Psychological Review
2009, Vol. 116, No. 4, 998–1002

0033-295X/09/$12.00 DOI: 10.1037/a0017113

Empirical and Computational Support for Context-Dependent Representations of Serial Order: Reply to Bowers, Damian, and Davis (2009)

Matthew M. Botvinick
Princeton University

David C. Plaut
Carnegie Mellon University

J. S. Bowers, M. F. Damian, and C. J. Davis (2009) critiqued the computational model of serial order memory put forth in M. Botvinick and D. C. Plaut (2006), purporting to show that the model does not generalize in a way that people do. They attributed this supposed failure to the model's dependence on context-dependent representations, translating this argument into a general critique of all parallel distributed processing models. The authors reply here, addressing both Bowers et al.'s criticisms of the Botvinick and Plaut model and the former's assessment of parallel distributed processing models in general.

Keywords: serial order, working memory, computational models

In Botvinick and Plaut (2006), we proposed a novel neural network model of short-term memory for serial order. We reply here to a critique of this work put forth by Bowers, Damian, and Davis (2009).

We are grateful to Bowers et al. (2009) for clearly acknowledging a distinctive strength of the Botvinick and Plaut (2006) model: its ability to account for the impact of domain-specific background knowledge on serial recall. As discussed at length in Botvinick and Plaut, immediate serial recall in humans is strongly affected by the structure of the sequences to be remembered. When this structure fits well with the statistics of previously encountered material, enhanced recall is generally observed. The fact that the Botvinick and Plaut model also displays this characteristic allows it to account naturally for effects of bigram frequency and phonotactic regularity and also to simulate a rich set of results from studies examining memory for sequences generated from artificial grammars (Botvinick, 2005; Botvinick & Bylsma, 2005).

Although Bowers et al. (2009) did acknowledge this aspect of our model, they then went on to portray it not as an asset but as a fatal flaw. The Botvinick and Plaut (2006) model, they argued, is "too sensitive" to the details of previous experience. In particular, they suggested, the model shows an insufficient ability to encode arbitrary new sequences, including sequences containing previously unencountered items. Bowers et al. attributed this supposed difficulty to the way in which the Botvinick and Plaut model represents sequence information. As detailed in our original article, the model relies on conjunctive representations of item and order, in which the way that an item is represented depends on the ordinal position at which it appears. Bowers et al. saw dire problems in such "context-sensitive" representations. Indeed, portraying context-sensitive representation as a hallmark of connectionist or parallel distributed processing (PDP) models, they asked the reader to accept their observations concerning the Botvinick and Plaut model as grounds for rejecting connectionist models in general.

In principle, we welcome the kind of theoretical debate Bowers et al. (2009) sought to generate. However, their specific assertions simply do not hold up to scrutiny. As we explain in the sections below, their critique of the Botvinick and Plaut (2006) model is directly contradicted by abundant empirical data and rests upon simulation results that are either innocuous or irrelevant. Their larger critique of connectionist modeling is predicated on false premises concerning the functional implications of context-dependent representation, as well as a basic misunderstanding of the connectionist approach. All in all, we worry that their critique of our work may have done more to obscure the important issues than to illuminate them.

Taking Account of the Empirical Data

Despite the skepticism Bowers et al. (2009) expressed toward context-dependent representations, once all the relevant empirical data are considered, it becomes very nearly impossible to deny the involvement of such representations in serial order memory. In particular, although Bowers et al. made no mention of it, there have been at least seven studies examining neural activity during serial recall in nonhuman primates (Barone & Joseph, 1989; Funahashi, Inoue, & Kubota, 1997; Inoue & Mikami, 2006; Kermadi & Joseph, 1995; Kermadi, Jurquet, Arzi, & Joseph, 1993; Ninokura, Mushiake, & Tanji, 2003, 2004), and a central finding in every one of these has been that item information is encoded in a way that depends on serial position. A representative finding from the work of Inoue and Mikami (2006) is shown in Figure 1 (for further discussion, see Botvinick & Watanabe, 2007).

Matthew M. Botvinick, Department of Psychology and Neuroscience Institute, Princeton University; David C. Plaut, Department of Psychology, Department of Computer Science, and Center for the Neural Basis of Cognition, Carnegie Mellon University.

Correspondence concerning this article should be addressed to Matthew M. Botvinick, 3-S-13 Green Hall, Princeton University, Princeton, NJ 08540. E-mail: matthewb@princeton.edu

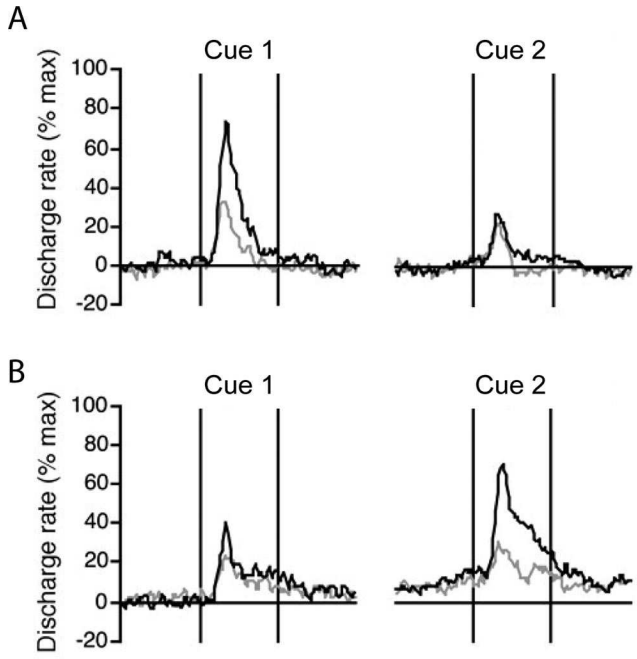

Figure 1. Response profiles for two prefrontal neurons, reported by Inoue and Mikami (2006), during sequential presentation of two visual shape cues. Both neurons displayed differential responses to preferred (black) and nonpreferred (grey) shapes, as well as differential responses across ordinal positions. The neuron contributing to the upper panels (A) responded preferentially to items occupying the first ordinal position; the neuron in the lower panels (B) responded preferentially to items in the second position. From "Prefrontal Activity During Serial Probe Reproduction Task: Encoding, Mnemonic and Retrieval Processes," by M. Inoue and A. Mikami, 2006, *Journal of Neurophysiology, 95,* p. 1014, Figure 4. Copyright 2006 by The American Physiological Society. Adapted with permission.

Unless Bowers et al. (2009) wish to assert that these neuroscientific data are spurious, or that they are not germane to human immediate serial recall, it seems unproductive to debate the relevance of context-dependent representations in this domain. Present evidence overwhelmingly suggests that the brain employs such representations. Barring the emergence of contradictory evidence, the scientific challenge is to determine how these representations are instantiated and updated during encoding, how they are maintained during retention, how they are read out during recall, and how they may be shaped by experience.

It is these questions that the Botvinick and Plaut (2006) model strives to address. Of course, its success in doing so must be judged against how well it accounts for the details of human performance in serial recall. Bowers et al. (2009) argued that the model fails in important ways and claimed to document its failure in a set of follow-up simulations. It is to these we now turn.

Simulation Results

Bowers et al. (2009) presented a series of seven simulations. It is surprising that the first six of these present no problem at all for the Botvinick and Plaut (2006) model. Together, these simulations simply show that the model generalizes to new sequences more readily when distributed, rather than localist, item representations are used. The fact that distributed representations support generalization has been a central message of connectionist modeling since the 1980s. To see this point substantiated yet again, in the context of serial recall, is satisfying, if not particularly surprising. In any case, we fail to see what challenge is implied to the Botvinick and Plaut account. Indeed, we noted in that article (p. 205) that our use of localist item representations was an implementational convenience, rather than a core aspect of the account, and indeed, some of the simulations reported in Botvinick and Plaut actually used distributed item representations.

This brings us to Bowers et al.'s (2009) Simulation 7. This simulation was intended to model a thought experiment, which lies at the heart of the critique. Bowers et al. asked the reader to imagine a scenario in which someone learns a new letter-name, "ree," and is asked to repeat back a short sequence containing this new item in the first ordinal position (e.g., "ree, B"). Having completed this task, they noted, the same individual should have no trouble then repeating back a sequence in which "ree" appears in another ordinal position (e.g., "B, ree"). In Simulation 7, Bowers et al. purported to show that the Botvinick and Plaut (2006) model cannot simulate this scenario. The model was first trained on a set of 25 items. A 26th item was then introduced, and the model received massive training on sequences in which this item appeared at position one. The model was then shown to have difficulty recalling sequences in which the new item appeared in other positions.

A little reflection should reveal that this simulation bears little relation to the scenario described in the thought experiment. The individual in that experiment did not receive massive training on sequences containing the novel item. Instead, he or she *immediately* repeated back one sequence ("ree-B"), and then another ("B-ree"). A proper simulation of the thought experiment would therefore involve presenting the model with a novel item and seeing whether it can recall that item at any ordinal position immediately, without any interposed training.

In fact, the model handles this situation just fine. To demonstrate this, we trained the model on items meant to represent syllables (reasoning that most letter names take this form). The input layer contained three sets of 10 units, respectively representing onset, nucleus, and coda.[1] A syllable was represented by activating one unit from each of these groups. The model was trained to recall sequences of syllables following the procedures used in Botvinick and Plaut (2006). It is important to note that one specific syllable was avoided during training. By analogy to Bowers et al.'s (2009) thought experiment, let us refer to this syllable as "ree." Following training, the model's weights were held constant, and it was tested on sequences in which ree appeared in position one (e.g., "ree, B"). The model recalled these sequences without error. Not surprising, the model then went on to perform perfectly on sequences in which ree appeared at other ordinal positions (e.g., "B, ree").

Naturally, we take no issue with the assertion that "a participant who recalls ree-B can also succeed on the sequence B-ree" (Bow-

[1] The model contained 75 hidden units and was trained on sequences from length one to three. Further simulation details are available from the lead author upon request.

ers et al., 2009, p. 17). What we have demonstrated here, contrary to the claims of Bowers et al., is that the Botvinick and Plaut (2006) model has absolutely no difficulty in simulating this scenario.[2]

False Premises Concerning Context-Dependent Representations

Having dispensed with the more concrete and specific points made by Bowers et al. (2009), we may now briefly address their main high-level theoretical assertions. Bowers et al. used their simulation results as a springboard to a broad critique of context-dependent representation. Their central thesis is that context-independent representations support the formation of arbitrary associations, whereas context-dependent representations do not. The only problem with this formulation is that it is not true. Indeed, both of its two premises are belied by existing computational models in the area of serial order memory. The first premise is contradicted, for example, by the model recently proposed by Burgess and Hitch (2006), which relies on links between context-independent item and position representations, but which shows recall performance strongly affected by earlier learning. The second premise, in turn, is contradicted by the primacy model proposed by Page and Norris (1998), which uses conjunctive representations of item and order, but which is capable of encoding arbitrary sequences. (Note that Bowers et al. incorrectly grouped the primacy model with models that employ context-independent representations of item and order. Unit activation in the primacy model depends on both the occurrence of a specific item and the ordinal position in which the item occurs. The model thus, by definition, employs a conjunctive code. Indeed, our 2006 model shares some important properties with the primacy model. The nature of the connection between the two models is suggested by Figure 5 from Botvinick & Plaut, 2006, which shows something very much like a primacy gradient.)

There is, of course, an interesting and important relationship between representational form and generalization behavior, a relationship that has been explored throughout the connectionist literature and well beyond. The nature of this relationship is, however, not as simple as the one Bowers et al. (2009) described.

A Straw-Man Critique of Connectionist Modeling

By criticizing the use of context-dependent representations, Bowers et al. (2009) intended to call into question connectionist modeling at large. However, by conflating connectionist/PDP modeling with context-dependent representation, Bowers et al. revealed a fundamental misunderstanding. They asserted that "a core claim of the PDP approach is that all knowledge is coded in a context dependent manner" (p. 7). This is simply untrue. In fact, the approach takes no specific stance on the degree to which the representation of an entity is dependent or independent of the contexts in which it occurs. Rather, one of the main tenets of the approach is to discover, rather than stipulate, representations (Plaut & McClelland, 2000). Internal representations are learned under the pressure of performing specific tasks, and the degree to which they exhibit context dependence is a consequence of basic network mechanisms, the learning procedure, and the structure of the tasks to be learned. For example, Plaut and Gonnerman (2000) trained a network to comprehend morphologically complex words that varied in semantic transparency (i.e., the degree to which the meaning of a word is consistent with the independent meanings of its component morphemes). When a set of words was embedded in an artificial language in which other words were transparent, the representations of their morphemes were largely context-independent (as reflected by patterns of morphological priming). By contrast, when the same words were embedded in a language in which other words were opaque, morphemes were given context-dependent representations in which the degree of dependence varied as a function of the word's transparency. Thus, contrary to Bowers et al.'s claims, PDP networks are not restricted to learning context-dependent representations but can develop representations that span the full range from context dependence to context independence. Indeed, Botvinick and Plaut (2006) specifically noted that their networks carried some context-independent information about item and order (p. 228) and discussed the potential role of weight-based associations between context-independent representations in serial order memory (p. 233).

Conclusions

In this brief reply, we have rejected the comments and conclusions of Bowers et al. (2009) on several grounds. First, we have noted the existence of abundant empirical data contradicting their arguments. Second, we have shown why their simulation results in fact present no challenge to our model. Third, we have questioned several basic premises underpinning their theoretical polemic.

If Bowers and colleagues are serious about modeling short-term memory for serial order, then the ball now bounces back to them and others who advocate for context-independent representations of item and order. Having failed to show that the Botvinick and Plaut (2006) model has unreasonable difficulty with arbitrary sequences, the burden of proof now falls upon Bowers and colleagues to demonstrate that models using context-independent representations can capture the effects of background knowledge that are so well captured by the Botvinick and Plaut model. Bowers and colleagues appealed to the notion that introducing "chunks" might allow this. This idea has been bandied about informally since at least the early 1980s, but we still await a serious attempt

[2] Out of curiosity, we reran Simulation 7 from Bowers et al. (2009), using the item representations just described. This yielded starkly different results from those reported by Bowers et al. Specifically, even after extensive training where ree only ever appeared in position one, the model had absolutely no trouble recalling that item at other ordinal positions. It turns out that the difficulty Bowers et al. observed in their Simulation 7 had nothing to do with serial order memory but related instead simply to item encoding. Upon initial presentation of the initially withheld item in their simulation, the model did not even correctly shadow the item during encoding, nor did the model correctly recall when it was presented in isolation (i.e., as a one-item list; J. Bowers, personal communication, March 13, 2009). The correlations among item features, given the structure of the training set, were evidently strong enough during initial training that the model had effectively been trained not to be able to encode the novel item, much as human subjects find it difficult to encode syllables that are phonotactically highly irregular, as shown, for example, by Brown and Hildum (1956). The results Bowers et al. reported thus reflect, in part, an idiosyncratic (and arguably ecologically invalid) aspect of their item representations.

to implement it in a runable model so that it can be tested against existing empirical data.

References

Barone, P., & Joseph, J. P. (1989). Prefrontal cortex and spatial sequencing in macaque monkey. *Experimental Brain Research, 78,* 447–464.

Botvinick, M. (2005). Effects of domain-specific knowledge on memory for serial order. *Cognition, 97,* 135–151.

Botvinick, M., & Bylsma, L. M. (2005). Regularization in short-term memory for serial order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 351–358.

Botvinick, M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review, 113,* 201–233.

Botvinick, M., & Watanabe, T. (2007). From numerosity to ordinal rank: A gain-field model of serial order representation in cortical working memory. *Journal of Neuroscience, 27,* 8636–8642.

Bowers, J. S., Damian, M. F., & Davis, C. J. (2009). A fundamental limitation of the conjunctive codes learned in PDP models of cognition: Comment on Botvinick and Plaut (2006). *Psychological Review, 116,* 986–997.

Brown, R. W., & Hildum, D. C. (1956). Expectancy and the perception of syllables. *Language, 32,* 411–419.

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55,* 627–652.

Funahashi, S., Inoue, M., & Kubota, K. (1997). Delay-period activity in the primate prefrontal cortex encoding multiple spatial positions and their order of presentation. *Behavioural Brain Research, 84,* 203–223.

Inoue, M., & Mikami, A. (2006). Prefrontal activity during serial probe reproduction task: Encoding, mnemonic and retrieval processes. *Journal of Neurophysiology, 95,* 1008–1041.

Kermadi, I., & Joseph, J. P. (1995). Activity in the caudate nucleus of monkey during spatial sequencing. *Journal of Neurophysiology, 74,* 911–933.

Kermadi, I., Jurquet, Y., Arzi, M., & Joseph, J. P. (1993). Neural activity in the caudate nucleus of monkeys during spatial sequencing. *Experimental Brain Research, 94,* 352–356.

Ninokura, Y., Mushiake, H., & Tanji, J. (2003). Representation of the temporal order of visual objects in the primate lateral prefrontal cortex. *Journal of Neurophysiology, 89,* 2868–2873.

Ninokura, Y., Mushiake, H., & Tanji, J. (2004). Integration of order and object information in monkey lateral prefrontal cortex. *Journal of Neurophysiology, 71,* 550–560.

Page, M., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105,* 761–781.

Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects compatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes, 15,* 445–485.

Plaut, D. C., & McClelland, J. L. (2000). Stipulating versus discovering representations [Commentary on M. Page, Connectionist modeling in psychology: A localist manifesto]. *Behavioral and Brain Sciences, 23,* 489–491.

Received March 23, 2009
Revision received June 1, 2009
Accepted June 4, 2009 ■

Postscript: Winnowing Out Some Take-Home Points

Matthew M. Botvinick
Princeton University

David C. Plaut
Carnegie Mellon University

Some arguments made in Bowers, Damian, and Davis's (2009) rebuttal can be dispensed with quickly. For example, they made much of the fact that the neurophysiological studies that have reported conjunctive coding for item and order have also often reported neurons that code for item independent of order and vice versa. However, this is immaterial; as noted both in our Reply and in our original article, the Botvinick and Plaut (2006; BP06) model contains units with the same coding profiles. We can also dispense quickly with the assertion that the primacy model of Page and Norris (1998) uses order-independent item representations. The units in that model do respond to specific items, but they assume different activation levels depending on the ordinal position at which items occur. That is, by definition, conjunctive coding for item and order.

This clears the way for us to consider the new simulations that Bowers et al. (2009) presented in their rebuttal. Their first two simulations sought, in part, to highlight a point we hoped was obvious from the simulation we presented in our earlier reply: The BP06 model can recall an item at an ordinal position where that item has not previously occurred *if* the item is represented in terms of a set of subordinate features, *and* each of those features has previously occurred at the relevant position. It was perhaps worthwhile for Bowers et al. to emphasize this point (see also Marcus, 1998), but it hardly seems a blow to the BP06 model. What they show is that, if the environment is diabolical enough to place a low-level feature in a position where it has never occurred during the millions of events that make up the entire history of experience of the system, poor recall performance will result. However, we doubt that the environment human learners inhabit is quite so adversarial. Moreover, even if it were, there is no evidence that humans can generalize on the basis of features that have no overlap at any level of representation with previously encountered features (see McClelland & Plaut, 1999, for discussion).

This leads us to one take-home point, which is that the adequacy of the BP06 model, like that of all parallel distributed processing (PDP) models, must be assessed in the context of an ecologically valid training set, a set of training examples that reflects the relevant statistical properties of a real-world learning environment (see Botvinick & Plaut, 2006). By focusing on highly contrived and implausible training contexts, Bowers et al. (2009) shed little light on the psychologically relevant properties of our model. As the saying goes, garbage in, garbage out.

Based on their final simulation, Bowers et al. (2009) went on to pose a false dilemma. They argued that PDP models can retain either the ability to encode arbitrary new patterns or the ability to benefit from previous experience with specific patterns, but not both. There is certainly a logical tradeoff between these two capacities, because as increasing influence is accorded to previous experience, there is an inevitable reduction in the flexibility needed to process novel events (see McClelland, McNaughton, & O'Reilly, 1995). However, this tradeoff represents a graded continuum, not a polar contrast. In their simulation using the BP06

model, Bowers et al. seem to have found a corner of parameter space where compositional coding predominates, and previous experience with specific feature combinations has little impact on recall performance. However, a cursory glance at the PDP literature makes clear that, given a reasonable training environment, PDP models are quite capable of striking a balance between flexibility on the one hand and sensitivity to structure on the other. Indeed, this point has been made in precisely the domain Bowers et al. staked out for their own demonstration: word versus nonword reading (see Plaut & McClelland, 1993).

There is some irony in the fact that Bowers et al.'s (2009) closing criticism of our model focused on the issue of sensitivity to previous learning, because, as we emphasized in BP06, this is an area where traditional "context-independent" models face substantial difficulty. To our gratification, the last couple of years have seen a resurgence of interest in the role of learning in short-term memory for serial order (see Thorn & Page, 2008). The critique of our model offered by Bowers et al. contributes to this welcome development by drawing attention to the inherent tradeoff between learning and flexibility. In so doing, however, the critique does little to challenge the viability of the BP06 model.

References

Botvinick, M., & Plaut, D. C. (2006). Such stuff as habits are made on: A reply to Cooper and Shallice (2006). *Psychological Review, 113,* 917–927.

Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology, 37,* 243–282.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why are there complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419–457.

McClelland, J. L., & Plaut, D. C. (1999). Does generalization in infant learning implicate abstract algebra-like rules? *Trends in Cognitive Science, 3,* 166–168.

Page, M., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105,* 761–781.

Plaut, D. C., & McClelland, J. L. (1993). Generalization with componential attractors: Word and nonword reading in an attractor network. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 824–829). Hillsdale, NJ: Erlbaum.

Thorn, A., & Page, M. (2008). *Interactions between short-term and long-term memory in the verbal domain*. Hove, United Kingdom: Psychology Press.

Correction to Smith and Ratcliff (2009)

In the article, "An Integrated Theory of Attention and Decision Making in Visual Signal Detection," by Philip L. Smith and Roger Ratcliff (*Psychological Review*, 2009, Vol. 116, No. 2, pp. 283–317), there is an error on p. 284 in the right-hand column. In the sentence "On each trial, the angular position of the cue, α, ($0 < \alpha \leq 360°$), was selected randomly; the uncued locations were at α + 120°", the plus sign should have been a plus/minus symbol. The correct sentence is presented below.

"On each trial, the angular position of the cue, α, ($0 < \alpha \leq 360°$), was selected randomly; the uncued locations were at $\alpha \pm 120°$."

DOI: 10.1037/a0016900