# Attention constraints and learning in categories

Rahul Bhui[1,2] and Peiran Jiao[3]
[1]Department of Psychology, Harvard University
[2]Department of Economics, Harvard University
[3]Department of Finance, Maastricht University

May 18, 2020

**Abstract**

When different stimuli belong to the same category, learning about their attributes should be guided by this categorical structure. Here, we demonstrate how an adaptive response to attention constraints can bias learning toward shared qualities and away from individual differences. In three preregistered experiments using an information sampling paradigm with mousetracking, we find that people preferentially attend to information at the category level when idiosyncratic variation is low, when time constraints are more severe, and when the category contains more members. While attention is more diffuse across all information sources than predicted by Bayesian theory, there are signs of convergence toward this optimal benchmark with experience. Our results thus indicate a novel way in which a focus on categories can be driven by rational principles.

## Introduction

Categorical thinking helps us deal with our complex world in many ways. The use of categories is often thought to be optimized for making predictions under various constraints on information processing (Anderson, 1991; Markman & Ross, 2003; Fryer & Jackson, 2008). When forming judgments of stimulus magnitude, a bias toward category averages and a neglect of individual differences has been shown, in several cases, to reflect an adaptive response to limited or noisy information (Huttenlocher, Hedges, & Vevea, 2000; Feldman, Griffiths, & Morgan, 2009; Hemmer & Steyvers, 2009; Brady & Alvarez, 2011). When noise can be controlled by acquiring data, a related tendency to focus learning on category averages may reflect an adaptive response to attention constraints.

Suppose, for example, that you are an investor gathering information, trying to figure out how much various stocks should be worth. You have limited time and effort and cannot learn about all of the countless stocks in the market. There are idiosyncratic factors which affect the value of each stock separately, and shared factors which have a common impact on all stocks in the same industry. Do you focus your learning on the former or the latter? The efficient allocation of limited attention can favor learning about shared qualities because they are informative about all category members, whereas idiosyncratic information pertains only to each individual (Peng & Xiong, 2006). This mechanism could help explain why investors consult financial indices while neglecting data from specific firms (Huang, Huang, & Lin, 2019), why children prefer to inquire about kinds of animals rather than particular instances (Cimpian & Park, 2014), and why people forming impressions of others favor stereotypic information over attending to individuating attributes (Fiske & Neuberg, 1990).

However, it is not known whether rational principles account for human judgment in this setting. Prominent perspectives in psychology, neuroscience, and economics maintain that attention is spent where it has maximum benefit (Gottlieb & Oudeyer, 2018; Caplin, 2016), but the extent to which its allocation is guided by such hierarchical statistical structure in the environment remains unclear (Radulescu, Niv, & Ballard, 2019). If attention is deployed optimally, one's focus on shared qualities should adjust flexibly to environmental features that affect the value of information attainable from each source.

Several predictions follow from this core idea. 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, 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. Finally, 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.

In this study, we conducted experimental tests of these predictions to investigate how learning in categories can be driven by the optimal allocation of attention. Our task was designed to let us directly measure attention using mousetracking, while precisely controlling the statistical structure of information via an abstract sampling paradigm. This setup allowed us to carefully test the implications of a formal Bayesian model of attention allocation.

People playing our "stock prediction game" had to estimate the values of various hypothetical stocks based on a stream of incoming information. These values were generated by a known categorical structure—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 from the average). These factors varied stochastically and independently across trials. In a given trial, participants could reveal noisy signals in every moment of time 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 were thus able to allocate their attention with high temporal precision to learn about the relevant variables. In this way, we were able to investigate the theoretical predictions by measuring the amount of time spent mousing over each factor, and evaluating how this changed when we manipulated the factors' prior variances, the time limit, and the number of stocks. Using this approach we found that people do adapt their degree of categorical focus approximately in line with rational principles.

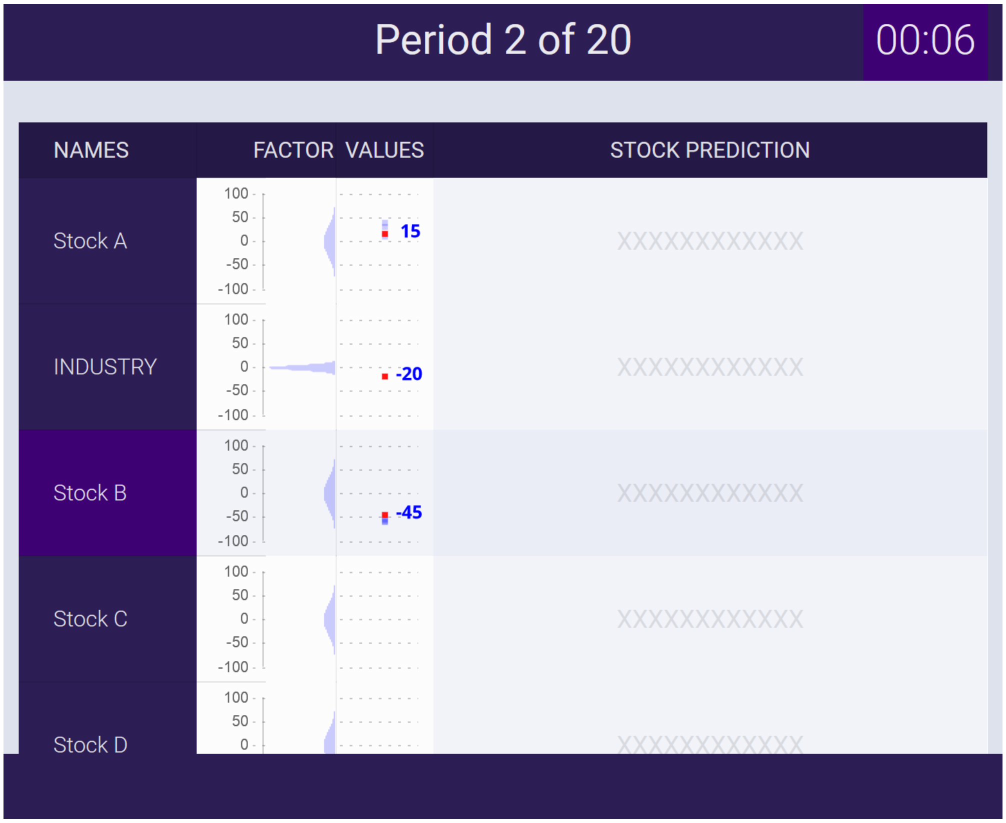


Figure 1: Screenshot of experiment.

| | $\sigma_{industry}$ | $\sigma_{stock}$ | $\sigma_{signal}$ | $n$ of stocks | *time* |
|---|---|---|---|---|---|
| Experiment 1 - low category variance | 5 | 30 | 10 | 5 | 12 s |
| Experiment 1 - high category variance | 30 | 5 | 10 | 5 | 12 s |
| Experiment 2 - long time limit | 30 | 5 | 10 | 5 | 20 s |
| Experiment 2 - short time limit | 30 | 5 | 10 | 5 | 8 s |
| Experiment 3 - few category members | 30 | 5 | 10 | 2 | 12 s |
| Experiment 3 - many category members | 30 | 5 | 10 | 8 | 12 s |

Table 1: Design parameters.

# Method

## Participants

Four hundred and thirty-eight participants from the United States were recruited on Amazon Mechanical Turk across three experiments (experiment 1, $n = 147$; experiment 2, $n = 146$; experiment 3, $n = 145$). 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 preregistration, and based on monetary constraints, informed partly by our pilot studies which suggested that this number should be sufficient to detect the hypothesized effects if present. Participants provided informed consent, and the study was approved by the Harvard Committee on the Use of Human Subjects. The preregistration can be found at https://osf.io/6mcqy/.

## Procedure

All three experiments used the same "stock prediction" paradigm (see Figure 1). In each trial, 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). These factors were generated independently in every trial from zero-mean Gaussian distributions portrayed on screen by sideways bell curves. For simplicity, all stock-specific factors had the same prior distribution.

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. 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 trial, participants recorded their estimate of each stock's total value (the sum of the two relevant factors) on sliders. They did not provide an estimate of the industry factor by itself. Upon submitting these estimates, they were shown the true stock values and the magnitude of their errors. At the end of the experiment, they were paid a bonus based on the average accuracy of their estimates according to a quadratic loss function of which they were informed.

We used a within-subjects design in each of the three 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. Experiment 1 varied the prior variances of the factors (simultaneously increasing $\sigma_{industry}$ from 5 to 30 and decreasing $\sigma_{stock}$ from 30 to 5), experiment 2 varied the time budget (decreasing time from 20 seconds to 8 seconds), and experiment 3 varied the number of stocks (increasing $n$ of stocks from 2 to 8). All of 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.

After reading the instructions, participants were provided with two self-paced practice trials in which they were told the true values of each factor and allowed unlimited time to sample informa-

| | Coefficient | Category focus | | |
|---|---|---|---|---|
| | | Posterior mean | Posterior 95% CI | $P(\beta > 0)$ |
| *Experiment 1* | Intercept | 0.067 | $[0.040, 0.095]$ | >0.999 |
| | Higher category variance | 0.053 | $[0.026, 0.080]$ | >0.999 |
| | $N$ | 100 | | |
| *Experiment 2* | Intercept | 0.129 | $[0.090, 0.167]$ | >0.999 |
| | Shorter time limit | 0.029 | $[-0.005, 0.064]$ | 0.951 |
| | $N$ | 100 | | |
| *Experiment 3* | Intercept | 0.039 | $[0.001, 0.079]$ | 0.976 |
| | Larger category size | 0.133 | $[0.093, 0.175]$ | >0.999 |
| | $N$ | 91 | | |

Table 2: Posterior estimates from Bayesian random effects models predicting relative attention to industry factor from experimental condition. $P(\beta > 0)$ denotes the posterior probability that the coefficient is positive. CI = credible interval.

tion. 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).

# Results

## Preregistered analysis

The following analyses 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 trials in which more than 20% of the available viewing time was spent not mousing over any factor.

The category focus—by which we mean the fraction of time spent mousing over the industry factor compared to the average stock factor—is displayed for each experiment in Figure 2. The results appear to be in line with the 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. The regressions were computed using the *brms* package in R with default weakly informative priors (Bürkner, 2017). They indicate positive effects of higher category-level variance ($P(\beta_{variance} > 0) > 0.999$), time pressure ($P(\beta_{time} > 0) = 0.951$), and category size ($P(\beta_{size} > 0) > 0.999$). Our preregistered test criterion of $P(\beta > 0) > 0.95$ is met in all three cases. Thus, participants appear to alter their patterns of attention as qualitatively predicted by the theory.

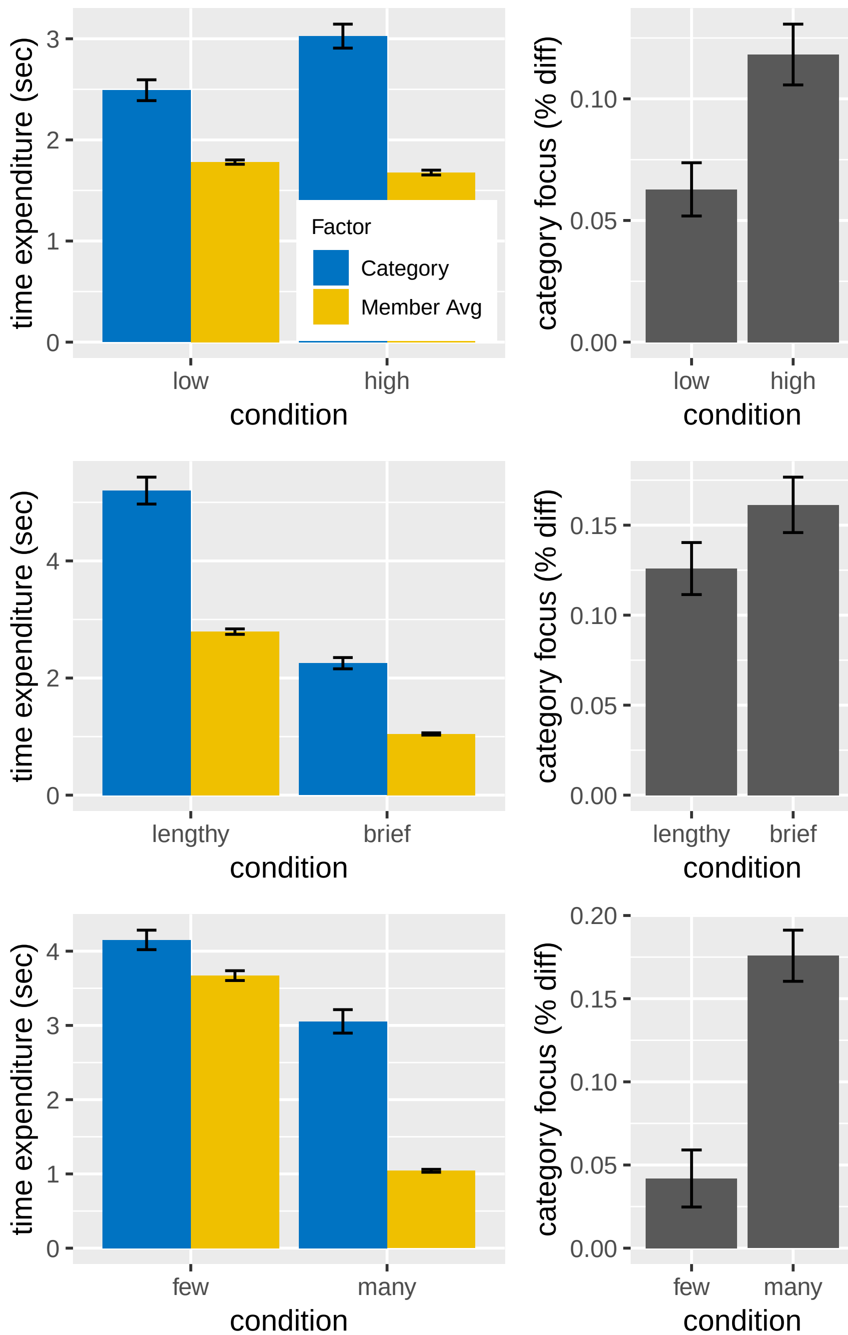


Figure 2: Attention allocation patterns. Mean time spent on each factor (left), and difference between proportions of time spent on category and average member (right). Low vs high category variance (experiment 1; top), lengthy vs brief time limit (experiment 2; middle), few vs many category members (experiment 3; bottom). Error bars depict 95% confidence intervals.

| | Coefficient | Relative proportion of time spent on category factor: Higher category variance | Lower category variance |
|---|---|---|---|
| *Experiment 1* | Intercept | 0.086 $[0.042, 0.131]$ (.999) | 0.042 $[0.000, 0.085]$ (.974) |
| | Trial | 0.010 $[0.003, 0.017]$ (.997) | $-0.004$ $[-0.009, 0.002]$ (.102) |
| | Block | 0.018 $[-0.046, 0.081]$ (.712) | 0.079 $[0.016, 0.140]$ (.995) |
| | Trial $\times$ Block | $-0.008$ $[-0.019, 0.002]$ (.049) | 0.002 $[-0.006, 0.010]$ (.685) |
| | | Shorter time limit | Longer time limit |
| *Experiment 2* | Intercept | 0.088 $[0.033, 0.146]$ (.999) | 0.102 $[0.048, 0.156]$ ($>$.999) |
| | Trial | 0.014 $[0.008, 0.021]$ ($>$.999) | 0.000 $[-0.007, 0.007]$ (.552) |
| | Block | 0.073 $[-0.008, 0.158]$ (.957) | 0.066 $[-0.008, 0.139]$ (.960) |
| | Trial $\times$ Block | $-0.013$ $[-0.023, -0.004]$ (.003) | $-0.004$ $[-0.013, 0.006]$ (.226) |
| | | Larger category size | Smaller category size |
| *Experiment 3* | Intercept | 0.100 $[0.056, 0.144]$ ($>$.999) | 0.023 $[-0.044, 0.087]$ (.770) |
| | Trial | 0.011 $[0.003, 0.020]$ (.995) | $-0.005$ $[-0.013, 0.003]$ (.128) |
| | Block | 0.036 $[-0.026, 0.100]$ (.867) | 0.101 $[0.013, 0.193]$ (.991) |
| | Trial $\times$ Block | 0.001 $[-0.011, 0.013]$ (.556) | $-0.007$ $[-0.018, 0.004]$ (.109) |

Table 3: Posterior estimates from Bayesian random effects models predicting relative attention to industry factor from trial and block number. 95% credible intervals in brackets and $P(\beta > 0)$ in parentheses.

### Exploratory analysis

While we find qualitative support for the theory, participants appear to fall short of the quantitatively optimal strategy laid out in the Appendix, which prescribes a particular level of category focus. As shown in Figure 3, their degree of category focus falls in between the Bayesian prediction and an equal split of their attention across all factors. Statistically, the average category focus is lower than the Bayesian benchmark in each case according to Bayesian regressions predicting category focus from individual-specific intercepts ($P(\beta < \bar{\lambda}) > .999$ in every condition; not reported). Attention is thus more diffuse across all information sources than the quantitative Bayesian benchmark. This deviation seems to be strongest at the outset of the experiment, generally in the first block participants encounter.

Note that this discrepancy cannot be naturally explained by a difference between objective and subjective signal precision. If people were able to only partially extract the information from the noisy signals, making their subjective precision lower than the objective precision, this would enhance rather than reduce the optimal focus on the category. Intuitively, blunted precision has an effect similar to that of a constraint on information processing capacity and amplifies the value of the focused strategy which makes efficient use of limited attention. Neither can the observations be reasonably captured by subjectivity in priors. A reduced category focus would require subjective prior variance to be higher than objective variance at the member level or lower at the category level. However, even doubling the former and halving the latter is not sufficient to totally account for the data, as shown in the Appendix.

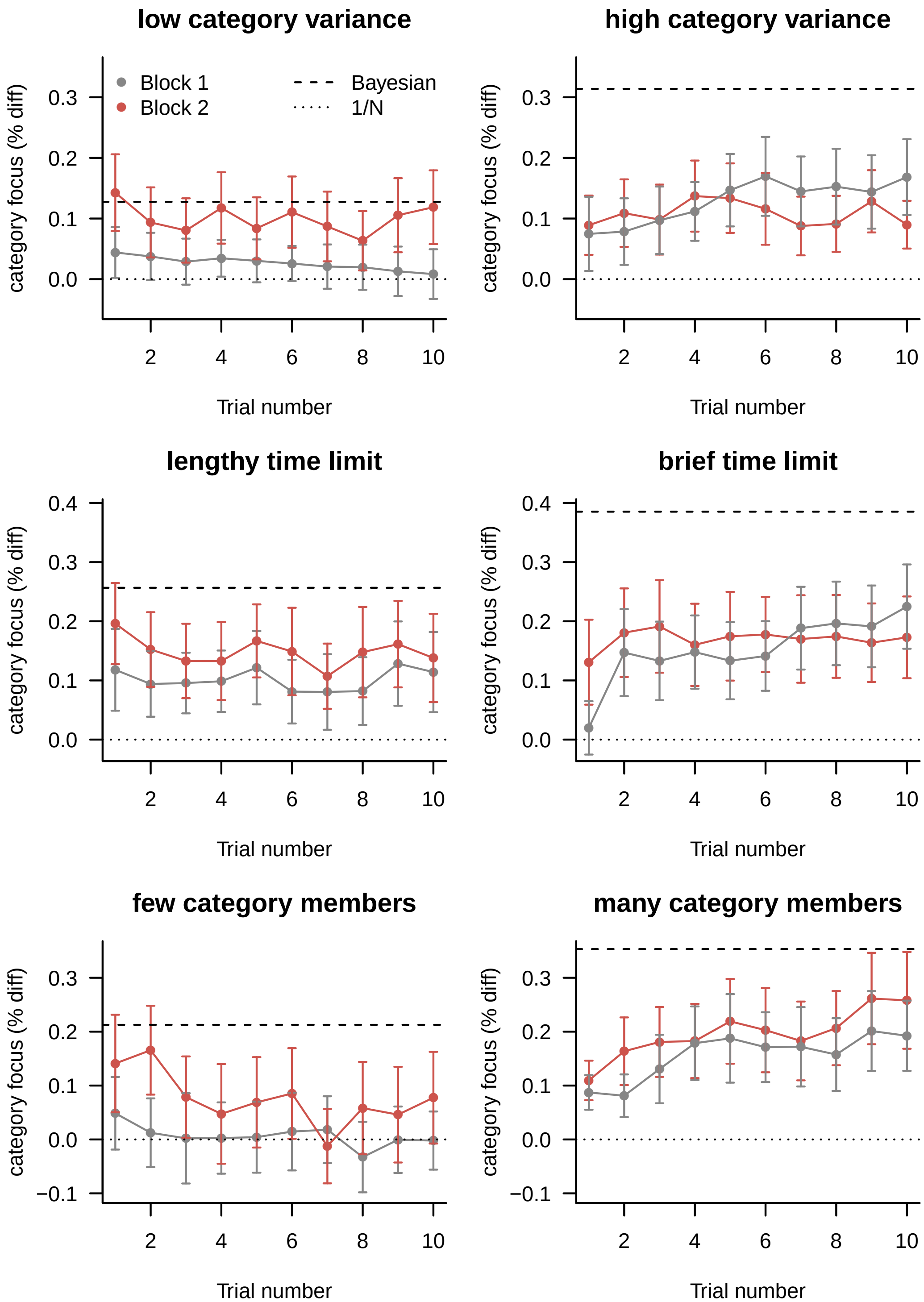


Figure 3: Attention allocation across trials. Color indicates which block was encountered first. Error bars depict 95% confidence intervals.

Do people learn toward the optimal strategy with experience? Across trials, there is some evidence of convergence, as indicated by the positive coefficients on the trial term in the regressions of Table 3. 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 more consistently move closer to the Bayesian benchmark. This is formally indicated by the generally positive coefficients on the block term in Table 3. Hence, the data taken together suggests that experience may bring people nearer to optimality.

## Discussion

The guiding role of hierarchical statistical structure in human judgment has been studied across a variety of domains such as reinforcement learning (Collins, 2018), planning (Tomov, Yagati, Kumar, Yang, & Gershman, 2020), decision making (Gershman, Malmaud, & Tenenbaum, 2017), visual perception (Gershman, Tenenbaum, & Jäkel, 2016), linguistics (Xu & Tenenbaum, 2007), and more (Tenenbaum, Kemp, Griffiths, & Goodman, 2011), and we help extend these investigations into the arena of attention. Related work on categorization has demonstrated that people selectively attend to features which are maximally diagnostic of category identity (Rehder & Hoffman, 2005; Matsuka & Corter, 2008; Blair, Watson, Walshe, & Maj, 2009; Blair, Watson, & Meier, 2009; Kim & Rehder, 2011; Meier & Blair, 2013), which could be consistent with Bayesian principles (Radulescu et al., 2019). This work has primarily focused on how categories are learned and inferred. Here, we take the hierarchical structure as given and consider how it can be used to guide attention for the purpose of forming stimulus judgments.

Our work is also motivated by a sizable literature across cognitive and decision science which casts attention in terms of optimal uncertainty reduction (Gottlieb & Oudeyer, 2018; Caplin, 2016). We use an abstract sampling paradigm which enables us to tightly control the structure of the information environment, and reveal attention transparently using mousetracking. These design features allow us to directly weigh the data against qualitative and quantitative benchmarks of optimality. The results indicate that people flexibly adjust their patterns of attention to category information in line with rational principles.

There may be some deviation from the precisely optimal strategy in the direction of a uniform attention split. A similar equity (aka $1/N$) heuristic has been documented in other allocation problems such as financial portfolio choice (Benartzi & Thaler, 2001) and parental effort investment (Hertwig, Davis, & Sulloway, 2002). In the former case, this bias toward naive diversification has been found to improve performance, potentially due to the way it approximately captures the principle of regularization or Bayesian shrinkage (Tu & Zhou, 2011). This rationale suggests that diffuse attention could reflect an inductive bias that improves performance when averaged across a broader spectrum of attention allocation problems, by preventing an agent from neglecting a potentially crucial information source. This speculative possibility awaits investigation. There is also some evidence of convergence toward the optimal strategy across both trials and blocks. These findings are consistent with the claim that rational inattention better describes judgment in repeated situations, where agents can discover the optimal strategy through experience (Maćkowiak, Matějka, & Wiederholt, 2018).

While we have focused on attention and learning, other mechanisms also contribute to an emphasis on categories, and may work in tandem with those described here. 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).

Much like investors, children inhabit a world in which the complexity of their environment dramatically outstrips their ability to process information. The adaptive framework we use here may help illuminate why young children preferentially seek out information about kinds of things rather than about concrete individuals (Cimpian & Park, 2014; Cimpian & Petro, 2014). This tendency seems somewhat sensitive to the potential for information gain, as it only emerges when talking to knowledgeable adults. If it indeed reflects an adaptive response to constraints on information processing, new predictions follow from the theory we study: children should be especially interested in category-level information when the category contains many members and when the individual differences between members are minimal. These predictions await empirical testing.

Future research can build on our work and address its limitations in several ways. First, other tests will be needed to determine whether our findings extend to more covert forms of attention (Carrasco, 2011). People might focus even more narrowly on a small number of information sources when they are not all overtly presented and some fail to come to mind. Second, more complex interactions between attention and decision making should be experimentally characterized. Certain facets of uncertainty may be especially beneficial to reduce in different problems, such as when information is gathered for the purpose of constructing a financial portfolio, and this may have far-reaching consequences (Peng & Xiong, 2006). Finally, although our study considered only instrumental reasons for acquiring information, non-instrumental motivations like curiosity are doubtless important. It is valuable to assess the extent to which these subjective preferences, too, may be ultimately grounded in rational principles (Gottlieb & Oudeyer, 2018; Gottlieb, Oudeyer, Lopes, & Baranes, 2013).

## Acknowledgments

We thank Florian Fröhlich for helpful research assistance. Support from the Pershing Square Fund for Research on the Foundations of Human Behavior is gratefully acknowledged.

## References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409.

Barberis, N., & Shleifer, A. (2003). Style investing. *Journal of Financial Economics*, *68*(2), 161–199.

Benartzi, S., & Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review*, *91*(1), 79–98.

Blair, M. R., Watson, M. R., & Meier, K. M. (2009). Errors, efficiency, and the interplay between attention and category learning. *Cognition*, *112*(2), 330–336.

Blair, M. R., Watson, M. R., Walshe, R. C., & Maj, F. (2009). Extremely selective attention: Eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1196.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392.

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.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525.

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.

Collins, A. G. E. (2018). Learning structures through reinforcement. In R. W. Morris, A. Bornstein, & A. Shenhav (Eds.), *Goal-Directed Decision Making: Computations and Neural Circuits* (pp. 105–123). Academic Press.

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*(4), 752.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 23, pp. 1–74). Elsevier.

Fryer, R., & Jackson, M. O. (2008). A categorical model of cognition and biased decision making. *BE Journal of Theoretical Economics*, *8*(1).

Gershman, S. J., Malmaud, J., & Tenenbaum, J. B. (2017). Structured representations of utility in combinatorial domains. *Decision*, *4*(2), 67.

Gershman, S. J., Tenenbaum, J. B., & Jäkel, F. (2016). Discovering hierarchical motion structure. *Vision Research*, *126*, 232–241.

Gottlieb, J., & Oudeyer, P.-Y. (2018). Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience*, *19*(12), 758–770.

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: Computational and neural mechanisms. *Trends in Cognitive Sciences*, *17*(11), 585–593.

Hemmer, P., & Steyvers, M. (2009). A Bayesian account of reconstructive memory. *Topics in Cognitive Science*, *1*(1), 189–202.

Hertwig, R., Davis, J. N., & Sulloway, F. J. (2002). Parental investment: How an equity motive can produce inequality. *Psychological Bulletin*, *128*(5), 728.

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.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*(2), 220.

Kim, S., & Rehder, B. (2011). How prior knowledge affects selective attention during category learning: An eyetracking study. *Memory & Cognition*, *39*(4), 649–665.

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.

Maćkowiak, B., Matějka, F., & Wiederholt, M. (2018). *Rational inattention: A disciplined behavioral model* (mimeo). CEPR Discussion Papers, DP13243.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, *129*(4), 592.

Matsuka, T., & Corter, J. E. (2008). Observed attention allocation processes in category learning. *Quarterly Journal of Experimental Psychology*, *61*(7), 1067–1097.

Meier, K. M., & Blair, M. R. (2013). Waiting and weighting: Information sampling is a balance between efficiency and error-reduction. *Cognition*, *126*(2), 319–325.

Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, *80*(3), 563–602.

Radulescu, A., Niv, Y., & Ballard, I. (2019). Holistic reinforcement learning: The role of structure and attention. *Trends in Cognitive Sciences*, *23*(4), 278–292.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*(1), 1–41.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.

Tomov, M. S., Yagati, S., Kumar, A., Yang, W., & Gershman, S. J. (2020). Discovery of hierarchical representations for efficient planning. *PLoS Computational Biology*, *16*(4), e1007594.

Tu, J., & Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, *99*(1), 204–215.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245.

# Appendix

## 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 decision maker is tasked with accurately estimating the values of several random variables, $v_i$, $i = 1, \ldots, n$, aiming to minimize the expected quadratic loss of his estimates. 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 judgments, the decision maker 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 $x_c^j$ or $x_{m_i}^j$, and all signals collectively as $x$. The decision maker 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 thus 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 thus

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

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

Due to independence of the components, the objective is equivalent to minimizing the sum of their expected posterior variances. Assuming an interior solution in what follows,

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

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

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

We evaluate how the category focus $\bar{\lambda}$ (i.e., the fraction of time spent on the category factor relative to the average fraction of time spent on the member-specific factors) changes based on various parameters:

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

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

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

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

$$\tag{10}$$

## 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)
- $-15$

## Sensitivity Analysis

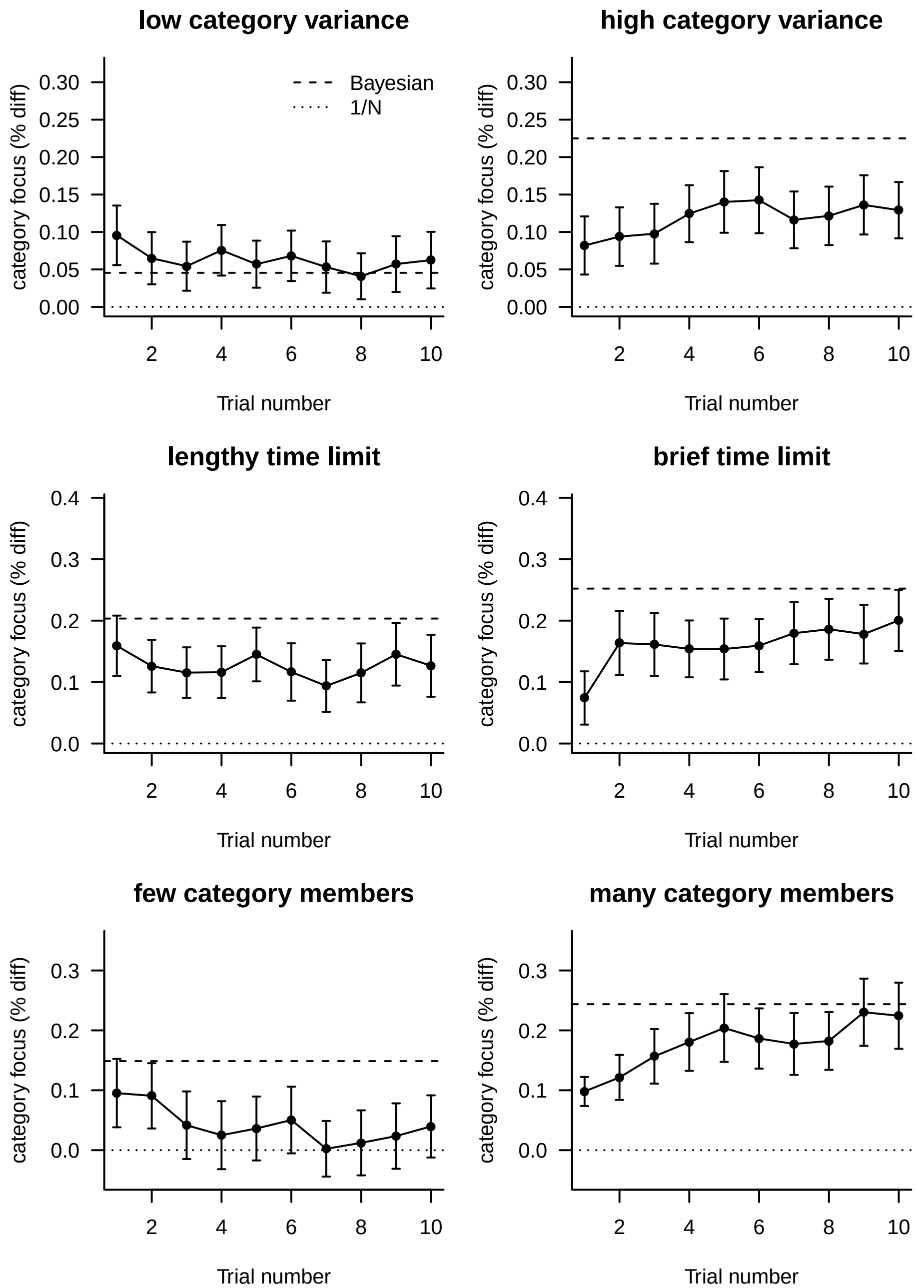


Figure 4: Attention allocation across trials with subjective member-level variance $= 2\sigma_{stock} = 10$ and subjective category-level variance $= \frac{1}{2}\sigma_{industry} = 15$. Error bars depict 95% confidence intervals. $P(\beta < \bar{\lambda}) > .999$ in all conditions except low category variance, according to Bayesian regressions predicting category focus from individual-specific intercepts (not reported).