A., Marsman, M., & Wagenmakers, E.-J. (2020). Bayesian rank-based hypothesis testing for the rank sum test, the signed rank test, and Spearman's  $\rho$ . *Journal of Applied Statistics*, 47(16), 2984–3006.
- Willemsen, M. C., & Johnson, E. J. (2011). Visiting the decision factory: Observing cognition with MouselabWEB and other information acquisition methods. In M. Schulte-Mecklenbeck, A. Kuehberger, & J. G. Johnson (Eds.), *A handbook of process tracing methods for decision research* (pp. 21–42). Psychology Press New York, NY, US.

## Appendix A

### Theoretical Model

We present a fully Bayesian variant of the attention allocation problem in Peng and Xiong (2006), streamlined for the purposes of this study. In the current setup, the agent is tasked with accurately estimating the values of several random variables,  $v_i, i = 1, \dots, n$ , aiming to minimize the expected quadratic loss of his predictions. The variables are generated according to a hierarchical structure, in which they are equal to the sum of two components: a shared category factor  $c$ , and an idiosyncratic member-specific factor  $m_i$ , so  $v_i = c + m_i$ . Both components are drawn from normal distributions,  $c \sim \mathcal{N}(\mu_c, 1/\tau_c)$  and  $m_i \sim \mathcal{N}(\mu_m, 1/\tau_m)$ .

To help form accurate predictions, the agent gains information by collecting noisy signals about these factors, which requires splitting his limited attention across them. In each moment of time, he draws a noisy signal from any one of the  $n + 1$  factors he chooses. These signals are generated from a normal distribution with mean equal to the true factor value ( $c$  or  $m_i$ ) and precision  $\tau_\varepsilon$ . We denote the  $j$ th signal from a given factor as  $x_c^{j_0}$  or  $x_{m_i}^{j_i}$ , and all signals collectively as  $x$ . The agent has a finite attention budget,  $\kappa$ , denoting the total number of signals available (in this case, the amount of time available for signal collection multiplied by the frequency at which signals are drawn). He must therefore choose which fraction  $\lambda$  of time to spend attending to the shared factor  $c$ . By symmetry, the remainder is spread evenly across the idiosyncratic factors  $m_i$ , so each receives fraction  $\frac{1-\lambda}{n}$  of the attention. We ignore the discreteness of signal timing for simplicity.

The posterior distributions of the factors are

$$c|x \sim \mathcal{N}\left(\frac{\tau_c \mu_c + \tau_\varepsilon \sum_{j_0} x_c^{j_0}}{\tau_c + \tau_\varepsilon \lambda \kappa}, (\tau_c + \tau_\varepsilon \lambda \kappa)^{-1}\right) \quad (1)$$

$$m_i|x \sim \mathcal{N}\left(\frac{\tau_m \mu_m + \tau_\varepsilon \sum_{j_i} x_{m_i}^{j_i}}{\tau_m + \tau_\varepsilon \left(\frac{1-\lambda}{n}\right) \kappa}, \left(\tau_m + \tau_\varepsilon \left(\frac{1-\lambda}{n}\right) \kappa\right)^{-1}\right). \quad (2)$$

Due to independence of the components, the objective is equivalent to minimizing the sum of their expected posterior variances (with the category factor being weighted more heavily than each member-specific factor). Assuming an interior solution in what follows,

$$\arg \min_{\lambda} V = \left[ (\tau_c + \tau_{\varepsilon} \lambda \kappa)^{-1} + \left( \tau_m + \tau_{\varepsilon} \left( \frac{1-\lambda}{n} \right) \kappa \right)^{-1} \right] \quad (3)$$

$$\frac{\partial V}{\partial \lambda} = - \frac{\tau_{\varepsilon} \kappa}{(\tau_c + \tau_{\varepsilon} \lambda \kappa)^2} + \frac{\frac{1}{n} \tau_{\varepsilon} \kappa}{\left( \tau_m + \tau_{\varepsilon} \left( \frac{1-\lambda}{n} \right) \kappa \right)^2} = 0 \quad (4)$$

$$\lambda^* = \left( \frac{1}{1 + \sqrt{n}} \right) \left( 1 + \frac{n \tau_m - \sqrt{n} \tau_c}{\tau_{\varepsilon} \kappa} \right) \quad (5)$$

We define the category focus,  $\bar{\lambda}$ , as the fraction of time spent on the category factor relative to the average fraction of time spent on the member-specific factors:

$$\bar{\lambda} = \lambda^* - \left( \frac{1 - \lambda^*}{n} \right) = \lambda^* \left( \frac{1 + n}{n} \right) - \frac{1}{n} \quad (6)$$

We evaluate how the category focus changes based on various parameters:

$$\frac{\partial \bar{\lambda}}{\partial \tau_c} = - \left( \frac{n + 1}{n + \sqrt{n}} \right) \left( \frac{1}{\tau_{\varepsilon} \kappa} \right) < 0 \quad (7)$$

$$\frac{\partial \bar{\lambda}}{\partial \tau_m} = \left( \frac{n + 1}{\sqrt{n} + 1} \right) \left( \frac{1}{\tau_{\varepsilon} \kappa} \right) > 0 \quad (8)$$

$$\frac{\partial \bar{\lambda}}{\partial \kappa} = \frac{\partial \bar{\lambda}}{\partial \tau_{\varepsilon}} = \left( \frac{n + 1}{n + \sqrt{n}} \right) \left( \frac{\tau_c - \sqrt{n} \tau_m}{\tau_{\varepsilon} \kappa} \right) < 0 \text{ when } \sqrt{n} \sigma_c^2 > \sigma_m^2 \quad (9)$$

These expressions respectively mean that category focus is decreasing in prior category precision and increasing in prior member-specific precision (i.e., increasing in prior category variance and decreasing in prior member-specific variance), and decreasing in attention capacity and signal precision when prior category variance scaled based on the number of members is greater than prior member-specific variance.

**Prediction dispersion.** To calculate the theoretical variance of the value predictions, note first that the category component of the predictions should remain the

same across all members and can thus be ignored. Denote the posterior mean of  $m_i$  conditional on  $x$  as  $\hat{m}_i = \frac{\tau_m \mu_m + \tau_\varepsilon \sum_{j_i} x_{m_i}^{j_i}}{\tau_m + \tau_\varepsilon \left(\frac{1-\lambda}{n}\right) \kappa}$ , and observe further that only the term  $\frac{\tau_\varepsilon \sum_{j_i} x_{m_i}^{j_i}}{\tau_m + \tau_\varepsilon \left(\frac{1-\lambda}{n}\right) \kappa}$  varies with the signals drawn (and moreover,  $\mu_m = 0$  in the present task). Suppose  $J = \left(\frac{1-\lambda}{n}\right) \kappa$  signals are received for each member (again ignoring signal discreteness for simplicity), and let  $a = \frac{J\tau_\varepsilon}{\tau_m + J\tau_\varepsilon}$  and  $\bar{x}_{m_i} = \frac{1}{J} \sum_{j_i} x_{m_i}^{j_i}$ , so  $Var(\hat{m}_i) = Var(a\bar{x}_{m_i}) = a^2 Var(\bar{x}_{m_i})$ . By the law of total variance,

$$Var(\bar{x}_{m_i}) = E[Var(\bar{x}_{m_i}|m_i)] + Var(E[\bar{x}_{m_i}|m_i]) \quad (10)$$

$$= E\left[\frac{\sigma_\varepsilon^2}{J}\right] + Var(m_i) \quad (11)$$

$$= \frac{\sigma_\varepsilon^2}{J} + \sigma_m^2 \quad (12)$$

$$a^2 Var(\bar{x}_{m_i}) = \left(\frac{J\tau_\varepsilon}{\tau_m + J\tau_\varepsilon}\right)^2 \left(\frac{\sigma_\varepsilon^2}{J} + \sigma_m^2\right) \quad (13)$$

$$= \frac{(J\tau_\varepsilon)^2}{(\tau_m + J\tau_\varepsilon)^2} \left(\frac{1}{J\tau_\varepsilon} + \frac{1}{\tau_m}\right) \quad (14)$$

$$= \frac{(J\tau_\varepsilon)^2}{(\tau_m + J\tau_\varepsilon)^2} \left(\frac{\tau_m + J\tau_\varepsilon}{\tau_m J\tau_\varepsilon}\right) \quad (15)$$

$$= \frac{J\tau_\varepsilon}{\tau_m(\tau_m + J\tau_\varepsilon)}. \quad (16)$$

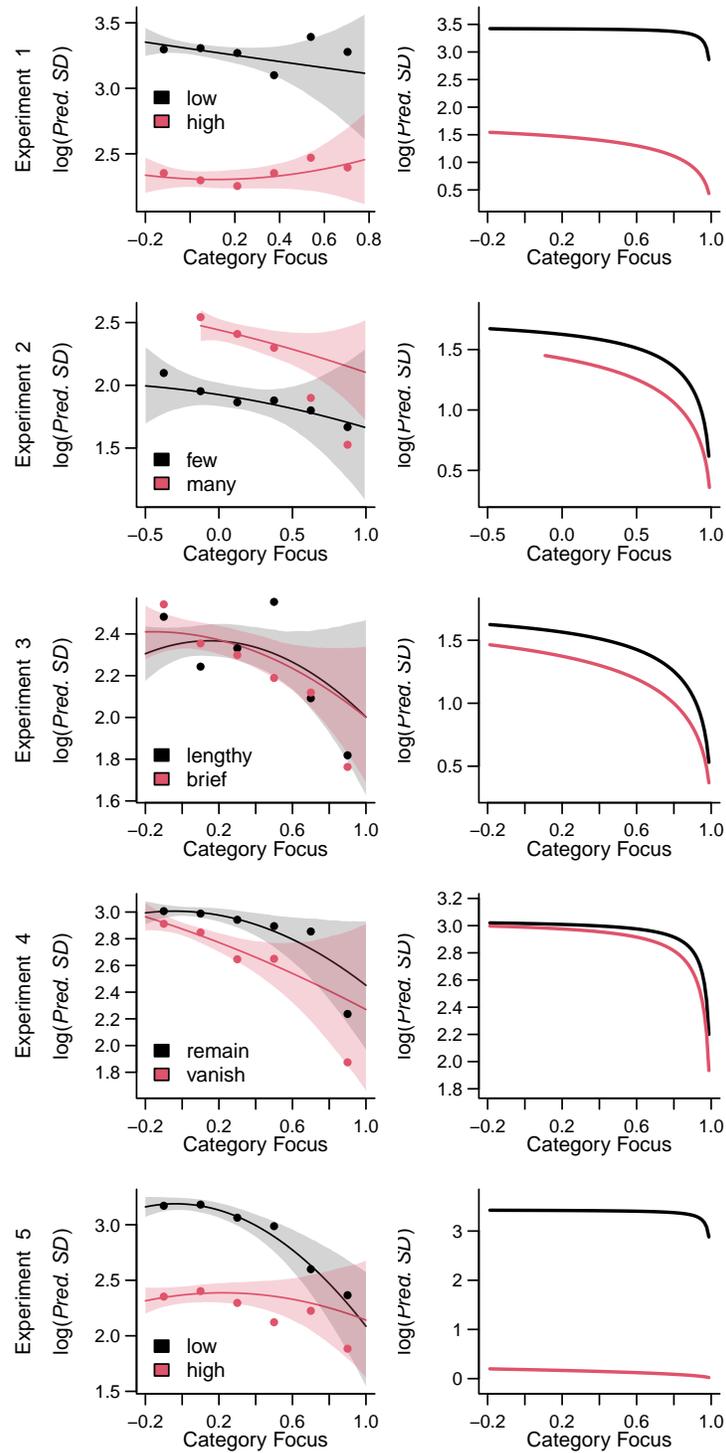
Observe that as the number of member-specific signals  $J$  grows large (small  $\lambda$ ), the variance of the posterior means will converge to the prior variance of the member-specific factor,  $\sigma_m^2$ , because that is the level of irreducible uncertainty. Further, when no member-specific signals are acquired ( $J = 0$ , large  $\lambda$ ), the variance of the posterior means equals zero, because there is nothing to distinguish each member and so the exact same prediction will be made for all of them. As category focus increases (and hence  $J$  decreases), the prediction variance will decrease from  $\sigma_m^2$  to 0.

**Prediction accuracy.** The theoretical error of the value predictions can be obtained by observing that the posterior variance is the Bayes estimator of the quadratic

loss (Roberts 2007). Thus, the error can be simply given as the posterior variance

$$(\tau_c + \tau_\varepsilon \lambda \kappa)^{-1} + \left( \tau_m + \tau_\varepsilon \left( \frac{1-\lambda}{n} \right) \kappa \right)^{-1}.$$

**Response noise.** There is also variation that may be caused by response noise due to the experiment’s slider interface. Assuming it is normally distributed for tractability, it can be included in the calculations for prediction error and variance by simply adding in its variance. In the main text, when plotting the theoretical predictions, we roughly calibrate this level of response noise to the data. Specifically, we set the standard deviation of response noise to be the 10th percentile of prediction standard deviations across all (included) periods in each experiment. This provides a balance between finding the minimal level of variation (corresponding to uncontrollable noise) while avoiding undue influence from outliers (either participants or periods with unusually low noise). In Figure S1, we show the same plot as in the main text connecting prediction dispersion and category focus, but without any response noise. This plot reveals that the main effect of response noise is to produce a floor on the prediction dispersion, which can attenuate the link to category focus when prediction dispersion is already low. For example, this attenuation is visible in the high category variance condition of Experiment 1.



Supplemental Material, Figure S1. Prediction Dispersion and Category Focus Without Response Noise. Relationship between category focus and dispersion in stock value predictions in the same period. (Left) Data, binned averages with quadratic regression lines and 95% credible intervals from Bayesian random effects models. (Right) Theory, without response noise.

## Appendix B

### Comprehension Check

1. If the industry factor is  $-8$ , and the stock-specific factor for A is 30, what is the total value of stock A?

- 30
- 38
- $-8$
- 22 (correct)

2. If the industry factor is 10, the stock-specific factor for A is  $-15$ , and the stock-specific factor for B is 5, what is the total value of stock A?

- 10
- 15
- $-5$  (correct)
- $-15$

## Appendix C

### Model Fitting

We fit an extended version of the theoretical model to capture how individual participants integrate signals to form their predictions. Suppose that  $J_0$  signals are obtained for the category factor and  $J_i$  signals are obtained for a given stock-specific factor. Using the notation from Appendix A, we suppose that the prediction of Stock  $i$ 's value is constructed as  $\hat{c} + \hat{m}_i$  where

$$\hat{c} = \frac{\tau_c \mu_c + \sum_{j_0} \tau_{\varepsilon,c}^{j_0} x_c^{j_0}}{\tau_c + \sum_{j_0} \tau_{\varepsilon,c}^{j_0}} \quad (17)$$

$$\hat{m}_i = \frac{\tau_m \mu_m + \sum_{j_i} \tau_{\varepsilon,m}^{j_i} x_{m_i}^{j_i}}{\tau_m + \sum_{j_i} \tau_{\varepsilon,m}^{j_i}} \quad (18)$$

are the posterior means of the category and stock-specific factors conditional on the acquired signals, with two modifications. First, we allow the subjective weights placed on signals from the category factor ( $\tau_{\varepsilon,c}$ ) and the stock-specific factors ( $\tau_{\varepsilon,m}$ ) to differ. Second, we allow these weights to differ depending on when the signal was received ( $\tau_{\varepsilon,c}^{j_0}$  and  $\tau_{\varepsilon,m}^{j_i}$ ). Specifically we add another parameter ( $\psi$ ) which allows a different weight to be placed on the last signal seen for each factor, to reflect its increased salience as it was highlighted and displayed numerically in Experiments 1–3:

$$\tau_{\varepsilon,c}^{j_0} = \begin{cases} \tau_{\varepsilon,c} & \text{when } j_0 < J_0 \\ \psi \tau_{\varepsilon,c} & \text{when } j_0 = J_0 \end{cases} \quad (19)$$

$$\tau_{\varepsilon,m}^{j_i} = \begin{cases} \tau_{\varepsilon,m} & \text{when } j_i < J_i \\ \psi \tau_{\varepsilon,m} & \text{when } j_i = J_i \end{cases} \quad (20)$$

The special case of  $\psi = 1$  represents no effect of salience.

Finally, to translate this point estimate into a response distribution that includes

noise due to extraneous factors such as the slider response paradigm, we assume the prediction is input with Gaussian response noise and hence is given by  $\mathcal{N}(\hat{c} + \hat{m}_i, \sigma_{response}^2)$ .

Accordingly, we fit 4 free parameters to the data for each individual: the weight placed on category-level signals ( $\tau_{\varepsilon,c}$ ), the weight placed on stock-specific signals ( $\tau_{\varepsilon,m}$ ), the last signal salience weight ( $\psi$ ), and the response noise ( $\sigma_{response}$ ). The remaining variables ( $\mu_c, \mu_m, \tau_c, \tau_m, x_c^{j_0}, x_{m_i}^{j_i}, J_0$ , and  $J_i$ ) are determined from the data or design parameters.

We fit the model using Bayesian parameter estimation, with the following weakly informative priors, where  $\tau_{\varepsilon,c}$  and  $\tau_{\varepsilon,m}$  are scaled up by 100:  $\tau_{\varepsilon,c} \sim Cauchy_{x>0}(1, 10)$ ,  $\tau_{\varepsilon,m} \sim Cauchy_{x>0}(1, 10)$ ,  $\psi \sim Cauchy_{x>0}(1, 10)$ , and  $\sigma_{response} \sim Cauchy_{x>0}(0, 25)$ , where  $Cauchy_{x>0}(x_0, \gamma)$  denotes the Cauchy distribution truncated below at zero with location parameter  $x_0$  and scale parameter  $\gamma$ . The summary statistics of the fitted parameters are in Table S1. See van Doorn, Ly, Marsman, and Wagenmakers (2020) about Bayesian signed rank and rank sum tests. Regressions of prediction dispersion on the parameters are in Tables S2 and S3.

	<i>Exp. 1</i>	<i>Exp. 2</i>	<i>Exp. 3</i>	<i>Exp. 4</i>	<i>Exp. 5</i>
Parameter	high low	many few	brief lengthy	vanish remain	high low
$\tau_{\varepsilon,c}$	0.353 [0.490] 1.492 [0.892]	0.391 [0.572] 0.478 [0.578]	0.443 [0.577] 0.389 [0.506]	0.263 [0.643] 0.446 [0.531]	0.202 [0.372] 1.122 [0.350]
$P(\text{diff.} > 0)$	(<.001)	(.103)	(.929)	(<.001)	(<.001)
$\tau_{\varepsilon,m}$	2.290 [1.267] 0.802 [0.825]	2.062 [0.985] 1.846 [1.143]	2.335 [0.992] 1.986 [0.715]	0.309 [0.565] 1.207 [1.036]	1.254 [1.104] 0.139 [0.323]
$P(\text{diff.} > 0)$	(>.999)	(.990)	(.999)	(<.001)	(>.999)
$\psi$	2.997 [1.777] 3.069 [1.528]	2.866 [1.756] 2.261 [1.800]	3.090 [1.421] 2.983 [1.536]	1.112 [0.917] 1.247 [0.803]	1.079 [1.183] 0.788 [0.607]
$P(\text{diff.} > 0)$	(.332)	(.996)	(.695)	(.088)	(.927)
$\sigma_{response}$	9.225 [4.317] 8.293 [4.265]	10.790 [6.719] 10.176 [5.895]	9.798 [5.096] 10.445 [5.074]	15.977 [3.586] 9.989 [4.000]	12.778 [3.981] 17.631 [4.511]
$P(\text{diff.} > 0)$	(.988)	(.792)	(.042)	(>.999)	(<.001)

Supplemental Material, Table S1

*Parameter Estimates. Means shown with standard deviations in brackets. The posterior probability that the difference across conditions is positive is shown in parentheses, according to a Bayesian signed rank test (Exps. 1–4) or a Bayesian rank sum test (Exp. 5).*

Variable	log(1 + Prediction SD)	
<i>Exp. 1</i>	Higher category variance	Lower category variance
Intercept	1.502 [1.295, 1.706] (>.999)	3.051 [2.879, 3.223] (>.999)
$\tau_{\varepsilon,c}$	0.085 [-0.014, 0.186] (.952)	0.087 [0.041, 0.135] (>.999)
$\tau_{\varepsilon,m}$	0.065 [0.026, 0.103] (>.999)	0.071 [0.021, 0.120] (.998)
$\psi$	0.040 [0.011, 0.067] (.998)	0.001 [-0.027, 0.028] (.517)
$\sigma_{response}$	0.057 [0.044, 0.070] (>.999)	0.006 [-0.006, 0.017] (.849)
Category focus	0.022 [-1.141, 1.194] (.515)	-1.372 [-2.914, 0.111] (.035)
Category focus <sup>2</sup>	0.112 [-1.157, 1.501] (.564)	0.541 [-0.858, 1.893] (.782)
<i>Exp. 2</i>	Larger category size	Smaller category size
Intercept	1.601 [1.327, 1.883] (>.999)	1.087 [0.832, 1.342] (>.999)
$\tau_{\varepsilon,c}$	-0.136 [-0.256, -0.019] (.014)	0.004 [-0.113, 0.126] (.517)
$\tau_{\varepsilon,m}$	0.116 [0.040, 0.191] (.999)	0.121 [0.055, 0.185] (>.999)
$\psi$	0.057 [0.012, 0.102] (.994)	0.039 [-0.002, 0.079] (.969)
$\sigma_{response}$	0.041 [0.029, 0.053] (>.999)	0.051 [0.037, 0.065] (>.999)
Category focus	-1.473 [-3.541, 0.756] (.090)	-1.333 [-3.381, 0.757] (.098)
Category focus <sup>2</sup>	-0.003 [-1.496, 1.525] (.496)	-0.912 [-3.256, 1.558] (.222)
<i>Exp. 3</i>	Shorter time limit	Longer time limit
Intercept	1.574 [1.327, 1.829] (>.999)	1.462 [1.208, 1.720] (>.999)
$\tau_{\varepsilon,c}$	-0.050 [-0.143, 0.039] (.141)	0.032 [-0.073, 0.133] (.720)
$\tau_{\varepsilon,m}$	0.077 [0.019, 0.136] (.995)	0.091 [0.006, 0.173] (.983)
$\psi$	0.070 [0.031, 0.110] (.999)	0.051 [0.015, 0.088] (.997)
$\sigma_{response}$	0.042 [0.031, 0.054] (>.999)	0.051 [0.040, 0.063] (>.999)
Category focus	-1.458 [-2.867, -0.036] (.022)	-1.206 [-2.645, 0.393] (.070)
Category focus <sup>2</sup>	-0.763 [-2.088, 0.563] (.126)	-1.579 [-3.006, -0.023] (.024)
<i>Exp. 4</i>	Data points vanish	Data points remain
Intercept	2.331 [2.129, 2.538] (>.999)	2.803 [2.665, 2.942] (>.999)
$\tau_{\varepsilon,c}$	-0.075 [-0.139, -0.012] (.011)	0.016 [-0.045, 0.076] (.693)
$\tau_{\varepsilon,m}$	0.088 [0.013, 0.164] (.990)	0.074 [0.038, 0.109] (>.999)
$\psi$	-0.032 [-0.074, 0.012] (.073)	-0.015 [-0.054, 0.023] (.228)
$\sigma_{response}$	0.034 [0.022, 0.046] (>.999)	0.011 [0.002, 0.020] (.990)
Category focus	-2.535 [-3.931, -1.171] (<.001)	-1.251 [-2.382, -0.104] (.017)
Category focus <sup>2</sup>	-0.540 [-1.951, 1.009] (.227)	-1.200 [-2.367, -0.047] (.020)

Supplemental Material, Table S2

*Regression Results, Exps. 1–4: Prediction Dispersion and Model Parameter Estimates. Posterior mean estimates from Bayesian random effects models with prediction dispersion regressed on fitted model parameters. 95% credible intervals in brackets and  $P(\beta > 0)$  in parentheses.*

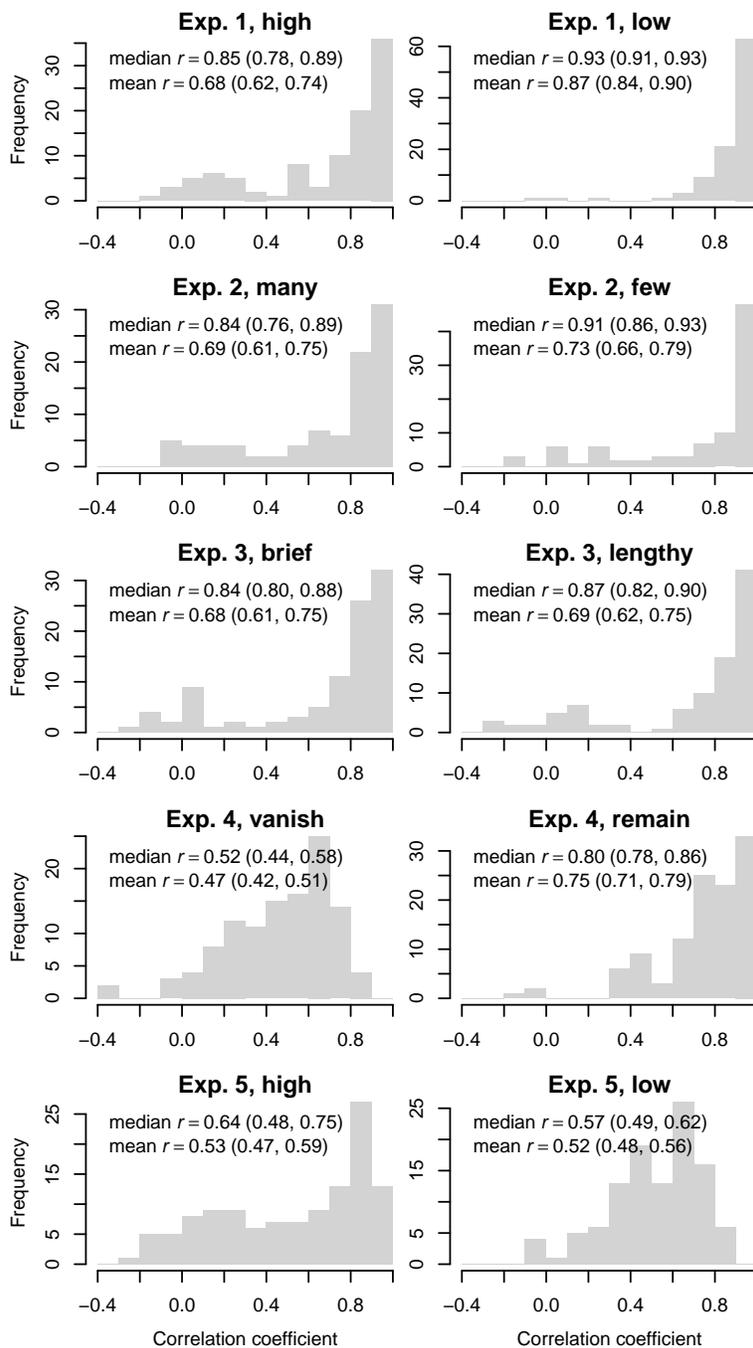
Variable	log(1 + Prediction SD)	
<i>Exp. 5</i>	Higher category variance	Lower category variance
Intercept	1.366 [1.143, 1.596] (>.999)	2.568 [2.365, 2.784] (>.999)
$\tau_{\varepsilon,c}$	-0.112 [-0.259, 0.036] (.069)	-0.069 [-0.202, 0.057] (.153)
$\tau_{\varepsilon,m}$	0.018 [-0.039, 0.071] (.745)	0.182 [0.040, 0.322] (.992)
$\psi$	0.010 [-0.034, 0.057] (.671)	-0.010 [-0.088, 0.067] (.406)
$\sigma_{response}$	0.077 [0.063, 0.091] (>.999)	0.037 [0.026, 0.047] (>.999)
Category focus	-0.109 [-1.733, 1.557] (.434)	-2.149 [-3.310, -0.966] (<.001)
Category focus <sup>2</sup>	-0.319 [-2.045, 1.527] (.361)	-2.029 [-3.293, -0.829] (<.001)

Supplemental Material, Table S3

*Regression Results, Exp. 5: Prediction Dispersion and Model Parameter Estimates. Posterior mean estimates from Bayesian random effects models with prediction dispersion regressed on fitted model parameters. 95% credible intervals in brackets and  $P(\beta > 0)$  in parentheses.*

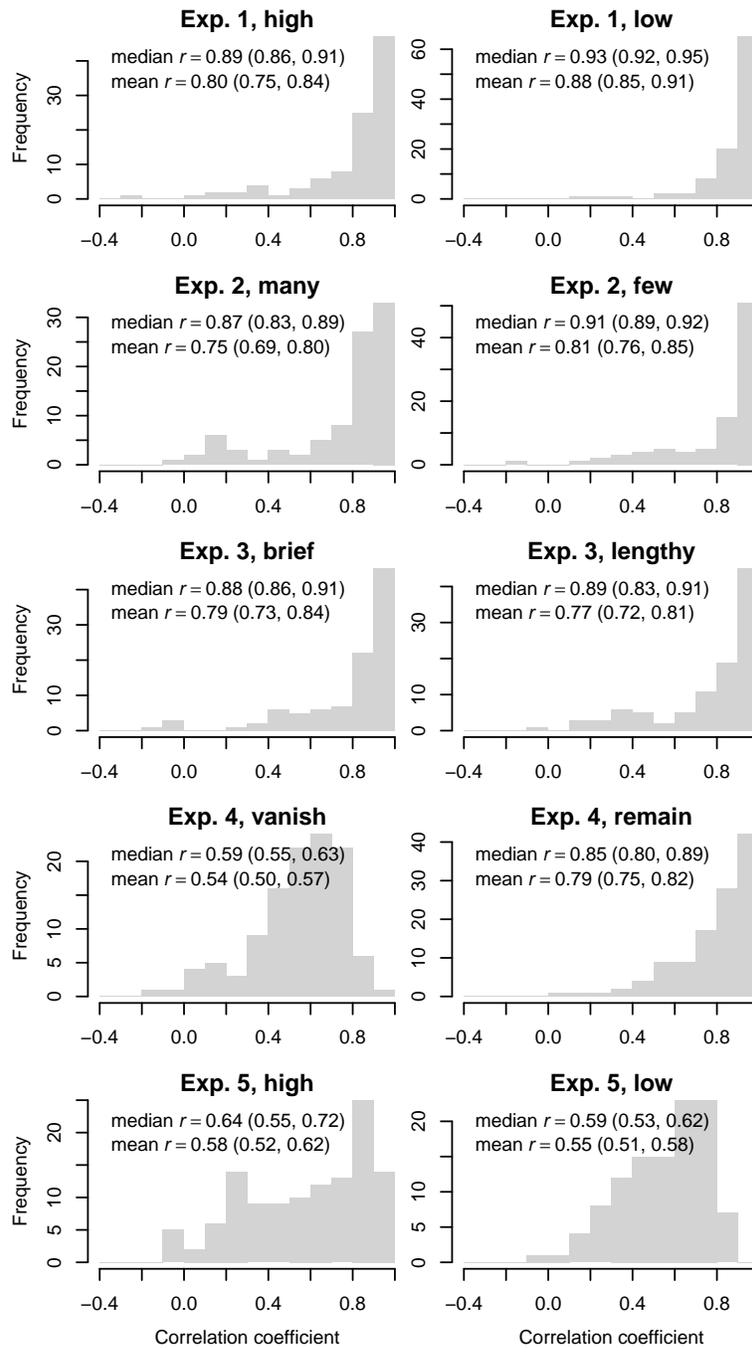
## Appendix D

### Correlations Between Participant Predictions and True Values



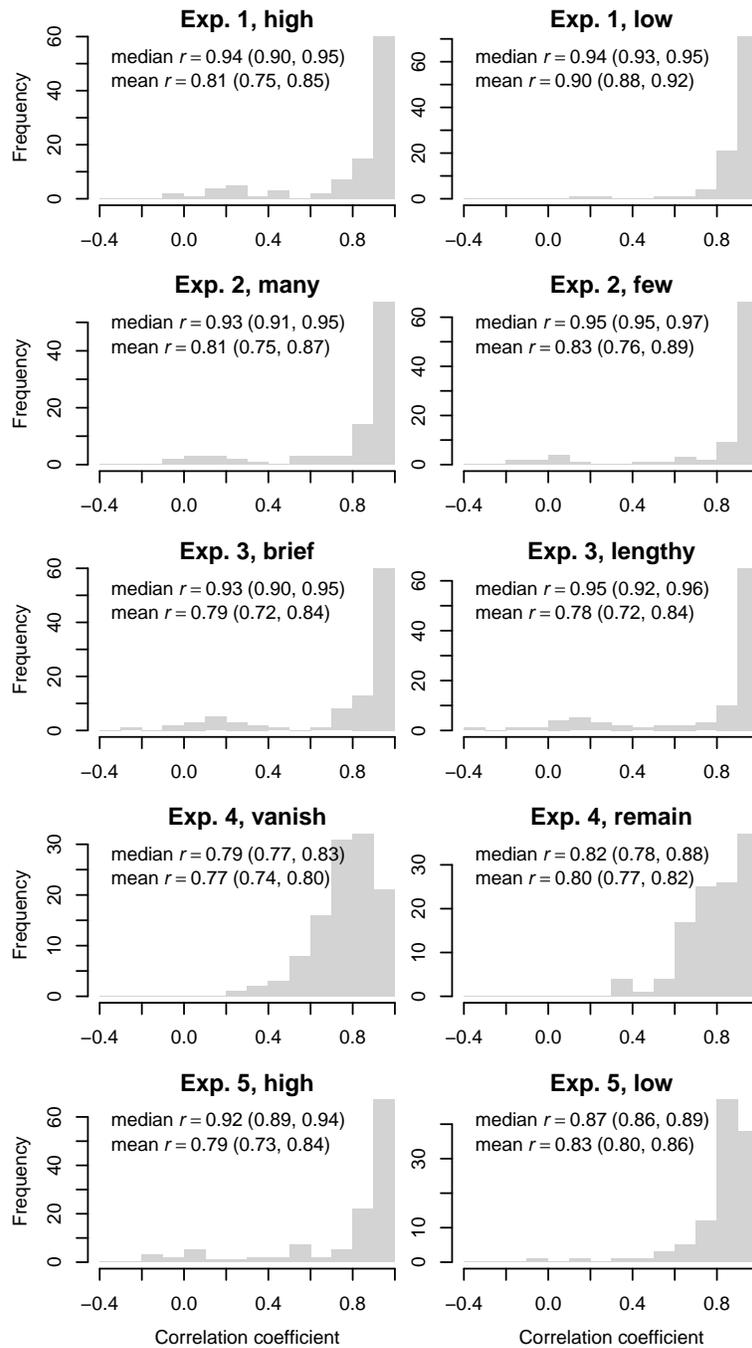
*Supplemental Material, Figure S2. Correlations Between Participant Predictions and True Values.* Distribution of correlation coefficients for all participants. Mean and median values reported with bootstrapped 95% confidence intervals in parentheses.

## Correlations Between Participant Predictions and Fitted Model Predictions



Supplemental Material, Figure S3. Correlations Between Participant Predictions and Fitted Model Predictions. Distribution of correlation coefficients for all participants. Mean and median values reported with bootstrapped 95% confidence intervals in parentheses.

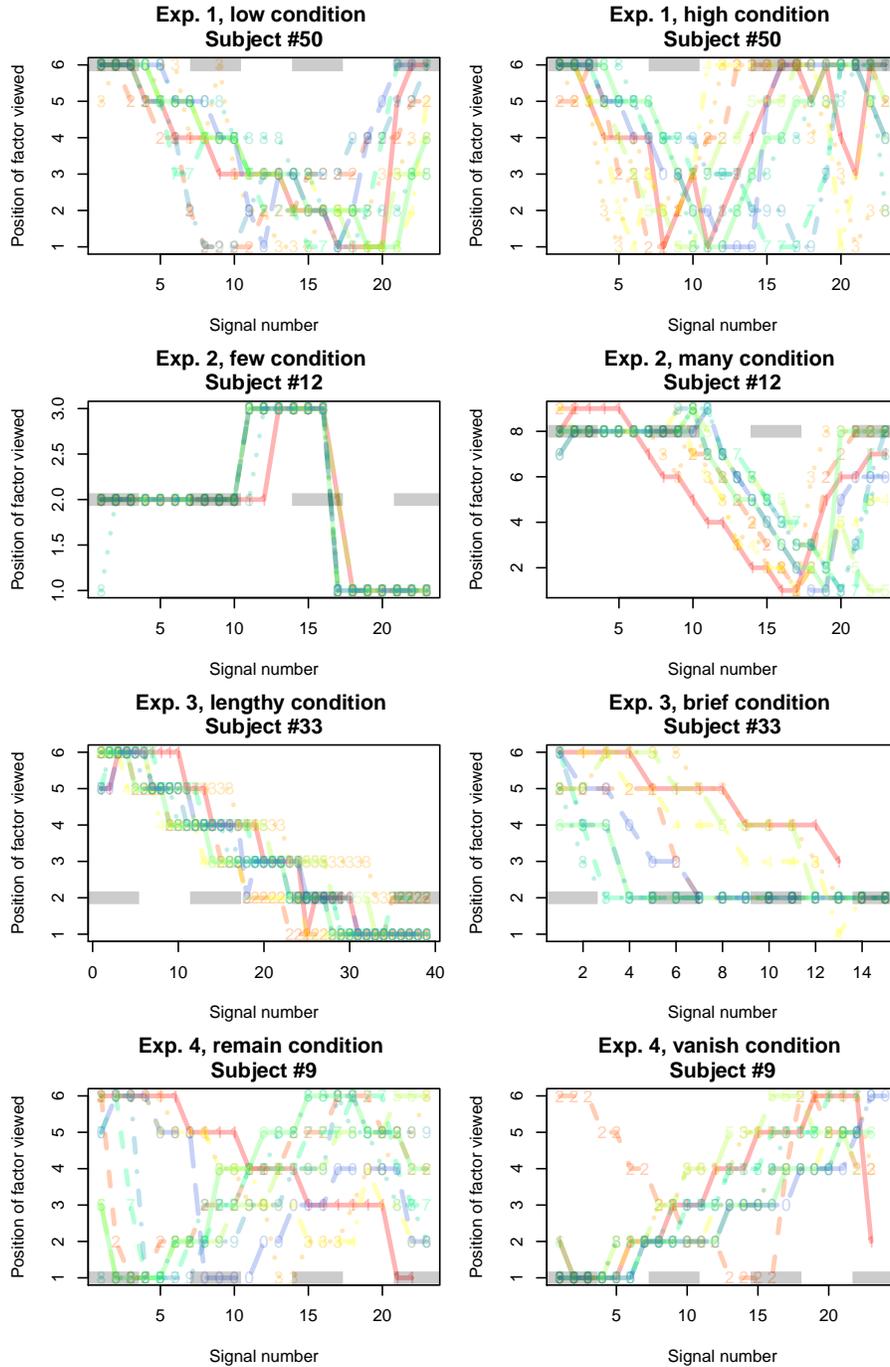
## Correlations Between Fitted Model Predictions and True Values



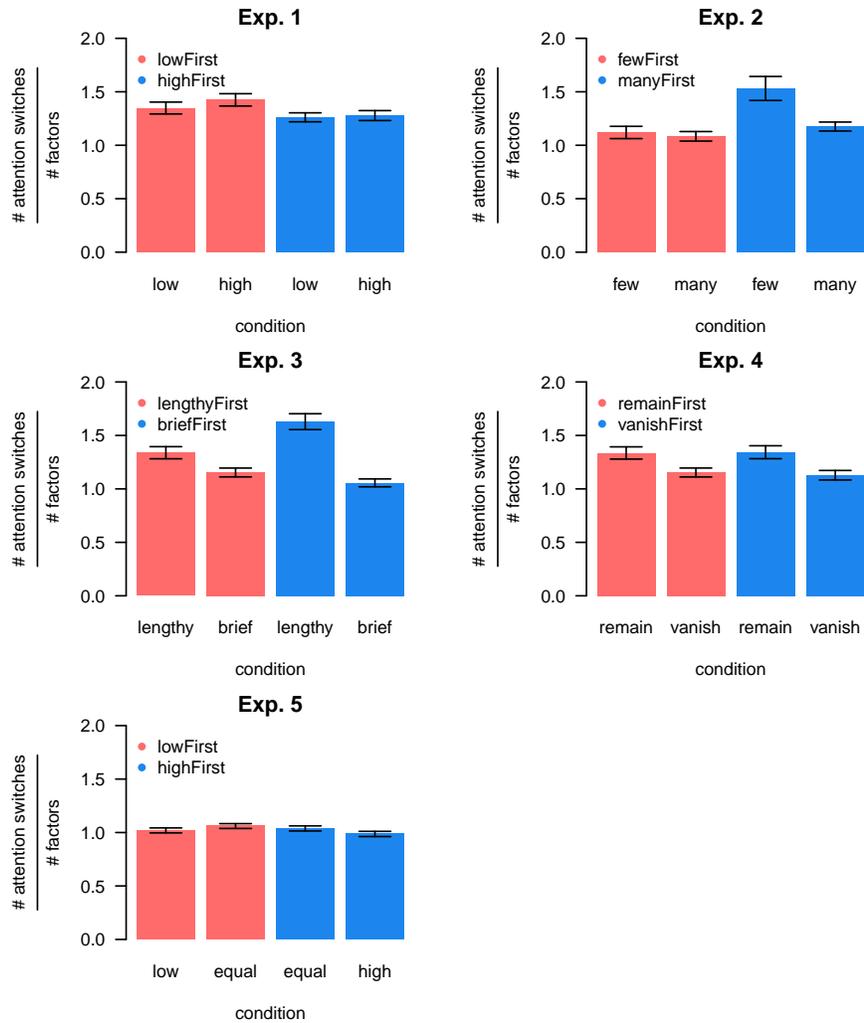
Supplemental Material, Figure S4. Correlations Between Fitted Model Predictions and True Values. Distribution of correlation coefficients for all participants. Mean and median values reported with bootstrapped 95% confidence intervals in parentheses.

# Appendix E

## Within-Period Attention Trajectories



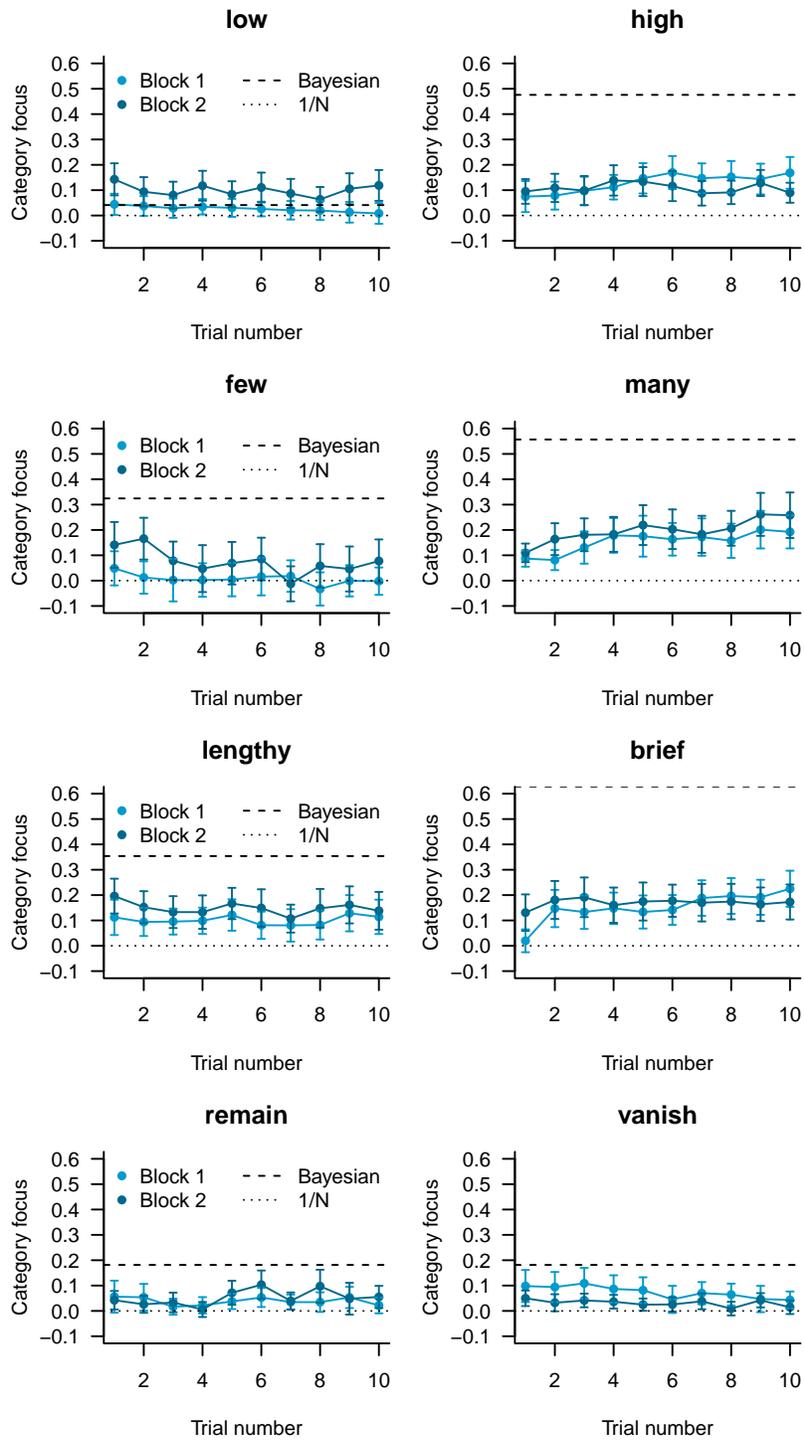
Supplemental Material, Figure S5. Examples of Attention Trajectories. Number denotes period number. Dashed gray line indicates the position of the industry for the participant.



*Supplemental Material, Figure S6. Number of Attention Switches.* Number of times that attention switches from one factor to another, relative to the total number of factors. Means with 95% confidence intervals.

Attention trajectories tend to follow a waterfall or cascade type pattern, going from top to bottom, as is visible in the example trajectories shown in Figure S5. There is sometimes also a return of attention to the industry factor at the end after some signals have been acquired from all of the factors. The number of attention switches relative to the number of factors is shown in Figure S6. A pure cascade pattern would yield a ratio of 1. The ratio is especially elevated when going from many stocks to few stocks (in Experiment 2) or from a short time limit to a long time limit (in Experiment 3). In both of these situations, the participant is moving from a more stringent environment to a more relaxed one, and so may be more lax with their attention trajectory.

## Across-Period Dynamics



Supplemental Material, Figure S7. Category Focus Across Periods and Blocks. Color indicates which block was encountered first. Error bars depict 95% confidence intervals.

Coefficient	Category focus	
	Higher category variance	Lower category variance
<i>Exp. 1</i>		
Intercept	0.084 [0.038, 0.129] (>.999)	0.041 [0.000, 0.084] (.975)
Period	0.010 [0.003, 0.017] (.998)	-0.004 [-0.009, 0.002] (.083)
Block	0.020 [-0.042, 0.085] (.740)	0.080 [0.020, 0.140] (.996)
Period $\times$ Block	-0.009 [-0.019, 0.001] (.040)	0.002 [-0.006, 0.010] (.698)
<i>Exp. 2</i>		
	Larger category size	Smaller category size
Intercept	0.099 [0.056, 0.143] (>.999)	0.025 [-0.036, 0.088] (.782)
Period	0.011 [0.003, 0.020] (.996)	-0.005 [-0.013, 0.004] (.126)
Block	0.036 [-0.027, 0.098] (.879)	0.098 [0.011, 0.188] (.985)
Period $\times$ Block	0.001 [-0.011, 0.013] (.569)	-0.007 [-0.019, 0.004] (.113)
<i>Exp. 3</i>		
	Shorter time limit	Longer time limit
Intercept	0.086 [0.027, 0.145] (.998)	0.102 [0.048, 0.154] (>.999)
Period	0.014 [0.008, 0.021] (>.999)	0.001 [-0.006, 0.008] (.594)
Block	0.075 [-0.010, 0.155] (.960)	0.066 [-0.009, 0.142] (.960)
Period $\times$ Block	-0.013 [-0.023, -0.004] (.003)	-0.004 [-0.014, 0.006] (.203)
<i>Exp. 4</i>		
	Data points vanish	Data points remain
Intercept	0.105 [0.071, 0.141] (>.999)	0.041 [0.008, 0.076] (.992)
Period	-0.007 [-0.011, -0.003] (<.001)	-0.001 [-0.007, 0.005] (.383)
Block	-0.065 [-0.113, -0.020] (.003)	-0.005 [-0.058, 0.046] (.424)
Period $\times$ Block	0.005 [0.000, 0.010] (.970)	0.005 [-0.005, 0.015] (.858)

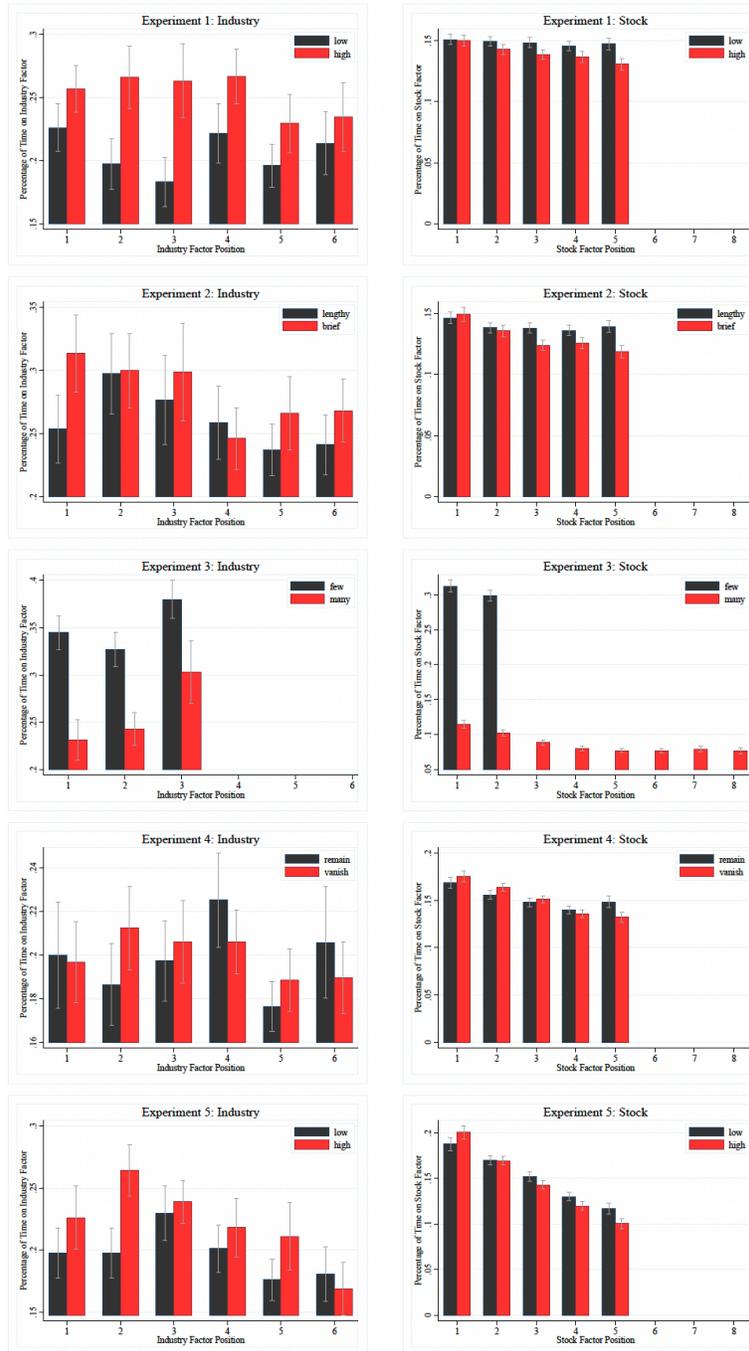
Supplemental Material, Table S4

*Across-Period Dynamics of Category Focus. Posterior estimates from Bayesian random effects models predicting category focus from period and block number. Block is a dummy variable indicating the second block. 95% credible intervals in brackets and  $P(\beta > 0)$  in parentheses.*

Across periods, there is some evidence of learning dynamics, as indicated by the positive coefficients on the period term in the regressions of Table S4. These dynamics appear only in the experimental conditions that demand a greater emphasis on the category, with hints of the opposite pattern in the other conditions. Across blocks, people tend to start off closer to the Bayesian benchmark in the second block, especially in the conditions fostering lower category focus. This is formally suggested by the generally positive coefficients on the block term in Table S4, although this effect could also simply reflect a spillover of strategies from one environment to the next.

# Appendix F

## Order Effects



Supplemental Material, Figure S8. Attention and Serial Position. Means shown with 95% confidence intervals.