Yoked Criteria Shifts in Decision System Adaptation: Computational and Behavioral Investigations

Blair C. Armstrong (blairarm@andrew.cmu.edu)
Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213 USA

Steve Joordens (joordens@psych.utoronto.ca)
Department of Psychology and the Centre for Computational Cognitive Neuroscience, University of Toronto Scarborough
1265 Military Trail, Toronto, ON M1C 1A4 Canada

David C. Plaut (plaut@cmu.edu)
Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213 USA

Abstract

We describe a theory of decision system adaptation in which yoked criteria shifts serve as a simple but powerful mechanism for rapidly minimizing errors without sacrificing speed. To support our theory, we implemented a connectionist model of lexical decision, wherein the state of a word perception network was "read" by a pair of decision units. The response criteria for these decision units were then subjected to yoked shifts to examine how, in the face of perceived errors, such a response mechanism might adjust performance. We also present the results of a lexical decision experiment that manipulated the truthfulness of the feedback participants received so as to trigger the error correction mechanism while keeping other task parameters constant. The results of the experiment largely parallel those of the simulation, suggesting that yoked decision shifts make an important contribution to error minimization in decision system adaptation.

Keywords: decision making, decision system adaptation, yoked criteria shifts, lexical decision, connectionist modeling.

An individual's ability to rapidly and correctly decide between two alternatives is critical to their survival and wellbeing. For example, a new driver may learn to brake or accelerate when faced with a yellow light under dry road conditions. However, if one day it snows their established decision behavior will need to be adapted to accommodate this fact. Thus, the driver must be capable both of deriving an initial calibration of their decision system, and of rapidly adapting this system in face of change.

The work reported in this paper focuses on how the updating of a previously well-calibrated decision system is accomplished. Motivating our work is an interesting pattern of effects reported by Gomez, Ratcliff, and Perea (2007), who varied the task participants completed (two-choice lexical decision vs. go/no go lexical decision) in a within-subjects design. After fitting their data with a diffusion model, the authors determined that changes of the decision criteria were key in accounting for performance differences across task blocks, with adjustments to most other parameters only having a modest effect.

Within the context of decision system adaptation, Gomez et al.'s (2007) findings might also suggest that rather than re-configure the system *de novo* when faced with different task demands, participants may adapt to the new task primarily by shifting their decision criteria. In many cases, such a shift may provide a rapid means of accommodating modest (and perhaps not so modest) changes without necessitating a potentially costly re-derivation of all of the parameters in the decision system.

If we assume that decision system adaptation occurs via shifts of decision criteria, this raises the question of how the criteria should be shifted. We examined this issue within the context of an abstract forced choice task wherein participants must make rapid and accurate *A* and *B* responses indicating the presence of stimuli *a* and *b*. Imagine that after becoming proficient at the task, some property of the task changes such that participants' find themselves incorrectly responding *A* to *b* items (e.g., all the *b*'s become more *a*-like). One logical adaptation would be to shift the *A* decision criterion so as to require the additional accumulation of evidence that an *a* was presented before making an *A* decision. However, such an adaptation would have the result of slowing overall reaction time (RT) because the average amount of evidence to be collected before making a particular decision will have increased—an effect which may not be adaptive if there are benefits associated with being fast.

The issue of overall increases in RT can be avoided, however, if both criteria are shifted in unison, such that when more evidence is required to make an *A* response, less evidence is required to make a *B* response. By keeping the average response criteria constant, overall accuracy and RT should remain (approximately) constant. Furthermore, this yoked shift should lead to 1) an increase in accuracy for *B* responses as less evidence must be accumulated to reach the *B* criterion, and by corollary that 2) RTs for *b* stimuli should decrease as accumulating less evidence should take less time; the converse—namely decreased accuracy and increased RT—would be predicted for *A* responses. It is worth noting that data compatible with these speed-accuracy relationships were reported by Wagenmakers, Ratcliff, Gomez, and McKoon (2008) in a lexical decision

experiment in which the proportion of words and nonwords was varied across blocks of trials.

To evaluate our proposal, we implemented a connectionist model of the lexical decision task to examine the effects of error correction via yoked criteria shifts. We also carried out a lexical decision experiment in which we manipulated the truthfulness of external feedback participants received so as to alter perceived errors and determine whether participants' responses adapted in a similar fashion as in the model.

Simulation

Our simulation builds upon previous connectionist models of word processing (Plaut, 1997) and decision making (Usher & McClelland, 2001), and how information from the former can be fed to a decision system to model lexical decision in a relatively comprehensive fashion (Joordens, Piercey, & Azarbehi, 2003). In particular, our simplified version of lexical decision consists of 1) a visual word processing network wherein an orthographic input gradually activates semantic representations[1], and 2) a pair of decision units able to measure the information content of semantics during the presentation of words and nonwords and use this to decide what type of stimulus was presented. Holding all other parameters constant, we implemented yoked shifts of the decision criteria to determine whether this causes the predicted decrease in accuracy and increase in RTs for one type of response, and the converse for the other.

Network Architecture. The network consisted of 48 orthographic input units, 200 hidden units, and 100 semantic output units. The orthographic units were subdivided into three slots of 16 units, each of which represented a single letter in a three-letter word. The hidden and semantic units integrated their net input over time (dt = 0.1) and their outputs were a sigmoidal function of their net input.

The orthographic units fed their activation to the hidden units, and the hidden units fed their activation to the semantic units. Additionally, the semantic units fed their activation back to the hidden layer. The hidden and semantic units also received input from a bias unit. For all but the bias connections, the initial weights were sampled randomly from a uniform distribution with a mean of 0.0 and a range of .25. The bias weights were sampled with a mean of -1.7 and a standard deviation of .25 to reduce the overall activation in the hidden and output units; these bias weights were not altered during training.

Training Patterns. The network was trained on 518 pairs of orthographic and semantic representations corresponding to all three-letter words in the MRC Psycholinguistics database (Coltheart, 1981; Wilson, 1987). Artificial representations for each letter in the alphabet were generated by randomly activating 4 features in a 16 feature vector, with the constraint that these representations differed from one another by at least 2 units. This ensured that each letter was represented somewhat distinctly while also partially recycling the orthographic units. The complete orthographic representation for a word consisted of the activation of each of its letter representations across the respective slots in the orthographic pool. To approximate the categorical structure of semantics, unique semantic representations for each of the words were generated as in Plaut (1997). First, 37 category prototypes with 15 of 100 semantic units active were generated. Each prototype was then distorted to generate a total of 14 category members by regenerating each of the its features with a probability of .05, and deciding to activate a regenerated feature with a probability of .15; these representations were further constrained such that they all differed from one another by at least three units. Semantic representations were randomly paired with orthographic representations to reflect the arbitrariness of orthographic-to-semantic mappings.

Training. The model was trained using recurrent back-propagation through time with a learning rate of 0.002 and momentum descent of 0.9 (set to 0.0 for the first 50 sweeps through the training examples). Before each example, the activation in all of the units in the network was set to 0.15. Each example consisted of clamping on an orthographic representation for 50 unit updates, and allowing this activation to percolate through the hidden and semantic units. Cross-entropy error was calculated during the last 10 unit updates, with units considered to be correctly activated or inactivated when they were within 0.1 of their target values. Weight changes based on error was applied after the presentation of the full example set. Training proceeded until all units in all examples were within 0.1 of their target values during the last 10 unit updates; this required approximately 10 000 sweeps through the training corpus.

Simulating Lexical Decision. Lexical decision was based on the information content of the semantic units, which we measured using stress S_j (Plaut, 1997), defined as:

$$S_j = a_j \log_2(a_j) + (1-a_j)\log_2(1-a_j) - \log_2(0.5)$$

Where a_j corresponds to a unit's activation. Stress is a nonlinear function of the degree to which a unit's activation deviates from 0.5. Given that the network was trained such that word stimuli would correctly activate or inactivate semantic features within a radius of 0.1, stress should be high for words. In contrast, nonwords should partially activate multiple semantic representations; this blended semantic representation should contain less extreme unit activations and hence produce lower stress. For clarity, in the present simulation overlap of the word and nonword stress distributions was minimized by selecting the 518 three-letter nonwords with the lowest stress after 50 unit updates.[2]

[1] For simplicity our simulation does not contain phonological or early visual representations, although a complete model of lexical decision would include such factors.

[2] This maximizes the network's ceiling performance, but the effects of yoked feedback are not strictly bound to these extreme nonwords.

Lexical decisions were made by 'word' and 'nonword' leaky integrator decision units (Usher & McClelland, 2001). These units approached their respective decision criteria by accumulating both excitatory external input, and inhibitory input from the competing unit and a leakage factor. Formally, a decision unit's activation a_j was defined as:

$$a_j = (1-\tau)a_{j(t-1)} + \tau(I_e - ka_{j(t-1)} - Ba_{i(t-1)}); \ \max(a_j, 0).$$

Where $a_{j(t-1)}$ corresponds to the unit's activation at the previous unit update, I_e corresponds to the unit's external input (the 'word' unit's external input was the semantic stress trajectory in the current example; the 'nonword' unit's external input was the average, or referent, trajectory across all experimental words and nonwords), k corresponds to a decay scaling factor, B corresponds to an inhibition factor, $a_{i(t-1)}$ corresponds to the activation of the competing decision unit at the previous unit update, and τ corresponds to a time integration constant. The resulting activations are bounded to not drop below zero, and a decision is defined as occurring once one of the units crosses pre-specified decision criteria (discussed below). In our simulations, only the decision criteria for the yes and no units were varied, with all other parameters remaining fixed ($a_i = a_j = 0.0$ at the onset of a trial, $k = 0.1$, $B = 0.7$, $\tau = 0.1$, to match the time integration in the orthography-to-semantics network).

This unit activation equation corresponds to a simplified version of the leaky integrator units described by Usher and McClelland (2001; Equation 4, p. 559) from which the Gaussian noise term has been dropped for simplicity. Thus, the only source of trial variability is due to variability in the stress trajectories of the words and nonwords. In cases where the word and nonword distributions minimally overlap, only the stress trajectories of words should be sufficiently above those of the referent trajectory to drive the 'word' unit above its decision criterion; conversely, only the stress trajectories for nonwords should be sufficiently below the referent trajectory for the referent trajectory to drive the 'no' unit above its decision criteria. However, as participants are pressed to respond more rapidly under difficult conditions, the increased overlap of the word and nonword trajectories should lead to increased errors.

We first simulated lexical decision results that roughly correspond to those of the truthful feedback blocks in the Experiment we report, in which participants are instructed to respond as quickly and as accurately as possible to a difficult lexical decision task and performance has dropped below ceiling. To do so, we employed a 'word' decision criterion of 0.355 and a 'nonword' decision criterion of 0.360, and collected response data for all 518 words and nonwords. These criteria were selected by decreasing the decision criteria so that the units were responding when there was still considerable overlap in the stress distributions for words and nonwords. In two additional conditions, yoked criteria shifts theorized to occur when there is a perceived decrease in relative accuracy for either words (i.e., increased 'word' decision criterion, decreased 'nonword' decision criterion) and nonwords (i.e., decreased 'word' decision criterion, increased 'nonword' criterion) were simulated by shifting the response criteria by 0.003 in opposite directions, and responses for all experimental words and nonwords were again collected.

Results and Discussion

As a manipulation check, before simulating difficult lexical decision we examined the model's lexical decision accuracy if allowed to process information across 50 unit updates (similar to a non-speeded condition); the network showed near perfect performance (overall accuracy > 98%). We then examined the effects of yoked criteria shifts relative to baseline performance in a difficult speeded lexical decision task, the results of which are reported in Table 1. The results show the predicted changes in accuracy and feedback after yoked criteria shifts. Relative to baseline, word decisions are slower and less accurate when the word decision criterion is increased and the nonword criterion is decreased; the converse is true for the converse manipulation. Given the low standard errors, we have forgone reporting detailed statistical analyses of the data.

Table 1. Accuracy and Reaction Time for the Simulation

	Condition											
	Baseline				W (+), NW (-)				W (-), NW (+)			
Lex	Acc	SE	RT	SE	Acc	SE	RT	SE	Acc	SE	RT	SE
W	.71	.02	17.82	.08	.64	.02	17.93	.08	.77	.09	17.24	.02
NW	.64	.02	19.94	.03	.73	.02	19.63	.03	.59	.04	20.36	.01

Lex = lexicality; Acc = accuracy; SE = standard error of the mean (stimuli); RT = reaction time (unit updates). W = word; NW = nonword.

Experiment

The behavioral experiment was designed to manipulate the position of the decision criteria while holding all other aspects of the task constant. To do so, we implemented a difficult version of lexical decision to drop performance below ceiling and to be able to observe criteria shifts, while also encouraging participants to rely on external feedback to calibrate their decision system. We then manipulated perceived errors to either words or nonwords via two forms of false feedback to determine if this produced the criteria shifts predicted by the simulation.

The experiment was divided into four conditions.[3] Conditions Ia and Ib contrasted truthful feedback versus concordant false feedback to nonwords (Ia) and words (Ib), respectively, by informing participants that they had correctly responded when they were in fact incorrect. Based on our simulations, we predicted that this would lead to a relative increase of the perceived accuracy for the type of stimulus receiving congruent false feedback and a relative decrease in the perceived accuracy for the type of stimulus

[3] An additional control condition not reported showed that providing feedback *per se* has no significant effect on performance.

receiving truthful feedback. Consequently, participants' decision criteria should be shifted such that less evidence was required to make decisions indicating the item was of the type receiving false feedback (leading to higher accuracy and faster responses), and more evidence was required to make responses indicating the item was of the type receiving truthful feedback (leading to lower accuracy and slower responses). For condition Ia, relative to truthful feedback, concordant false feedback for nonwords should lead to faster and more accurate feedback for words and slower and less accurate responses for nonwords; the converse should be true in condition Ib.

Conditions IIa and IIb contrasted the effects of truthful feedback versus discordant false feedback to nonwords (IIa) and words (IIb), by indicating that participants had responded incorrectly to a particular item when they were in fact correct. Interestingly, although superficially different, our proposed account treats the effects of discordant feedback for a given type of item as functionally equivalent to that of concordant false feedback for that type of item. To understand why, consider what type of error participants believe they have made when they receive discordant false feedback to nonwords (IIa). Essentially, providing feedback that their nonword response was incorrect is equivalent to providing feedback that they incorrectly responded 'nonword' to a word item. Thus, to minimize this type of error, we predict that they will decrease their word decision criteria and increase their nonword decision criteria, exactly as they did in condition Ia. Our predictions for each subsection of condition II are therefore identical to the corresponding subsection of condition I.

Method

Participants. Undergraduate students in the introductory psychology course at the University of Toronto Scarborough participated in the experiment; 52 in condition Ia, 54 in condition Ib, 53 in condition IIa, and 52 in condition IIb. All participants had normal or corrected to normal vision and participated in only one of the conditions.

Aparatus. Computers running E-prime 1.1.4.1 (Schneider, Eschman, & Zuccolotto, 2002) were used to execute the experiment. Each machine displayed output on a 15" Dell CRT monitor at a refresh rate of 85 Hz, and was equipped with headphones for the presentation of auditory feedback. Participants responded on a standard keyboard.

Stimuli and Design. The word stimuli were sampled from the MRC Psycholinguistics Database (MRC, 2005), and consisted of 160 nouns between four and six characters in length with a written frequency between 1 and 400 in the Kucera-Francis norms (mean = 55, SD = 66, skew = 2.9). The nonwords were generated by sampling a second set of non-overlapping word stimuli from the database constrained by the aforementioned criteria, and replacing a single consonant with another random consonant to make a nonword not in the database. (e.g., FATHER → NATHER). This produced nonwords with wordlike orthotactic structure so as to exacerbate task difficulty.

For each participant, the stimuli were randomly divided into two blocks of 160 items for use in the truthful feedback and false feedback blocks. The order of stimuli within these blocks was also randomized.

Procedure. Participants were instructed to decide whether the characters on the screen formed a word or a nonword by pressing "z" or "/", respectively, and were provided with a demonstration trial. They were instructed to respond to each trial as quickly and as accurately as possible.

Each trial consisted of six steps: (1) a 250 ms blank field, (2) a 500 ms fixation cross, (3) a 50 ms presentation of a lowercase character string, 4) a 50 ms mask consisting of three lines of 10 random characters filling the line where the probe string was presented, and the lines above and below it, (5) a response screen, and (6) 1000 ms of feedback, as detailed below. At the end of each trial, the next trial began automatically; the procedure required approximately 40 minutes. Note the very short duration of the probe and the presentation of the character mask, which were used to lower performance from ceiling and encourage participants to rely on external feedback to detect errors.

Feedback consisted of either 1) "CORRECT" and a bell sound, or "INCORRECT" and a buzzer sound. Feedback reflected response accuracy during the truthful feedback block. During false feedback, 50% of eligible items (i.e., incorrect responses for a particular type of item during concordant false feedback; correct responses for a particular type of item during discordant feedback) resulted in false external feedback. Only half of the eligible trials received false feedback to make the manipulation difficult to detect.

Following the experiment, participants completed a debriefing questionnaire to determine whether they were aware of the systematic change in feedback accuracy; according to the debriefing, none were.

Results

Prior to analysis, trials were binned based on lexicality (word vs. nonword), order of blocks (truthful block first vs. last), feedback block (truthful vs. false), and decision accuracy (correct vs. incorrect). Accuracy analyses only included trials with RTs greater than 200 ms and within 2.5 standard deviations of the bin's mean RT (92% of trials). Correct trials meeting these restrictions were included in the RT analyses. For efficiency, each condition's descriptive statistics and the results of a mixed ANOVA with two within-subjects variables (lexicality, feedback block) and a between-subjects variable (order of feedback) are presented in Tables 2 and 3. All significant effects have $p < .05$.

Within-condition Accuracy. In condition Ia, we observed a lexicality by feedback interaction consistent with the predicted effect of a yoked criteria shift. Explored further,

via t-tests, we confirmed the predicted effects of words becoming marginally more accurate (t_{53} = 1.66, p = .1), and nonwords becoming significantly less accurate during false feedback (t_{53} = 2.09). Additionally, we observed a main effect of lexicality (words more accurate than nonwords) and a feedback by order of feedback interaction. We explored this latter effect further in separate 2 (lexicality: word vs. nonword) x 2 (feedback: truthful vs. false) within-subjects ANOVAs for each order of feedback presentation; both these analyses showed main effects of feedback such that participants were more accurate during the second block of trials (truthful feedback first: $F_{1,27}$ = 4.62; false feedback first: $F_{1,25}$ = 9.74).

In condition Ib, we observed a two-way lexicality by feedback interaction and a three-way interaction between lexicality, feedback, and order of feedback. To explore this latter interaction, we conducted separate 2 (lexicality: word vs. nonword) x 2 (feedback: truthful vs. false) within-subject ANOVAs for each order of block presentation. When truthful feedback was presented first, we observed the expected lexicality by feedback interaction ($F_{1,21}$ = 12.51) with words becoming less accurate and nonwords becoming more accurate during the false feedback block (words: t_{21} = 2.76; nonwords: t_{21} = 2.95). However, no lexicality by feedback interaction or other effects were observed when truthful feedback was presented second.

In condition IIa, we observed a lexicality by feedback interaction consistent with the predicted effect of yoked criteria shifts. To explore this interaction, we conducted t-tests on the words and nonwords in the truthful feedback and false feedback conditions, which confirmed that under false feedback responses were significantly more accurate for words (t_{52} = 4.10) and significantly less accurate for nonwords (t_{52} = 3.40). Additionally, we observed a main effect of lexicality (words being more accurate than nonwords) and a feedback by order of feedback interaction which further analysis via separate 2 (lexicality: word vs. nonword) x 2 (feedback: truthful vs. false) within-subject ANOVAs for each order of block presentation showed to be the result of participants becoming significantly more accurate in the second block (main effect of feedback, truthful feedback first: $F_{1,31}$ = 5.44; false feedback first: $F_{1,20}$ = 15.20).

In condition IIb, we observed lexicality by feedback interaction and a three-way interaction between lexicality, feedback, and order of feedback. We explored this latter interaction further via separate 2 (lexicality: word vs. nonword) x 2 (feedback: truthful vs. false) ANOVAs for each order of block presentation. When truthful was presented first, the predicted lexicality by feedback interaction was observed, ($F_{1,27}$ = 26.15) such that responses were less accurate for words (t_{27} = 3.33) and more accurate for nonwords (t_{27} = 5.71) under false feedback. However, when false feedback was presented first there was no significant lexicality by feedback interaction ($F_{1,23}$ < 1), and there were main effects both of lexicality, such that nonwords were responded more accurately than words ($F_{1,23}$ = 7.00), and of feedback, such that participants were faster during the second block which consisted of truthful feedback ($F_{1,23}$ = 9.89).

Table 2. Accuracy and Reaction Times in the Experiment

			Condition							
			Ia				Ib			
OF	B	L	Acc	SE	RT	SE	Acc	SE	RT	SE
TFF	TF	W	.69	.02	811	17	.67	.03	790	14
		NW	.58	.02	935	18	.61	.02	864	14
	FF	W	.74	.02	674	15	.58	.04	689	13
		NW	.59	.04	776	15	.68	.02	702	13
FFF	TF	W	.73	.03	655	14	.62	.04	602	11
		NW	.58	.04	726	15	.71	.03	631	11
	FF	W	.74	.03	705	13	.62	.04	674	10
		NW	.49	.04	864	18	.71	.02	794	14
			IIa				IIb			
			Acc	SE	RT	SE	Acc	SE	RT	SE
TFF	TF	W	.68	.03	714	13	.65	.03	784	14
		NW	.59	.02	812	12	.57	.02	885	15
	FF	W	.77	.03	588	11	.57	.04	738	13
		NW	.57	.03	693	12	.68	.03	733	13
FFF	TF	W	.70	.03	616	9	.64	.04	725	15
		NW	.58	.03	739	12	.73	.03	715	13
	FF	W	.74	.03	723	12	.58	.03	877	14
		NW	.45	.03	882	12	.68	.02	835	13

Note. In condition Ia, 28 participants received truthful feedback first; 26 false feedback first. In condition Ib, 22 participants received truthful feedback first; 29 false feedback first. In condition IIa, 32 participants received truthful feedback first; 31 false feedback first. In condition IIb, 28 participants received truthful feedback first; 24 false feedback first. OF = Order of feedback blocks; TFF = truthful feedback first; FFF = false feedback first; B = block; TF = truthful feedback; FF = false feedback; L = lexicality; W = word; NW = nonword; Acc = accuracy; SE = standard error of the mean; RT = reaction time (ms)

Table 3: F-Statistics for the 2x2x2 ANOVAs in the Experiment

	Condition							
	Ia		Ib		IIa		IIb	
	Acc	RT	Acc	RT	Acc	RT	Acc	RT
lex	42.46*	35.07*	3.00†	17.68*	56.87*	72.32*	5.13*	< 1
lex*ofb	1.73	< 1	1.31	1.23	1.42	1.88	3.11*	6.31*
feedback	< 1	< 1	< 1	< 1	< 1	< 1	4.27*	< 1
fb*ofb	14.00*	16.40*	< 1	60.48*	16.93*	81.75*	10.62*	33.42*
lex*fb	5.44*	1.02	11.91*	< 1	25.43*	1.48	18.03*	14.15*
lex*fb*ofb	1.21	3.01†	12.33*	14.02*	1.87	< 1	12.47*	4.21*
ofb	< 1	1.11	< 1	4.41*	1.37	1.40	1.96	< 1

Note. Tests have 1 degree of freedom treatment. Conditions Ia through IIb have 52, 50, 51, and 50 degrees of freedom error. lex = lexicality; ofb= block order; fb = feedback; Acc = accuracy; RT = reaction time. † p < .1; * p < .05

Within-condition RT. In condition Ia, we observed an effect of lexicality (words faster), and a feedback by order of feedback interaction. Separate 2 (lexicality: word vs. nonword) x 2 (feedback: truthful vs. false) ANOVAs for each block presentation order showed this latter effect to be due to faster RTs in the second block (main effect of feedback, truthful first $F_{1,27}$ = 11.02; false first $F_{1,25}$ = 5.74).

In condition Ib, we observed a main effect of lexicality (words faster than nonwords), a main effect of order of feedback (faster for truthful feedback), a feedback by order of feedback interaction, and a lexicality by feedback by order of feedback interaction. To explore these interactions further, separate 2 (lexicality: word vs. nonword) x 2

(feedback: truthful vs. false) within-subjects ANOVAs for each order of block presentation were effectuated. All of the effects in these ANOVAs were significant (truthful first, lexicality $F_{1,21}$ = 3.50; feedback $F_{1,21}$ = 45.65; interaction $F_{1,21}$ = 6.76; false first, lexicality: $F_{1,28}$ = 19.32; feedback $F_{1,28}$ = 25.84; interaction $F_{1,28}$ = 8.92) and indicated that responses were on average faster in the second block, and differentially faster for nonwords.

In condition IIa, we observed a main effect of lexicality (words faster than nonwords), and a feedback by order of feedback interaction which separate 2 (lexicality: word vs. nonword) x 2 (feedback: truthful vs. false) within-subjects ANOVAs for each order of block presentation revealed to be the result of faster RTs in the second block of the experiment (truthful feedback first, feedback $F_{1,31}$ = 51.00; false first, feedback: $F_{1,30}$ = 34.23).

In condition IIb, only the interaction effects were significant. To explore these interactions further separate 2 (lexicality: word vs. nonword) x 2 (feedback: truthful vs. false) within-subjects ANOVAs for each order of block presentation were effectuated. When truthful feedback was presented first, all of the effects were significant (lexicality $F_{1,27}$ = 11.34; feedback $F_{1,27}$ = 16.18; interaction $F_{1,27}$ = 15.10), whereas there was only an effect of feedback when false feedback was presented first; these effects indicated that responses were faster in the second block, and in the case of truthful feedback that responses were on average significantly faster for words, and grew differently faster for nonwords during false feedback.

Discussion

Based on the yoked criteria shift theory of decision system adaptation we proposed and demonstrated via computational simulation, we derived a series of predicted accuracy and RT effects for each of the different feedback manipulations. In the accuracy data, the predicted effects were always present when both variants of false feedback were provided for nonwords; for words, the predicted effects of feedback were also observed, but only when false feedback was preceded by truthful feedback. In the RT data, none of the predicted RT shifts occurred. However, no unpredicted RT shifts running contrary to the yoked criteria shift account were observed either. This suggests that participants did shift their decision criteria as predicted, but traded off variations in speed for greater variations in accuracy.

In addition to this highly (though not perfectly) consistent adaptation predicted by the yoked criteria shifts, a number of other effects were observed throughout the different conditions with varying degrees of reliability. In particular, there were several similarities in the types of effects observed when false feedback was provided to nonwords and words, with the former being a cleaner match to the simulation data. These additional effects, although in some cases probably worthy of verification via replication, may provide an additional set of constraints for the development of more detailed models of decision system adaption.

General Discussion

Decision system adaptation to perceived changes in accuracy is critical in changing environments. The computational and behavioral results we have reported provide converging evidence that one simple yet powerful mechanism for effectuating such adaptations in a calibrated decision system are yoked shifts of decision criteria.

In the present work, we have intentionally kept our simulation and behavioral analyses relatively simple so as to facilitate relating them to our theory. We are currently examining whether some of the phenomena unexplained by the current simulation (e.g., predicted effects of feedback not occurring when false feedback is given to words before truthful feedback) could be accounted for by yoked criteria shifts if we equate the simulated referent trajectory to the stimuli classifications participants perceive to be correct, and by matching the wordlikeness distributions of the simulated word and nonword stimuli to those used in the behavioral experiment.

Acknowledgments

This research was supported by a NSERC CGS award to B.C.A., an NSERC Discovery grant to S.J., and an NIH award to D.C.P. We thank the anonymous reviewers for their helpful comments.

References

Coltheart, M. (1981) The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33*, 497-505.

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go lexical decision task. *Journal of Experimental Psychology*: *General, 136*(3), 389-413.

Joordens, S., Piercey, C. D., & Azarbehi, R. (2003). From word recognition to lexical decision: A random walk along the road of harmony. In F. Detje, D. Dörner, & H. Schaub (Eds.), *Proceedings of the 5th International Conference on Cognitive Modeling,* 141-146.

MRC Psycholinguistics Database. (2005). Retrieved February 1, 2005, from psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, *12*, 767-808.

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

Wagenmakers, E-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, *58*, 140-159.

Wilson, M. (1987) MRC Psycholinguistic Database: Machine Usable Dictionary, Version 2. Memo - IKBS Section, Rutherford Appleton Lab.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime* (Version 1.1.4.1) [Computer software] Pittsburgh, PA: Psychology Software Tools, Inc.