THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2017
VOL. 70, NO. 3, 361–365
http://dx.doi.org/10.1080/17470218.2016.1159706

Routledge
Taylor & Francis Group

RAPID COMMUNICATION

On the blessing of abstraction

Samuel J. Gershman

Department of Psychology and Center for Brain Science, Harvard University, Cambridge, MA, USA

ABSTRACT

The "blessing of abstraction" refers to the observation that acquiring abstract knowledge sometimes proceeds more quickly than acquiring more specific knowledge. This observation can be formalized and reproduced by hierarchical Bayesian models. The key notion is that more abstract layers of the hierarchy have a larger "effective" sample size, because they combine information across multiple specific instances lower in the hierarchy. This notion relies on specific variables being relatively concentrated around the abstract "overhypothesis". If the variables are highly dispersed, then the effective sample size for the abstract layers will not be appreciably larger than for the specific layers. Moreover, the blessing of abstraction is counterbalanced by the fact that data are more informative about lower levels of the hierarchy, because there is necessarily less stochasticity intervening between specific variables and the data. Thus, in certain cases abstract knowledge will be acquired more slowly than specific knowledge. This paper reports an experiment that shows how manipulating dispersion can produce both fast and slow acquisition of abstract knowledge in the same paradigm.

ARTICLE HISTORY
Received 8 November 2015
Accepted 25 February 2016
First Published Online 10 March 2016

KEYWORDS
Bayesian inference; Learning to learn; Induction; Abstraction

One reason to acquire abstract knowledge is that it facilitates the acquisition of new specific knowledge. This "learning to learn" (Harlow, 1949) is evident in many domains. For example, by the age of 24 months children learn that shape tends to be homogeneous within object categories (the "shape bias"; Heibeck & Markman, 1987; Landau, Smith, & Jones, 1988), allowing them to extend a category label to novel, similarly shaped objects after seeing a single category exemplar. Human motor control similarly benefits from learning abstract task structure (Braun, Aertsen, Wolpert, & Mehring, 2009). In rats, repeatedly reversing which of two actions is rewarded leads to progressively faster reversal, even after a single error (Buytendijk, 1930; Dufort, Guttman, & Kimble, 1954).

Learning to learn is puzzling: how can abstract knowledge facilitate learning if it must also be learned from specific examples? Hierarchical Bayesian models (HBMs) offer a way out of this puzzle, by formalizing how learning can occur simultaneously at multiple levels of abstraction (Gershman & Niv, 2015; Kemp, Goodman, & Tenenbaum, 2010; Kemp, Perfors, & Tenenbaum, 2007; Lucas & Griffiths, 2010). Specific variables are constrained by abstract variables by virtue of a probabilistic relationship between the two. For example, the distribution of dog sizes is centred on the prototypical dog size, an example of an "overhypothesis" in the terminology of Goodman (1955). Learning at both levels is accomplished by using Bayes' rule to form a joint posterior distribution over hypotheses and overhypotheses.

Importantly, learning in HBMs can sometimes be faster at more abstract levels—a phenomenon dubbed the "blessing of abstraction" by Goodman, Ullman, and Tenenbaum (2011), who contrasted it with connectionist approaches to knowledge acquisition that build from the specific to the more abstract (e.g., Hinton, Osindero, & Teh, 2006). The blessing of abstraction is similar to phenomena occurring at a much faster timescale in visual perception, where a coarse "gist" is extracted before fine-grained details (Hegdé, 2008; Hochstein & Ahissar, 2002). Abstract-

CONTACT Samuel J. Gershman gershman@fas.harvard.edu Department of Psychology, Harvard University, 52 Oxford St., room 295.05, Cambridge, MA 02138, USA.
I am grateful to Tomer Ullman and Josh Tenenbaum for helpful conversations.

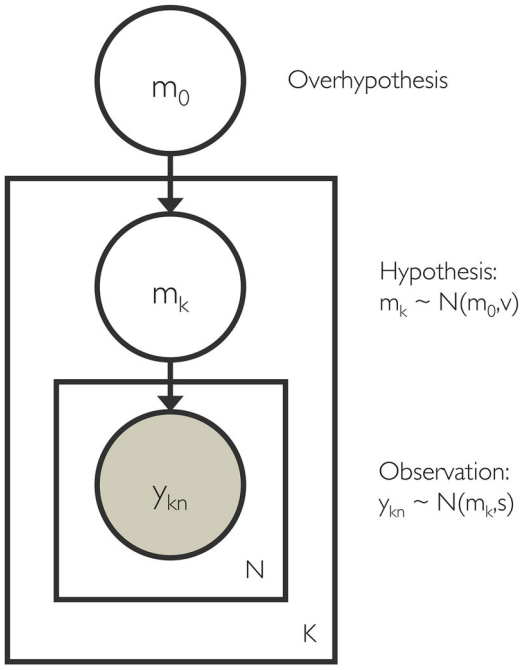

Figure 1. Generative model showing the probabilistic relationships between variables in the experiment. Unshaded nodes represent latent variables, shaded nodes represent observed variables, and plates denote replicates. K is the number of buttons and N is the number of observations for each button.

before-specific learning and coarse-to-fine processing arise in HBMs because the abstract variables of the hierarchy typically connect to multiple specific variables (e.g., the category "dogs" includes many specific dogs). By combining information across specific variables, the sample size for abstract variables is effectively larger than for specific variables.

Crucially, the blessing of abstraction depends on the dispersion of specific variables around the central tendency induced by the overhypothesis.[1] Dispersion could be controlled by the topology of the HBM (a pyramidal structure, with abstract variables at the top of the pyramid, will tend to produce the blessing) or its functional form (parameters governing the variance of specific variables conditional on abstract variables). While dogs tend to be similar sizes (low dispersion), the distribution of plant sizes can range from tiny flowers to enormous trees. In the latter case, the overhypothesis will not strongly constrain generalization to new examples, and we might expect learning at the abstract level to be slower than at the specific level. Indeed, from an information-theoretic perspective, data will tend to constrain specific variables more than abstract variables, because there are fewer sources of stochasticity inter-

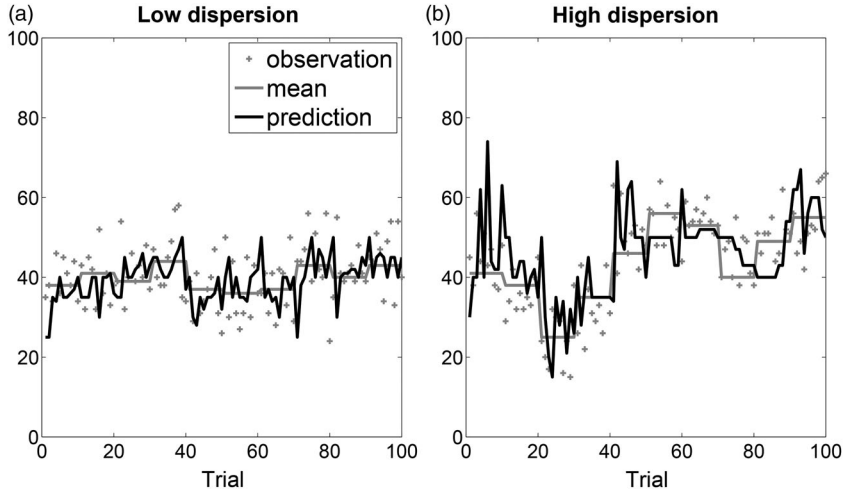

Figure 2. Example sequences of observations, the mean on each block, and human predictions (taken from two different participants). (A) Low dispersion condition. (B) High dispersion condition.

[1]More precisely, it depends on the dispersion at the specific level relative to the dispersion at the abstract level.

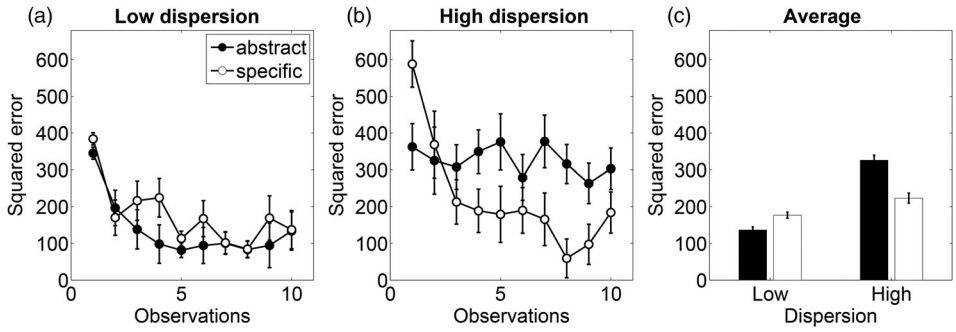

Figure 3. Experimental results. (A) Learning curves for the low dispersion condition. The "abstract" learning curve shows the mean squared error on the first trial of every block. The "specific" learning curve shows the mean squared error on each trial of the first block. Error bars represent standard errors. (B) Learning curves for the high dispersion condition. (C) Errors averaged across all points on each learning curve.

vening between specific variables and the data (Goel, 1983).

The study reported in this paper tests this basic prediction of HBMs by manipulating dispersion and comparing the learning curves for abstract and specific inferences. We predicted that the blessing of abstraction would only occur when dispersion was low, and that it would reverse for high dispersion.

Experimental study

Participants were presented with a simple prediction game: predict the scalar numerical output of a button. On each trial, participants made a prediction and then received feedback. Participants were faced with 10 different buttons, interacting with each one 10 times. To measure the blessing of abstraction, we computed two learning curves: an abstract learning curve (the squared error on the first trial of every button) and a specific learning curve (the squared error on each trial for the first button). The abstract learning curve gives us insight into the learned inductive bias, before any specific information about a button has been experienced. The specific learning curve gives us insight into learning prior to the formation of any inductive bias.

Method

Participants

Two hundred and seventeen participants (117 in the low dispersion condition, 100 in the high dispersion condition, 56% male, ages 21–44) were recruited for the experiment through the Amazon Mechanical Turk web service. The participants were paid 1 dollar for their participation. The experiment was approved by the Harvard Institutional Review Board.

Materials and procedure

The experimental interface consisted of a coloured button and a text-entry box in which participants entered their prediction on a 0 to 100 scale. Each block of trials had a different randomly coloured button. Participants entered their prediction and then clicked the button to receive feedback. Participants completed 10 blocks of 10 trials each, lasting a total of approximately 15 minutes. Participants in both conditions were instructed as follows:

> In this task, your job is to predict the pay-offs of slot machines (symbolized by coloured buttons). You will be shown 10 different slot machines, 10 times each. You will first be asked to guess the pay-off (between 0 and 100) and rate your confidence in your guess (1 = least confident, 10 = most confident). You will then receive feedback about the pay-off. The slot machine pay-offs are noisy, so no slot machine will give the same pay-off every time.

Each button k was associated with a Gaussian distribution $\mathcal{N}(m_k,s)$ over observation y_{kn} on trial n, where m_k is the mean and $s = 25$ is the variance. The mean was drawn from $\mathcal{N}(m_0,v)$ where $m_0 = 40$ is the global mean across all buttons, and v is the global variance. The global variance was manipulated between-subject: $v = 36$ in the low dispersion condition and 144 in the high dispersion condition. The generative process just described is displayed as a graphical model in Figure 1. Several example sequences of

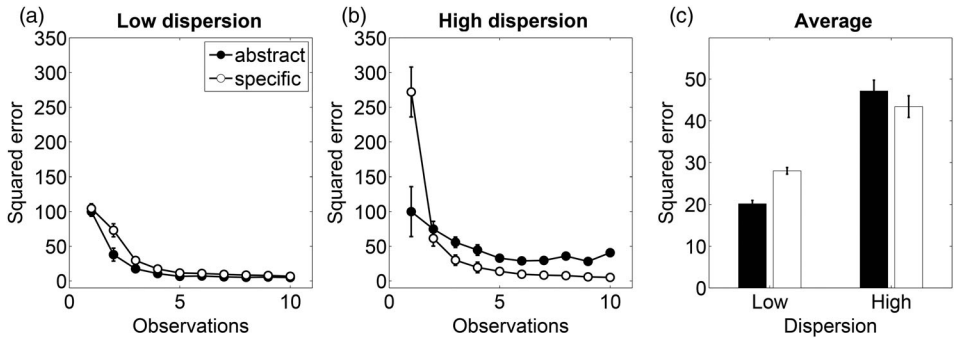

Figure 4. Hierarchical model predictions. Same format as Figure 3. (A) Learning curves for the low dispersion condition. (B) Learning curves for the high dispersion condition. (C) Errors averaged across all points on each learning curve.

observations and human predictions are shown in Figure 2.

To measure the specific learning curve, we computed the mean squared error (across participants) between the mean (m_k) on the first block and the participants' predictions on each trial. The abstract learning curve was computed similarly, but the global mean (m_0) was used instead of the block-specific mean.

Results

The learning curves for the low and high dispersion conditions are shown in Figure 3, revealing the expected result: abstract learning is slightly faster than specific learning in the low dispersion condition, but this pattern reverses in the high dispersion condition. To assess this result quantitatively, we averaged each learning curve (shown in Figure 3c) and performed a 2 (low vs. high dispersion) by 2 (abstract vs. specific) ANOVA. The main effects were not significant ($p > .35$) but there was a significant interaction [$F(1,430) = 12.67$, $p < .001$]. The interaction was also significant when only looking at the final datapoint on each learning curve [$F(1,430) = 4.4$, $p < .05$]. Post-hoc tests showed that abstract error was significantly lower than specific error in the low dispersion condition [$t(116) = 2.50$, $p < .05$] and significantly higher in the high dispersion condition [$t(99) = 3.80$, $p < .001$].

To compare the experimental results with the predictions of an HBM, we implemented the ideal Bayesian learner for this task (due to space limitations, details are omitted). The model predictions are shown in Figure 4. These predictions were made using the true generative parameters (no free parameters), though the predictions are generally robust to deviations from these parameters. The modelling results demonstrate that the ideal Bayesian learner captures the key interaction between dispersion and level of abstraction. We compared the HBM predictions to the predictions of a non-hierarchical model which assumed that each block was learned independently. As shown in Figure 5, the non-hierarchical model did not capture the

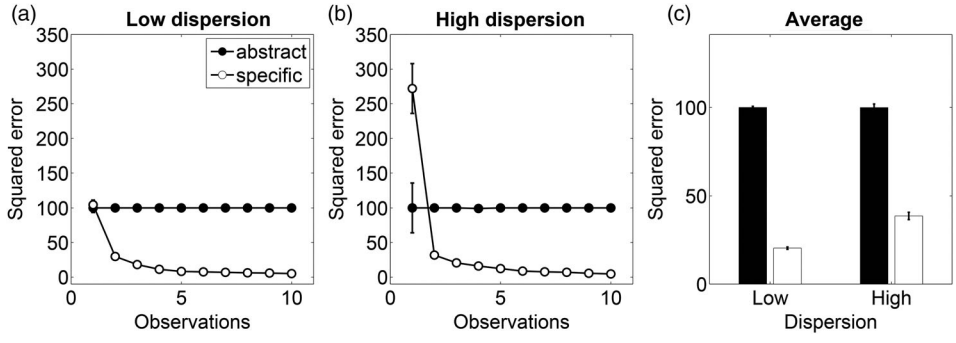

Figure 5. Non-hierarchical model predictions. Same format as Figure 3. (A) Learning curves for the low dispersion condition. (B) Learning curves for the high dispersion condition. (C) Errors averaged across all points on each learning curve.

interaction effect, because it has no means of learning at the abstract level.

Discussion

The results of this study provide important boundary conditions on the blessing of abstraction. When the dispersion of specific variables was low, we observed a blessing of abstraction (faster learning of the global mean relative to the block-specific mean), but when dispersion was high, we observed a reversal of the blessing (slower learning of the global mean). This finding is consistent with HBMs (Gershman & Niv, 2015; Goodman et al., 2011; Kemp et al., 2007, 2010; Lucas & Griffiths, 2010), in that the statistical aggregation of information across specific variables is only beneficial when the dispersion is low. When dispersion is high, the benefits of this aggregation are swamped by the additional uncertainty contributed by noise.

While our results are consistent with HBMs, they are not uniquely predicted by them. For example, connectionist models can also be designed to learn at multiple levels of abstraction (e.g., Rogers & McClelland, 2004). An appropriately configured exemplar model, with similarities determined by button identity, may also predict our findings. However, it is not exactly fair to compare Bayesian and exemplar models, since these models are formulated at different levels of analysis. Bayesian models describe the rational solution to an inductive inference problem, whereas exemplar models are mechanistic descriptions of the underlying psychological process. Indeed, exemplar models can be viewed as psychological implementations of Bayesian inference (Shi, Griffiths, Feldman, & Sanborn, 2010). Tying together rational and mechanistic theories is an important task that is not directly addressed in this paper.

ORCID


Samuel J. Gershman http://orcid.org/0000-0002-6546-3298

References

Braun, D. A., Aertsen, A., Wolpert, D. M., & Mehring, C. (2009). Motor task variation induces structural learning. *Current Biology*, *19*, 352–357.

Buytendijk, F. (1930). Über das umlernen. *Archives Néerlandaises de Physiologie de l'Homme et des Animaux*, *15*, 283–310.

Dufort, R. H., Guttman, N., & Kimble, G. A. (1954). One-trial discrimination reversal in the white rat. *Journal of Comparative and Physiological Psychology*, *47*, 248–249.

Gershman, S. J., & Niv, Y. (2015). Novelty and inductive generalization in human reinforcement learning. *Topics in Cognitive Science*, *7*, 391–415.

Goel, P. K. (1983). Information measures and Bayesian hierarchical models. *Journal of the American Statistical Association*, *78*, 408–410.

Goodman, N. (1955). *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.

Goodman, N. D., Ullman, T. D., & Tenenbaum, J. B. (2011). Learning a theory of causality. *Psychological Review*, *118*, 110–119.

Harlow, H. F. (1949). The formation of learning sets. *Psychological Review*, *56*, 51–65.

Hegdé, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Progress in Neurobiology*, *84*, 405–439.

Heibeck, T. H., & Markman, E. M. (1987). Word learning in children: An examination of fast mapping. *Child Development*, *58*, 1021–1034.

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*, 1527–1554.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*, 791–804.

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2010). Learning to learn causal models. *Cognitive Science*, *34*, 1185–1243.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*, 307–321.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*, 299–321.

Lucas, C. G., & Griffiths, T. L. (2010). Learning the form of causal relationships using hierarchical Bayesian models. *Cognitive Science*, *34*, 113–147.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Shi, L., Griffiths, T. L., Feldman, N. H., & Sanborn, A. N. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review*, *17*, 443–464.