Title
Control of Response Initiation: Mechanisms of Adaptation to Recent Experience

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 26(26)

ISSN
1069-7977

Authors
Mozer, Michael C.
Kinoshita, Shachiko
Davis, Colin J.

Publication Date
2004

Peer reviewed

Control of Response Initiation: Mechanisms of Adaptation to Recent Experience

Michael C. Mozer
Department of Computer Science &
Institute of Cognitive Science
University of Colorado

Sachiko Kinoshita
MACCS &
Department of Psychology
Macquarie University

Colin J. Davis
MACCS
Macquarie University
Sydney, NSW 2109

Abstract

In most cognitive and motor tasks, speed-accuracy trade offs are observed: Individuals can respond slowly and accurately, or quickly yet be prone to errors. Control mechanisms governing the initiation of behavioral responses are sensitive not only to task instructions and the stimulus being processed, but also to the recent stimulus history: when stimuli can be characterized on an easy-hard dimension (e.g., word frequency in a naming task), an easy item is responded to more slowly when intermixed with hard items than when presented among other easy items; likewise, hard items are responded to more quickly when intermixed with easy items. We propose a mathematical theory with three components: a model of temporal dynamics of information processing, a decision criterion specifying when a response should be initiated, and a mechanism of adaptation to the stimulus history. Performance during the course of an experimental trial is cast in terms of a utility function that increases with accuracy and decreases with response time. We assume a decision criterion that initiates a response at the point in time that maximizes expected utility. We posit that the effect of the stimulus history arises because information concerning recent trial difficulty is incorporated into the utility estimate. We present further behavioral studies to validate predictions of the theory.

Consider a simple task in which you are asked to name the sum of two numbers, such as 14+8. Given sufficient time, you presumably produce the correct result; however, under speed pressure, mistakes can occur. In most all cognitive and motor tasks, such empirical *speed-accuracy trade offs* are observed: Individuals can respond slowly yet accurately, or quickly and be prone to errors. Speed-accuracy trade offs are due to the fact that evidence accumulates gradually in response systems over time (Rabbitt & Vyas, 1970). Responses initiated earlier in time will be based on lower quality information, and hence more likely to be incorrect. This paper addresses a simple yet fundamental form of cognitive control—the mechanism that governs the initiation of a behavioral response, and therefore, where an individual operates on the speed-versus-accuracy continuum. In the following section, we describe data that place constraints on the nature of control mechanisms. We describe shortcomings of existing theoretical frameworks that have tried to account for these data. We then present a framework that successfully explains key phenomena and makes further predictions which we have verified through additional behavioral studies.

The Blocking Effect

To understand the control mechanism that initiates responses, consider the variables that affect its operation. The mechanism is influenced by task instructions: individuals can choose to emphasize speed or accuracy. The mechanism is also influenced by recent performance: participants often slow down after producing an error (Rabbit & Vyas, 1970). Finally, even in the absence of errors, the mechanism is sensitive to the recent stimulus environment (Kiger & Glass, 1981): when items are presented in a sequence or *block*, reaction time (*RT*) and error rate to an item depends on the immediately preceding items.

This *blocking effect* is generally studied by manipulating item difficulty. Some items are intrinsically easier than others, e.g., 10+3 is easier than 5+8, whether due to practice or the number of cognitive operations required to determine the sum. By definition, individuals have faster RTs *and* lower error rates to easy problems. However, the RTs and error rates are modulated by the composition of a block. Consider an experimental paradigm consisting of three trial blocks: just easy items (*pure easy*), just hard items (*pure hard*), and a mixture of both in random order (*mixed*). When presented in a mixed block, easy items slow down relative to a pure block and hard items speed up. Thus, the control mechanism that initiates responses uses information not only from the current stimulus, but also adapts to the stimulus environment in which it is operating. Table shows a typical blocking result for a word reading task, where word frequency is used to manipulate difficulty. Based on our review of the blocking-effect literature (e.g., Lupker, Brown & Columbo, 1997; Lupker, Kinoshita, Coltheart, & Taylor, 2000; Taylor & Lupker, 2001), we summarize the central, robust phenomena as follows.

TABLE 1. RTs and Error Rates for Blocking study of Lupker, Brown, & Columbo (1997, Experiment 3)

	Pure Block	Mixed Block	Difference
Easy Item	488 ms (3.6%)	513 ms (1.8%)	+25 ms (–1.8%)
Hard Item	583 ms (12.0%)	559 ms (12.2%)	–24 ms (+0.2%)

(1) Easy items are faster *and* less error prone than hard.
(2) When intermixed, easy items slow down and hard items speed up. However, the convergence of RTs for easy and hard items in a mixed block is not complete. Thus, RT depends both on the stimulus type and the composition of the block.
(3) Speed-accuracy trade offs are observed: a drop in error rate accompanies easy-item slow down; a rise in error rate accompanies hard-item speed up.
(4) Blocking effects occur across diverse paradigms, including naming, arithmetic verification and calculation, target search, and lexical decision. They are obtained when stimulus or response characteristics alternate from trial to trial (Lupker et al., 2000). Thus, the blocking effect is not associated with a specific stimulus or response pathway, but rather is a general phenomenon of response initiation.
(5) Overt responses are necessary for obtaining blocking effects, but overt errors are not.
(6) A signature of the effect concerns the relative magnitudes of easy-item slow down and hard-item speed up. Significantly more speed up than slow down is never observed. The trend is that speed up is less than slow down—indeed, some studies show no reliable speed up—although equal magnitude effects are observed.
(7) The effects of stimulus history are local, i.e., the variance in RT on trial n due to trial $n–k$ decreases rapidly with k. Dependencies for $k>2$ are not reliable (Taylor & Lupker, 2001).

Explanations for the Blocking Effect

The blocking effect demonstrates that the response time depends not only on information accruing from the current stimulus, but also on recent stimuli in the trial history. Therefore, any explanation of the blocking effect must specify how control processes, which determine the point in time at which a response is initiated, are sensitive to the composition of a block. Various mechanisms of control adaptation have been proposed.

Domain specific mechanisms. Most of the proposed mechanisms are domain specific. For example, Rastle and Coltheart (1999) describe a model with two routes to naming, one lexical and one nonlexical, and claim that the composition of a block affects the emphasis that is placed on the output of one route or the other. Meyer, Roelofs, and Levelt (2003) manipulate word length and explain blocking effects in terms of a control process, sensitive to block composition, that decides when to initiate a naming response—either after the motor program for the first syllable has been generated, or after the motor program for the entire word has been generated. Because of the ubiquity of blocking effects across tasks, domain-specific accounts are not compelling. Parsimony is achieved only if the adaptation mechanism is localized to a stage of response initiation common across stimulus-response tasks.

Rate of processing. Kello and Plaut (2003) have proposed a *rate-of-processing* explanation, according to which control processes adjust a gain parameter on units in a dynamical connectionist model. The parameter determines the steepness of the sigmoid curve. Technically, the gain does not affect rate of processing, i.e., it does not simply rescale time. Increasing the gain does result in more rapid convergence, but it also yields a higher error rate; thus the account should more appropriately be framed in terms of adapting the *rate of convergence.* Simulations of this model have explained the basic blocking effect, but not the complete set of phenomena we listed previously. Of greater concern is the fact that the model predicts that naming *duration* decreases with increased speed pressure, which doesn't appear to be true (Damian, 2003; Kinoshita, unpublished).

Evidence criterion. A candidate mechanism with intuitive appeal is the trial-to-trial adjustment of an *evidence criterion*, which specifies the level of evidence that must accumulate in support of a decision before the response is initiated. Random walk and diffusion models have such a parameter, often called the response criterion (Ratcliff, 1978). According to this account, the evidence criterion is determined by recent trial history: if previous trials were easy, the criterion is set low, if previous trials were hard, the criterion is set high. Thus, the criterion would be lowest in a pure-easy block, intermediate in a mixed block, and highest in a pure-hard block. When the criterion is high, RTs are slower but error rates are lower, resulting in slow down of easy items and speed up of hard items in a mixed block.

Taylor and Lupker (2001) illustrate that adaptation of an evidence criterion can—at least in some models—yield incorrect predictions concerning the blocking effect. Strayer and Kramer (1994) attempted to model the blocking effect with an adaptive response criterion in the diffusion model. They managed to fit their blocking data but the account had two fatal shortcomings. First, they allowed different criteria for easy and hard items in a mixed block, which makes no sense because the trial type was not known in advance, and setting differential criteria depends on knowing the trial type. Second, they used a nonstandard blocking paradigm in which the trial difficulty depended on whether an item was presented in a pure or mixed block, easy items being more difficult and hard items being less difficult in a mixed block.

In spite of these problems, we were also convinced an evidence-criterion-adjustment explanation should be feasible. We used a somewhat different model of temporal dynamics and response initiation than Strayer and Kramer (to be described shortly), but like Strayer and Kramer, the model had an adaptive parameter that determined the trade off between speed and accuracy. After six frustrating months of exploration, we admitted defeat: The model was unable to obtain the right qualitative pattern of results; either the blocking effect was an order of magnitude smaller than that observed in experiments, or went in the wrong direction. Several variants of the model came close, but were not robust; tiny changes to parameter values yielded qualitative effects on the pattern of results.

The failure of an evidence-criterion-adjustment account is not surprising from another perspective. On logical grounds, the relative importance of speed versus accuracy should be determined by task instructions and pay offs. Item difficulty is independent and unrelated factor. Consistent with this logical argument is the finding that manipulating instructions to emphasize speed versus accuracy does not produce the same pattern of effects as altering the composition of a block (Dorfman & Glanzer, 1988).

Adaptation to the statistics of the environment. Having ruled out three possible explanations, we sketch a fourth alternative, which is based on the premise that the goal of cognition is optimal and flexible performance across a variety of tasks and environments. In service of this goal, control mechanisms must be sensitive to the statistical structure of the environment, e.g., stimulus characteristics and configurations, response contingencies, etc. Previous models of control have exploited this assumption. For example, Treisman and Williams (1984) and Mozer, Colagrosso, and Huber (2002) considered a sequential choice task involving two response alternatives, and proposed that control mechanisms estimate prior probabilities of the two responses. If one response is more frequent, the larger prior induces a bias toward that response, which typically boosts performance.

Blocking effects can be explained via a related hypothesis. Because RTs depend on whether a trial is easy or hard, the control mechanisms responsible for response initiation must utilize an estimate of the item difficulty, or the quality of information available to response processes. If this estimate is unreliable (noisy), and if control mechanisms make the ecological assumption that the current trial is similar to recent trials, the estimate can be made more reliable by incorporating estimates from recent past trials. We elaborate this idea in a mathematical model of response initiation, and show that it can explain the key blocking phenomena listed earlier as well as other puzzling phenomena.

The ASE Model

We refer to our model as ASE, which stands for *Adaptation to the Statistics of the Environment*. Although the key claim of the model concerns the *mechanism of control adaptation* based on recent experience, we must make two additional sets of assumptions, one set concerning the *temporal dynamics* of information processing, and another set concerning the *decision criterion* for response initiation Although the specific assumptions we make are not critical, they must be made explicitly to fully flesh out the model.

Temporal dynamics. We need a way to characterize the temporal dynamics of information processing in tasks such as naming. The particular model of temporal dynamics is not critical, as long as it has the property that the quality of information available for responding increases gradually and monotonically over time.

We chose the probabilistic information transmission (PIT) model of Mozer, Colagrosso, and Huber (2002, 2003). To summarize the key properties relevant for the current work, the model consists of a cascaded series of processing *pathways* whose details are determined by the task being modeled. For example, to model a word naming task, we use a perceptual pathway that maps visual word forms to an internal semantic/lexical representation, and a response pathway that maps the internal representation to a distinct verbal naming response. Each pathway is a dynamic Bayesian network, and the conditional probability distributions in the model are specified by the nature of the mapping, the state of expertise being modeled, and the similarity structure among elements of representation. Given a stimulus presentation, the output of the model is a probability distribution over response alternatives as a function of time (Figure 1a). The response chosen at a particular time is a sample from the distribution (the model cannot choose the most probable response). The time course of processing depends on information transmission probabilities in the model. Easy, high frequency, and well practiced items have higher transmission probabilities, and hence are conveyed more rapidly.

This model is a generalization of random walk models and has several advantages. It provides a mathematically principled means of handling multiple alternative responses (necessary for naming) and similarity structure among elements of representation, and characterizes perceptual processing, not just decision making. The counter model (Ratcliff & McKoon, 1997) or connectionist integrator models (e.g., Usher & McClelland, 2001) could also serve us, although the PIT framework has an advantage in that it operates using a currency of probabilities—versus more arbitrary units of *count* or

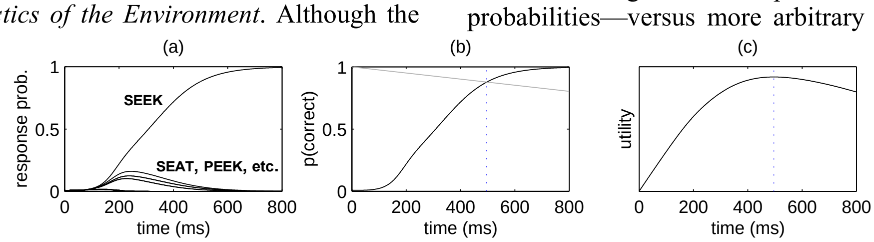

FIGURE 1. (a) Output of the probabilistic information-transmission model for presentation of the stimulus, e.g., SEEK, on a word naming task. (b) Treating the most probable output is an estimate of accuracy (light line is the time threshold). (c) Utility function based on most probable output (dashed line = time of maximum utility)

activation—which leads to explicit, interpretable decision criteria and adaptation mechanisms, and requires fewer additional assumptions to translate model output to predictions of experimental outcomes.

Decision criterion. To model blocking effects, we must make an explicit assumption concerning the decision criterion used for response initiation. A simple speed criterion (i.e., respond at α milliseconds following stimulus onset) or accuracy criterion (i.e., respond when the error rate is below α) is inadequate, because easy items are both faster and more accurate than hard items in pure blocks. Ratcliff's (1978) diffusion model uses an evidence threshold, which effectively yields an accuracy criterion that declines over time. We adopt this notion, as illustrated by the gray line in Figure 1b. The line is characterized by one free parameter, the slope κ. This criterion can be recast in an optimization framework: A response is initiated at the point in time that maximizes *utility*, where utility increases with expected accuracy and decreases with time (Figure 1c). Previous psychological theory has suggested that individuals can choose the optimal point at which to respond (Mozer et al., 2002; Rabbitt & Vyas, 1970; Triesman & Williams, 1984).

To summarize, we propose a theory premised on five key assumptions. (1) Transmission of stimulus information to response systems is gradual and accumulates over time. (2) Control mechanisms respond at the point in time that maximizes a utility measure that depends on both expected accuracy and time. (3) During ongoing processing, the system is able to compute an estimate of its response accuracy for the current stimulus. (4) This estimate is unreliable. (5) If control systems make the ecological assumption that the current trial is similar in difficulty to recent trials, the accuracy estimate can be made more reliable by incorporating estimates from recent trials.

An accuracy estimate can be obtained from the PIT dynamics by assuming that the most probable output at a point in time is correct (Figure 1b); we refer to this as curve as the *current accuracy trace (CAT)*. Given the response criterion (grey line, Figure 1b), a response initiation time can be determined (dashed line).

If the model's transmission probabilities are noisy, the CAT is a high-variance estimate of accuracy, because the assumption that the most probable response state is correct may be wrong. The suggestion of noise is not arbitrary, but rather is a central claim of the diffusion model, and has been key to explaining a variety of RT data. To overcome this noise source, it is sensible for control mechanisms to rely not solely on the CAT, but on accuracy traces from recent trials. We claim that the model maintains a *historical accuracy trace (HAT)*, and the trace used for estimating utility—the *mean accuracy trace (MAT)*—is a weighted average of CAT and HAT, i.e., $\mathrm{HAT}(n) = \lambda\mathrm{CAT}(n-1) + (1-\lambda)\mathrm{HAT}(n-1)$, where n is an index over trials, and $\mathrm{MAT}(n) = \theta\mathrm{CAT}(n) + (1-\theta)\mathrm{HAT}(n)$; λ and θ are averaging weights. Figure 2a depicts the CAT, HAT, and MAT. The two solid curves represent CATs for easy and hard trials, as well as the MATs for pure blocks. The dotted curve represents the expected HAT in a mixed block—an average of easy and hard CATs. The dashed curves represent the MATs for easy and hard trials in a mixed block, formed by averaging the HAT and corresponding CAT. Because the CAT and HAT are time-varying functions, the notion of averaging is ambiguous; possibilities include averaging the accuracy of points with the same time value and times of points with the same accuracy value. It turns out that the choice has no qualitative impact on the simulation results we present. The essential requirement is that the computation to determine response-initiation time can be performed in real time, including identification of the utility maximum.

Modeling Blocking Effects

Figure 2b provides an intuition concerning the model's ability to replicate the basic blocking effect. The mean RT for easy and hard items in a pure block is indicated by the point of intersection of the CAT with the time threshold. The mean RT for easy and hard items in a mixed block is indicated by the point of intersection of the MAT with the time threshold. The easy item slows down, the hard item speeds up. Because the rate of processing is not affected by the blocking manipulation, the error rate will necessarily drop for easy items and rise for hard items. Although the RTs for easy and hard items come together, the convergence is not complete as long as $\theta > 0$. The theory thus explains the first three phenomena of Section 1. The fourth phenomenon, that the effects occur across diverse paradigms, is consistent with the theory: the theory concerns the response curves, but not the stimulus or response modalities or domains that underlie the curves. Consequently, cross-task blocking effects are implied by the theory. The theory is consistent with the observation that blocking effects occur even in the absence of overt errors, because the theory is neutral with regard to error production, and only if response mechanisms are engaged (phenomenon 5). If responses are not produced, the

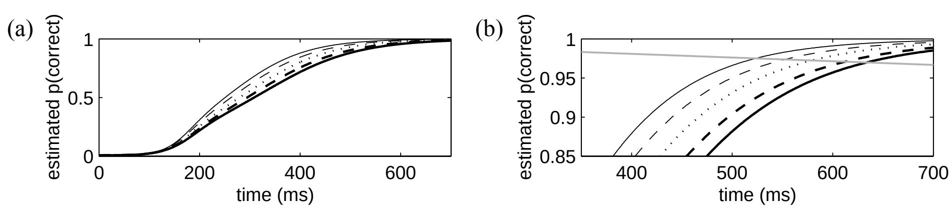

FIGURE 2. (a) easy and hard CAT (solid), mixed block HAT (dotted), and easy and hard MAT (dashed); (b) close up of the traces, along with the decision threshold

TABLE 2. Expt. 1 of Taylor & Lupker (2001): Human data and simulation

	Human Data			Simulation		
	Pure	Mixed	Difference	Pure	Mixed	Difference
Easy	519 ms (0.6%)	548 ms (0.7%)	29 ms (0.1%)	524 ms (2.4%)	555 ms (1.7%)	31 ms (–0.7%)
Hard	631 ms (2.9%)	610 ms (2.9%)	–21 ms (0.0%)	634 ms (3.0%)	613 ms (3.7%)	–21 ms (0.7%)

TABLE 3. Context experiment: Human data and simulation

	Human Data			Simulation		
	Same Context	Diff. Context	Switch Effect	Same Context	Diff. Context	Switch Effect
Easy	432 ms	488 ms	56 ms	437 ms	493 ms	56 ms
Hard	514 ms	467 ms	–47 ms	514 ms	470 ms	–44 ms

response-accuracy curves need not be generated, and the averaging process that underlies the effect cannot occur.

The fact that hard-item speed up is never greater than easy-item slow down (phenomenon 6) turns out to be a key diagnostic. Our initial candidate models tended to yield more speed up than slow down because the magnitude of RT change was proportional to the RT, and hard RTs are larger than easy RTs. Empirically, the error-averaging model we propose never yields more speed up than slow down. As shown in Figure 2b, the mixed-block MAT (dashed) hugs the pure-block MAT (solid) more tightly for hard than easy items. The asymmetry is due to the fact that the easy CAT reaches asymptote before the hard CAT. The model produces more symmetric blocking effects when responses are initiated at a point where both easy and hard CATs are ascending at the same rate (leading to high error rates, unlike the behavioral data). However, we were unable to find model parameters that produced the invalid pattern of more speed up than slow down.

Beyond providing qualitative explanations for key phenomena, the model fits specific experimental data. Taylor and Lupker (2001, Expt. 1) instructed participants to name high frequency words (easy items) and nonwords (hard items). Table 2 compares mean RTs and error rates for human participants and the simulation. One should not be concerned with the error-rate fit, because measuring errors in a naming task is difficult and subjective. (Over many experiments, error rates show a speed-accuracy trade off.) Taylor and Lupker further analyzed RTs in the mixed block conditional on the context—the 0, 1, and 2 preceding items. Figure 3 shows the RTs conditional on context. The model's fit is excellent. Trial *n* is most influenced by trial *n–1*, but trial *n–2* modulates behavior as well; this is well modeled by the exponentially decaying HAT.

Simulation details. Parameters of the PIT model were chosen to obtain pure-block mean RTs comparable to those obtained in the experiment and asymptotic accuracy of 100% for both easy and hard items. We added noise to the transmission rates to model item-to-item and trial-to-trial variability, but found that this did not affect the expected RTs and error rates. We fixed the HAT and MAT averaging terms, λ and θ, at 0.5, and picked κ to obtain error rates in the pure block of the right order. Thus, the degrees of freedom at our disposal were used for fitting pure block performance; the mixed block performance (Figure 3) emerged from the model.

Asymptotic Effect of Context

In the standard blocking paradigm, the target item is preceded by a context in which roughly half the items are of a different difficulty level. We conducted a behavioral study in which the context was maximally different from the target. Each target was preceded by a context of ten items of homogeneous difficulty, either the *same* or *different* difficulty as the target. This study allows us to examine the asymptotic effect of context switching. We performed this study for two reasons. First, Taylor and Lupker (2001) obtained results suggesting that a trial was influenced by only the previous two trials; our model predicts a cumulative effect of all context, but diminishing exponentially with lag. Second, several candidate models we explored predict that with a strong context, speed up of hard is significantly larger than slow down of easy; the model we've described does not.

The results are presented in Table 3. The model provides an excellent fit to the data. Significantly larger context effects are obtained than in the previous simulation (~50 ms in contrast to ~25 ms), and—given the strong context—the easy items become slower than the

FIGURE 3. RTs from human subjects (black) and simulation (white) for easy and hard items in mixed block, conditional on 0, 1, and 2 previous item types. Last letter in a string indicates the current trial and first letters indicate context. Thus, "EHH" means a hard item preceded by another hard item preceded by an easy item.

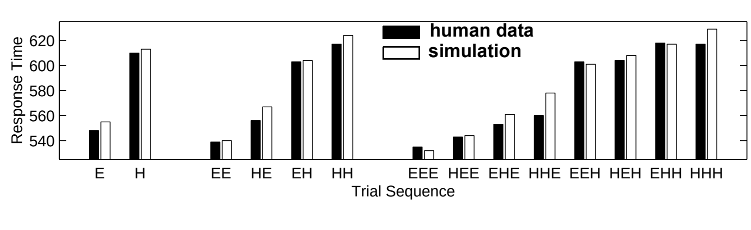

TABLE 4. Simulation of validity-modulating masked priming effect

	Repetition-Prime Trial	Unrelated-Prime Trial	Priming Effect
20% valid	560 ms	585 ms	25 ms
80% valid	515 ms	580 ms	65 ms

hard (although this effect is not statistically reliable in the experimental data). Further, both data and model show more slow down that speed up, a result that allowed us to eliminate several competing models. For this simulation, we fit parameters of the PIT model to the same-context results. We also treated the MAT averaging constant, θ, as a free parameter on the rational argument that this parameter can be tuned to optimize performance: if there is not much variability among items in a block, there should be more benefit to suppressing noise in the CAT using the HAT, and hence θ should be smaller. We used 0.35 for this simulation, in contrast to 0.5 for the first simulation.

Reinterpreting Other Experimental Findings

In many studies, contrasts are made between experimental blocks whose composition varies in terms of the proportion of easy and hard items. In such cases, our model may provide an alternative interpretation of experimental results. Consider a subliminal priming study in which participants are asked to perform lexical decision on a target string preceded by a masked prime (Bodner & Masson, 2001). The prime and target could be identical or unrelated. Although the prime was subliminal—not accessible for report—a repetition priming effect is observed: lexical decision RT to a target is faster if the prime is identical to the target. Subliminal repetition priming effects are common in the literature, but what is surprising in this study is that *prime validity* influences priming: The priming effect is larger when the prime and target are identical on 80% of trials (high validity) than when they are identical on 20% of trials (low validity). Bodner and Masson suggest that "recruitment of the prime resource to assist target processing should be more likely when the...prime validity...is higher." (p. 618). This counterintuitive explanation implies that the prime is analyzed deeply: its match to the target is determined, prime validity is estimated, and the estimate is available for strategic control.

Our model offers an alternative account. The repetition prime makes a trial easy because the prime activation supports the target, and the unrelated prime makes a trial relatively hard. Low and high validity conditions are thus mixed blocks containing 20% and 80% easy trials, respectively. We ran a simulation to show that these mixtures yield a blocking effect consistent with the reduction of priming in the low validity condition (Table 4). In the model, the prime influences the time course of information transmission, which modulates the model's response-initiation criterion on future trials—a simpler, more elegant account than Bodner and Masson's.

Conclusions

Theories in cognitive science occasionally hand the problem of control to a homunculus. More commonly, control processes are left unspecified. And when implemented, control generally involves explicit, active, and sophisticated mechanisms. We have described a model that achieves an interesting sort of control—sequential adaptation of the speed-accuracy trade off. However, the mechanism that gives rise to this adaptation is passive and in a sense dumb; it essentially reestimates the statistical structure of the environment by updating an expectation of task difficulty. Our hope and belief is that many aspects of cognitive control can be explained away by such simple, passive mechanisms, eventually eliminating the homunculus from cognitive science.

REFERENCES

Bodner, GE, & Masson, ME (2001). Prime validity affects masked repetition priming: Evidence for an episodic resource account of priming. *Journal of Memory & Language, 45,* 616–647.

Damian, MF (2003). Articulatory duration in single word speech production. JEP: LMC, *29*, 416-431.

Dorfman, D., & Glanzer, M. (1988). List composition effects in lexical decision and recognition memory. *J. Mem. & Lang.*, 27, 633–648.

Kello, CT & Plaut, DC (2003). Strategic control over rate of processing in word reading: A computational investigation. *Journal of Memory and Language*, *48*, 207-232.

Kiger, JI, & Glass, AL (1981). Context effects in sentence verification. *JEP:HPP*, *7*, 688-700.

Lupker, SJ, Brown, P, & Colombo, L (1997). Strategic control in a naming task: Changing routes or changing deadlines? *JEP:LMC*, *23*, 570–590.

Lupker, S, Kinoshita, S, Taylor, T, & Coltheart, M. (Nov. 2000) Is a time criterion used when naming pictures and computing sums? *Annual meeting of the Psychonomic Society*, New Orleans.

Meyer, AS, Roelofs, A, & Levelt WJM (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language*, *48*, 131–147.

Mozer, MC, Colagrosso, MD, & Huber, DE (2002). A rational analysis of cognitive control in a speeded discrimination task. In NIPS XIV (pp. 51-57). Cambridge, MA: MIT Press.

Mozer, MC, Colagrosso, MD, & Huber, DE (2003). Mechanisms of long-term repetition priming and skill refinement: A probabilistic pathway model. In *Proceedings of the Twenty Fifth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum Assoc.

Rabbitt, PMA, & Vyas, SM (1970). An elementary preliminary taxonomy for some errors in laboratory choice RT tasks. *Acta Psych.*, *33*, 56–76.

Ratcliff, R. A theory of memory retrieval. *Psych. Rev.*, 1978, 85, 59-108.

Ratcliff, R., & McKoon, G. (1997). A counter model for implicit priming in perceptual word identification. *Psych. Review*, *104*, 319–343.

Taylor, TE, & Lupker, SJ (2001). Sequential effects in naming: A time-criterion account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 117–138.

Treisman, M & Williams, TC (1984). A theory of criterion setting with an application to sequential dependencies. *Psych. Review*, *91*, 68–111.

Usher, M, & McClelland, JL (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, *108*, 550-592.