

The relationship between decision and action: Simulating response dynamics in categorization

Rick Dale (radale@memphis.edu)

Department of Psychology, University of Memphis, Memphis, TN, 38152

Abstract

A neural network simulation of taxonomic categorization is presented. Parameters of the model permit exploration of different interactivity between cognitive processing and a simple output component. By matching the simulation with human data, the model reveals that fluid interaction between cognition and action may be what produces graded response dynamics observed in human categorization experiments.

Keywords: Neural network, categorization, action, dynamics.

Introduction

Cognition is often thought to happen in places between the sensors and effectors, with sharp delineations at these boundaries (e.g., Fodor, 1983; Pylyshyn, 2000). This common account envisages cognition collapse its decisions onto the effectors, directing various motor systems in an all-or-none fashion. For example, well-known theories of attention posit central processing separate from systems responsible for guiding action (e.g., Posner & Petersen, 1990). In rapid, frequent, and metabolically cheap motor output, such as the ballistic saccades of the eyes, this perspective seems most apt. When the eyes are drawn by an external stimulus, the processing that has led to the saccade is programmed over a couple hundred milliseconds, and once it reaches premotor regions responsible for saccadic movement, fires off a motor instruction that may differ from moment to moment only because of noise. In other words, once the command has been issued, and the eyes instructed on their movement, a ballistic, linear motion to the target should be observed (e.g., Becker, 1991).

Recently, however, a growing body of research has shown that motor systems interact more richly with cognitive processes. Even in oculomotor systems, the trajectory of a saccade may vary in curvature depending on stimulus context (e.g., Sheliga, Riggio, & Rizzolatti, 1994). Perhaps more compelling is evidence that *manual* responses exhibit a range of dynamic characteristics that reveal processing. For example, force and velocity of responses *after* initiation vary concomitantly with frequency in a lexical decision task (e.g., Balota & Abrams, 1995), and response and stimulus probability in simple reaction-time tasks (Mattes, Ulrich, & Miller, 2002; see also Tipper, Howard, & Jackson, 1997). More recently, Dale, Kehoe, and Spivey (in press; see also Spivey, Grosjean, & Knoblich, 2005) show that computer-mouse trajectories indicate competition between categories when classifying atypical animal exemplars. The dynamic characteristics of the response, in the form of mouse trajectories, revealed an attraction towards a featurally-

similar category label (e.g., *fish*) for atypical exemplars (e.g., *whale*) compared to control trials (e.g., *cat*).

These results show that processing flows in systematic ways into motor behavior, rather than simply collapsing onto them to generate a categorical response. They may indeed recommend a “cascadic flow” perspective on cognition that sees information flow continuously from sensors to effectors (McClelland, 1979; Spivey et al., 2005). Such an account naturally predicts the appearance of these “post-decision” response dynamics, where the purported “decision” event is defined as the point at which motor movement is initiated.

These discoveries of dynamic response characteristics suggest numerous lines of inquiry. One such question, and the focus of the present paper, is the following: What kind of interaction between cognition and action gives way to dynamic post-decision characteristics of responses? In this paper, we employ a neural network model of taxonomic categorization to explore this question. We aim to further elaborate the possible relationship and interaction between cognition and action using a localist attractor network that categorizes animal exemplars.

In what follows, we briefly review research on categorization and the role typicality plays in common accounts of categorization. We then present a neural network simulation that permits comparison of different relationships between internal processing of animal exemplars, and an output component that models an explicit response. We explore this by investigating a parameter space that specifies the network’s decision-action relationship. Network conditions that vary this relationship are compared in how well they fit with human experiments that show these graded action dynamics.

Categorization and Typicality

In the current study, we make use of the graded nature of category structure: Categories, whatever their origin, have members that lie along typicality gradients. For example, an animal can be more or less typical as a member of an animal category, such as of mammals. The members of the mammal category thus reveal a typicality gradient, with cats being typical, and whales being considerably less typical. This results in a variety of experimental effects. Typical members are recognized faster, more consistently, have many features in common, and can facilitate language comprehension (see Murphy, 2002, for an excellent review).

Dale et al. (in press) make use of this property of category structure to uncover post-decision response dynamics in human participants. Computer-mouse trajectories were

recorded during a simple categorization task. Participants categorized an animal exemplar by clicking the mouse on one of two category choices. Mouse-movement trajectories consisted of a movement from the bottom-center of the screen, to the correct target on the left or right (beside which was a competing category label). Target trials used atypical animals (e.g., *whale*) and included an incorrect competitor category that had considerable overlap in terms of semantic and visual features (e.g., *fish*). Though participants responded by clicking the appropriate category (*mammal*), mouse-movement trajectories exhibited substantial attraction toward the competitor category.

This taxonomic categorization task may be a particularly good cognitive process to continue exploring these properties of motor output. While many previous studies of action involve very simple processes (e.g., cued fixation or reaching, Sheliga et al., 1994; Tipper et al., 1997), categorization is a relatively “higher-order” process, involving more processing of exemplars prior to the manual response. More importantly for the present paper, theories of category structure have often made use of feature-based explanations. With very few exceptions, theories of categorization have proposed evaluation of object features, and in some manner comparing these features to prototypes

or exemplars in memory (Medin, 1989). This aspect of categorization theories makes this cognitive process amenable to constraint-based neural network architectures.

Before setting forth the current simulation, it is important to note that there exist a number of quantitative models for categorization, including basic retrieval and recognition mechanisms (e.g., Nosofsky, 1987), category learning and organization (e.g., Love, Medin, & Gureckis, 2004), and even typicality (e.g., Hirahara & Nagano, 2003) and the time course of categorization (Nosofsky & Palmeri, 1997; Lamberts, 2000). Despite such a broad range of models and their application, none makes a distinction between the decision mechanism and a component that implements the overt response. For example, Nosofsky and Palmeri (1997) and Lamberts (2000) offer prominent models for the time course of the *decision* leading up to the response, but not the dynamics of the response itself. We make use of a relatively transparent neural network model that facilitates the inclusion of a motor-response component – the response dynamics of the model is a natural extension of the architecture of the system itself. The model includes decision and response components that are very similar in their operation, permitting manipulation of parameters to explore the relationship between the categorization process, and the response dynamics it generates.

A second important point about the model is that it finesses some debate about the nature of categorization. In particular, exemplar-based and prototype-based theories of categorization are in ongoing competition (e.g., recently, Smith & Minda, 2000). We avoid these issues, and for simplification assume a basic prototype for each animal category used: Mammals, fish, etc., are assumed to have ideal, familiar, frequent, etc. features that define a prototypical exemplar. While this simplifies the presentation of the model here, the way we integrate a response component in the model may be extended to previous models discussed above.

Simulation

Network Architecture

We use normalized-recurrence to simulate the time course of categorization. This hand-coded localist attractor architecture has been used to model a range of cognitive processes, including phoneme perception, spoken-word recognition, online sentence processing, and in modeling the time course of visual search (Spivey & Dale, 2004; Spivey et al., 2005). This architecture is *localist* because individual units stand for specific features of the animal exemplars. For example, when having the network categorize *cat*, one unit in the network might be active to represent the feature representing the animal’s habitat, <land>. The architecture is an *attractor* network because iterated updates of its unit activations lead the network towards a stable state. By applying a set of activation-update rules, the activations of the network’s units are expected to asymptote on particular

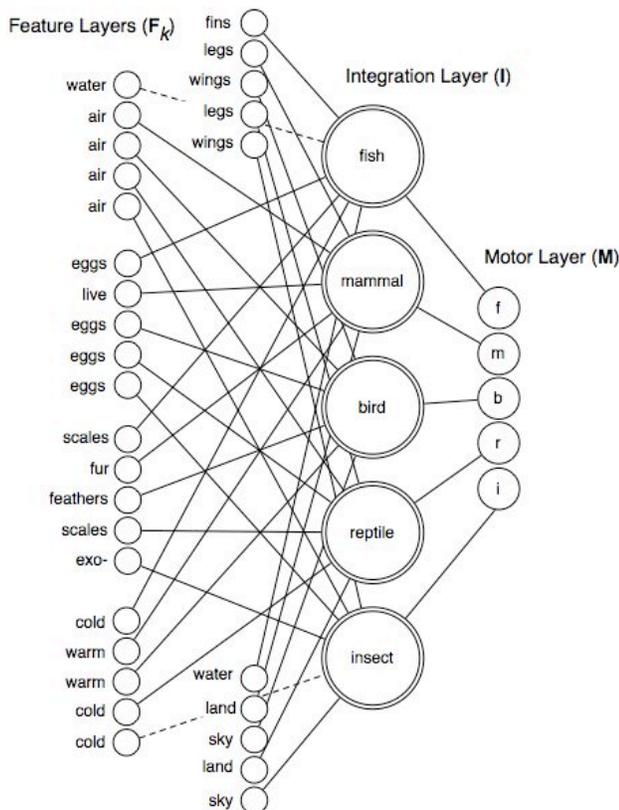


Figure 1: The hand-coded normalized-recurrence model for categorization, with 6 banks of 5 features feeding into the integration (category) layer.

values, usually with one unit obtaining maximal activation, and competing units approaching minimal activation.

The network functions according to two basic computational principles. First, multiple feature units simultaneously constrain the network's behavior. Second, this parallel feature processing is integrated in a layer of units representing alternative outcomes of the model, in which one unit achieves maximal activation over time. Fig. 1 presents the current model. It consists of an array of feature layers, in which each unit represents a particular property of animal exemplars. These layers feed into an integration layer, in which the outcome of categorization is assessed by iterated parallel processing of the constraints.

To further illustrate how the model captures the time course of categorization, we can present an exemplar to the model by setting the appropriate features in the layers to an activation of 1, and all those not relevant to 0. For example, when setting features representing the exemplar *whale*, the values for LIMBS features would have 1 for <fins>, and 0 for all other units. The activity of all these layers serves as input at the integration layer by taking a sum over all relevant nodes. For the processing of *whale* features, the net input to the *mammal* node would be the sum of the activation levels of the units representing mammalian features – in this case, 0 for <legs>, but 1 for <air> as source of oxygen, and so on. When processing *whale*, the *fish* node will thus receive some net input from the <fins> feature unit.

The integration layer is then updated by normalization: The units are made to sum to 1 all together by dividing these net input values by the total net input to the integration layer across all units. This normalized activation then feeds back into the feature layers, these are then normalized, and the cycle is continued until the integration layer becomes stable. Fig. 2 represents such a simulation run for *whale*, in which repeated iteration results in stable and maximal activation of the unit for *mammal* in the integration layer.

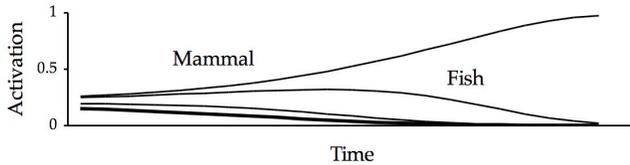


Figure 2: Integration node activation when categorizing *whale*.

Formal presentation of this model's functioning is straightforward. Net activation into the i -th integration unit is the total sum of its relevant feature units across all the feature layers (the i -th unit in each layer). From this net input, the output from this integration unit is obtained by dividing this input activation by the total sum of activation into all integration units. With F_k representing the k -th feature layer in the set of layers F , and I the integration layer,

$$net_{I,i,t} = \sum_F a_{F_k,i,t-1} \quad a_{I,i,t} = net_{I,i,t} / \sum_I net_{I,j,t}$$

where $net_{i,t}$ and $a_{i,t}$ represent net input and activation of the i -th unit of the given layer at time step t . The value $net_{1,2,t}$ is the net input to the second unit of the integration layer, receiving input from the second feature unit of all 6 layers of F . The subsequent activation of a feature unit is determined by the sum of its previous activation and that activation multiplied by the corresponding input from the integration layer. Each feature layer then outputs a normalized activation, as in the integration layer. This process continues until one unit in the integration layer (or motor layer) reaches criterion activation (see below).

$$net_{F_k,i,t} = a_{F_k,i,t-1} + a_{I,i,t} \cdot a_{F_k,i,t-1}$$

$$a_{F_k,i,t} = net_{F_k,i,t} / \sum_{F_k} net_{F_k,j,t}$$

There are a number of benefits to this simple architecture. First, its inner-workings are directly scrutable. While it is important to seek scaled-up systems that fit data from categorization (e.g., Love et al., 2004), simple models that capture core theoretical principles may serve as explicit and transparent accounts for basic patterns of behavioral data. One such core theoretical principle, a second benefit of this model, is conceiving cognitive processes as subject to simultaneous informational constraints. This constraint-based approach to the time-course of categorization fits with perspectives on related cognitive phenomena, particularly semantic phenomena (e.g, among many, McRae, de Sa, & Seidenberg, 1997). Finally, this transparency and theoretical property are implemented in a simple system that operates through iterated updates of its activations. This makes normalized-recurrence particularly suitable for capturing temporal properties of cognition, an important goal for models of a cognitive process such as categorization (Nosofsky & Palmeri, 1997).

In order to map this categorization model onto a simulated response, we supplement it with an additional bank of nodes that receives input from the integration layer. This "motor" bank of units, denoted M , represents the outcome response, and also exhibits a time course over iterations of the model (see Fig. 1). Just as the integration layer approaches a stable decision through input from feature layers, the motor layer does so through input from the integration layer. At a given time t_α , some number of time steps over which integration and feature layers have interacted, these M units begin to receive activation from the integration layer in a manner similar to how the feature layers receive integration activation.

$$net_{M,i,t} = a_{M,i,t-1} + a_{I,i,t_\alpha+t-1} \cdot a_{M,i,t}$$

The activation of the i -th unit $a_{M,i,t}$ is then similarly normalized. The parameter t_α may be varied, allowing M to receive input from the categorization decision at different points in time. We choose to vary this parameter by

observing the “confidence” of the categorization decision – the maximal activation found in the integration layer, denoted here as α . The integration and feature layers may therefore be permitted to interact for a period of time (up to iteration t_α) before \mathbf{M} receives input from the integration layer when the maximal activation in the integration layer is α or higher.

As outlined further below, this allows a number of parameters specifying the interaction between \mathbf{I} and \mathbf{M} to be modified. Firstly, the point in time at which \mathbf{M} begins to receive input from \mathbf{I} , and begins to update its activation towards a stable output response, can be modified. Secondly, \mathbf{M} may be included in the set of layers \mathbf{F} that feed into the integration layer. This allows the process of categorization to be influenced by available responses in the task.

Procedure

In this simulation, we compared three different initial network conditions. 3 *atypical-competitive* conditions involved initializing a non-prototypical exemplar (*whale, penguin, bat*), and a featurally-similar competing response possibility (*fish, fish, bird*). 9 *atypical-noncompetitive* conditions were different only in activating a less saliently competing response unit (e.g., *bird* for *whale*). Finally, 8 composed a *typical* condition, (proto)typical feature values were activated, with a randomly selected competing response option. Each trial in the conditions was selected to have some overlap with Dale et al. (in press), in which human participants were subjected to similar trial types.

The beginning of a simulation run involved setting feature layer units to their relevant values, and turning the two relevant response units to .5 to have equibaised initial response options.¹ According to the equations above, activation then feeds into the integration layer, and back, until the motor layer reached criterion activation (.95).

As mentioned above, the parameters of this attractor network may be modified so as to vary the relationship between the categorization units in \mathbf{I} and the output units \mathbf{M} . Two separate sets of runs using these network conditions were conducted to explore this relationship. Each of these conditions is outlined individually below.

Varying α . The point in time at which activation flows from \mathbf{I} to \mathbf{M} can also be varied. This can be accomplished by varying the confidence threshold (α) at which the integration layer begins to influence the motor layer. As mentioned, t_α is defined as the point at which one integration layer unit has activation of α or greater in each subsequent iteration. We do this while maintaining the non-integral activation values in layer \mathbf{I} . We chose three levels of threshold α that span a reasonable range of possible activation of the

integration units: 0, .4, and .8. This parameter will reveal in the model the amount of processing that may lead to or diminish dynamic post-decision dynamics in the activation of \mathbf{M} . We hold the interaction parameter (β ; see next section) constant at 1 while modifying t_α .

Varying Interaction between \mathbf{I} and \mathbf{M} . Finally, we examine whether interaction between integration and motor layers contributes to graded output in the model, and as observed in human experiments. To do this, we add a term to the net input to unit i in \mathbf{I} :

$$net_{I,i,t} = \sum_F^k a_{F_k,i,t-1} + \beta \cdot a_{M,i,t-1}$$

This permits activation in \mathbf{M} , set at the beginning of the simulation run, to impinge on the time course of the categorization decision – activation in \mathbf{I} . The interaction parameter β permits variable interaction from \mathbf{M} to \mathbf{I} , and we use a broad range of values again: 0, .5, and 1. We hold t_α constant using a threshold of $\alpha = 0$.

Results

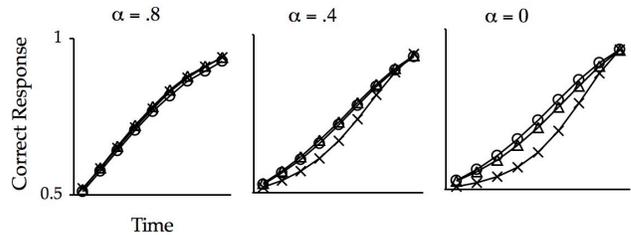


Figure 3: Dynamic profile of response activation (\mathbf{M}) when modifying the point at which categorization flows into output (cross = atypical-competitive; triangle = atypical-noncompetitive; circle = typical).

Varying α . All run conditions were normalized into 10 time steps so they could be overlaid (Spivey et al., 2005; Dale et al., in press). Figures present the iteration after competing response units are set at .5, showing the subsequent 9 normalized time steps before the motor units reach criterion of .95 or greater. Fig. 3 presents the effect of increasing or decreasing the threshold at which \mathbf{I} begins to feed into \mathbf{M} . The normalized time course of run conditions becomes more similar as this threshold increases. The earlier the integration layer feeds into motor, the greater the competition effect seen in the atypical-competitive activation change. As an additional check of this pattern, we ran two further parameter values having t_α established at integration thresholds 0, .2, .4, .6, and .8. The average difference between atypical-competitive and typical trajectories in the three middle time steps diminishes significantly as this parameter is increased ($r = -.97, p < .005$).

Varying Interaction between \mathbf{I} and \mathbf{M} . Fig. 4 shows that feedback from \mathbf{M} into \mathbf{I} is not required to generate the

¹ To match human experiments in Dale et al. (in press), where parameters permitted, activation flowed in for 3 time steps from response options (since these were shown first to subjects).

graded patterns in the atypical-competitive condition. However, increasing feedback from **M** generates a slight competition in the atypical-noncompetitive runs. Again, we ran another two parameter values ($\beta = .25$ and $.75$), and tracked the average difference between the middle three time steps for atypical-noncompetitive and typical conditions. This difference increases significantly with β ($r = .98, p < .005$). In other words, there emerges a disparity in the output dynamics of typical and atypical-noncompetitive runs when one allows motor and integration layers to more fluidly interact.

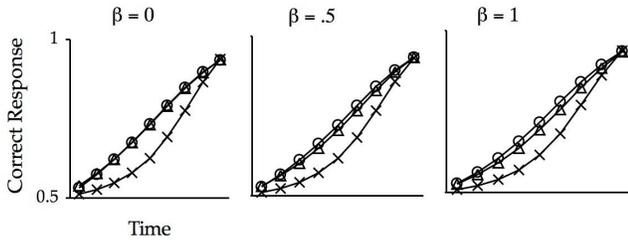


Figure 4: Dynamic profile of correct response (**M**) when modifying the feedback it receives from output units.

Mapping to Human Data. To investigate the relationship between the simulation and human data, we analyzed x-coordinates in manual trajectories drawn from Dale et al. (in press). A number of measures may be used, such as y-coordinate and Euclidean distance to target. However, the x-coordinate more closely represents proximity to correct or incorrect target (since, in the experiment, y-coordinates are the same for each category response label). In addition, differences in x-coordinates have served as the basis for comparing response dynamics in previous work (Spivey et al., 2005; Dale et al., in press).

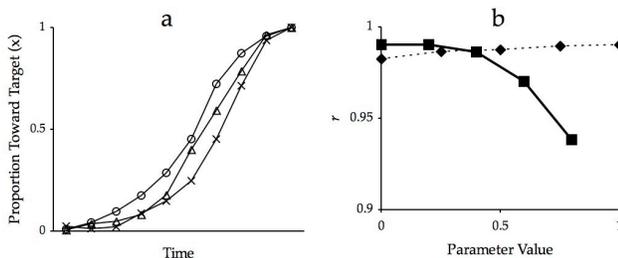


Figure 5a: Dynamic profile of x-coordinate change (in proportion to final, correct target) in human experiment. 5b: As the interaction parameter β (dotted line) is increased, a better fit is achieved with normalized human data. As the required confidence threshold α (solid line) is increased, the opposite holds.

Data from the lexical categorization tasks in Dale et al. (in press) are shown in Fig. 5a. This graph depicts similarly normalized time courses of the same trial types in human experiments as the manual response's x-coordinate reaches its target. The figure bears resemblance to network conditions that involve interaction between integration and

motor layers in both directions. In fact, in both perceptual (i.e., pictures) and lexical categorization, Dale et al. (in press) observed curved motor trajectories in trials akin to atypical-noncompetitive network conditions here, along with the more intuitive effect of atypical-competitive conditions. In a correlation of corresponding time steps, these normalized x-coordinate proportions show a strong relationship with **M** activation in the high interaction ($\beta = 1$) and early motor input condition ($\alpha = 0; r = .99, p < .0001$). When conducting the same regression analysis over the other parameter values, one obtains increases in fit in the expected directions (see Fig. 5b). The best fit with human data is obtained when interaction between integration and motor layers is maximized: When β is large (high interaction between **I** and **M**), and α is small (early input to **M**).

General Discussion

Normalized-recurrence provides a simple yet ideal arena for exploring the interaction among constraints that underlie taxonomic categorization. These constraints involve the parallel processing of features, and an accumulation of information that guides the categorization decision and resultant response. This model indeed resembles several recent simulations that also seek to characterize the time course of categorization (e.g., Lamberts, 2000). While the current model is simple, it offers some explicit insight into the possible relationship between the categorization decision, and the resultant response itself. In the model, patterns of competition observed in human data are obtained when both the decision process and the motor output concurrently and continuously interact. The kind of approach used here may suggest ways in which existing models (e.g., Nosofsky & Palmeri, 1997; Lamberts, 2000) might integrate action parameters, thereby providing a fuller picture of the cognitive process: Accounting for not just the hypothesized internal processes, but information accumulation all the way into the observable response behavior and its time course. For example, the Nosofsky and Palmeri (1997) model involves a random-walk process where exemplars race each other to help categorize a test exemplar. Including a motor component may simply involve integrating a second random-walk process in which possible category responses race each other for selection. Interactive parameters could then be similarly explored.

One possible promissory note about such models is a synthesis of “pre-decision” process models, and the experimentally observed relationship between process and response characteristics. Growing research on the embodiment of cognition (e.g., Barsalou, 1999) shows that parameters of action can impinge on the cognitive processing eventually leading into it (e.g., Glenberg & Kaschak, 2002). The current model provides some simple means by which computational mechanisms of these effects can be devised.

The current model has some obvious limitations. Firstly, it is extremely simple. The model simply embodies basic

computational principles regarding constraint-based feature processing, and the continuous integration of this information. However, it is perhaps surprising that such simple assumptions can succeed in generating the kind of response patterns observed in the human experiments: Both competitive and noncompetitive trials and their graded response characteristics can be captured in this system.

Secondly, as discussed earlier, the model does not seek to resolve current debate regarding opposing theories of the categorization process. While the model may prima facie have difficulty fitting knowledge-based results of categorization, its properties serve as an exploration of the time course of categorization given certain conditions. Those assumed here are at least sufficient to model basic categorization tasks. While this is promising, others seek to scale up computational models of categorization and category learning (Love et al., 2004). The current model may again provide some motivation for integrating output-based dynamics to model responses.

Thirdly, the model's parameters clearly cannot be directly mapped to neurophysiology. Nevertheless, substantial evidence from the neurophysiology of the motor system reveals both simultaneous and continuous competition between possible actions – properties similar to this paper's normalized-recurrence network (Spivey & Dale, 2004, for some review).

Despite these limitations, rather than intricately capturing categorization and its underlying neural substrate, the model provides an idealized computational system that explores the effects of two interacting systems. One system acts through informational constraints to gradually decide on a relevant category; another acts through this decision component at a given later time to mediate between only two possible stable states (the correct vs. incorrect end-point actions). In this model's context, fluid interaction between these components is what generates the response dynamics observed in human experiments.

Acknowledgments

The author would like to thank Morten Christiansen and Michael Spivey for valuable input and criticism.

References

Balota, D., & Abrams, R. (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1289-1302.

Becker, W. (1991). Saccades. In R. H. S. Carpenter (Ed.), *Eye Movements, Vol. 8 in Vision and visual Dysfunction*. CRC Press.

Dale, R., Kehoe, C., & Spivey, M. (in press). Graded motor responses in the time course of categorizing atypical exemplars. *Memory and Cognition*.

Fodor, J.A. (1983). *Modularity of mind*. Cambridge, MA: Bradford.

Glenberg, A., & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9, 558-565.

Hirahara, M. & Nagano, T. (2003). A neural network for the typicality effects. *Proceedings of the International Joint Conference on Neural Networks*, 4, 2502-2505.

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, 107, 227-260.

Love, B., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, 111, 309-332.

Mattes, S., Ulrich, R., & Miller, J. (2002). Response force in RT tasks: Isolating effects of stimulus probability and response probability. *Visual Cognition*, 9, 477-501.

McClelland, J. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86, 287-330.

McRae, K., de Sa, V.R., & Seidenberg, M.S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, 126, 99-130.

Medin, D. L. (1989). Concepts and concept structure. *American Psychologist*, 44, 1469-1481.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 13, 87-108.

Nosofsky, R. M., & Palmeri, T.J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 104, 266-300.

Posner, M. L., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13, 25-42.

Pylyshyn, Z. (2000). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22, 341-365.

Sheliga, B., Craighero, L., Riggio, L. & Rizzolatti, G. (1997). Effects of spatial attention on directional manual and ocular responses. *Experimental Brain Research*, 114, 339-351.

Smith, J., & Minda, J. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: LMC*, 26, 3-27.

Spivey, M.J., & Dale, R. (2004). The continuity of mind: Toward a dynamical account of cognition. In B. Ross (Ed.), *Psychology of Learning and Motivation* v.45 (pp. 85-142). Elsevier Academic Press.

Spivey, M., J. Grosjean, M., Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, 102, 10393-10398.

Tipper, S., Howard, L., & Jackson, S. (1997). Selective reaching to grasp: Evidence for distractor interference effects. *Visual Cognition*, 4, 1-38.