Ahmad Sohrabi and Robert L. West

Positive and Negative Congruency Effects in Masked Priming: A Neuro-computational Model Based on Representation Strength and Attention

Positive priming effects have been found with a short time between the prime and the target, while negative priming effects (i.e., a congruent prime causes longer RTs) have been found with a long time between the prime and the target. In the current study, positive and negative priming effects were found using stimuli that have strong and weak representations, respectively, without changing the time between prime and target. A model was developed that fits our results. The model also fits a wide range of previous results in this area. In contrast to other approaches our model depends on attentional neuro-modulation not motor self-inhibition.

In masked priming tasks, a brief masked stimulus (the prime) can affect the processing of the stimulus that follows (the target). A prime, a mask, and a target are presented sequentially and the task is to make a decision on the target. The result is usually a Positive Congruency Effect (PCE), also known as the positive compatibility effect. In PCE, the prime increases the performance on the target if they are congruent and decreases the performance if they are incongruent[1-4]. Conversely, a negative priming effect has been found, called the Negative Congruency Effect (NCE). This effect is also known as the negative compatibility effect, where paradoxically the prime increases the performance on the target if they are incongruent and decreases the performance if they are congruent[4-14]. The PCE has been shown with a short mask-target Stimulus Onset Asynchrony (SOA), while the NCE has been shown with a longer mask-target SOA (e.g., 100 ms). To explain these results, some researchers[4,15,16], based on Event Related Potential (ERP) measurements and computational modelling, argue that when SOA is short, response selection can already take place during the initial response activation phase; this is reflected as an early increase in ERP for the congruent compared to incongruent trials, and this should result in the congruency effects in the form of a PCE. When SOA is longer, responses have to be selected during the subsequent inhibitory phase. This is reflected as a late decrease in ERP for congruent compared to incongruent trials, and this should be demonstrated as a negative effect (i.e., NCE). In these studies, the reduction in ERP activity has been attributed to a motor self-inhibition, causing the NCE effect. The mask causes this inhibition to be reversed, by removing the sensory evidence for the corresponding response and initiating its suppression.

The current study includes an experiment and a neuro-computational model. While previous studies have shown PCE and NCE by using short and long mask-target SOA, respectively, the present experiment shows PCE and NCE by using strong and weak stimulus representations, and a fixed (and relatively short) mask-target SOA. It has been shown that the format of the stimuli (Arabic numerals versus written numerals) affects priming[3]. It has also been shown that the distance between number magnitudes can affect priming when the task involves saying whether the target number is higher or lower than a set comparison number[17,18]. In our study we used these manipulations to manipulate the "strength" of the stimuli. We interpreted manipulating Arabic versus written numbers as manipulating of how well learned the representations were. To further investigate this we had subjects memorize associations between arbitrary symbols and numerical values and tested them in same priming paradigm. We interpreted both manipulations (numerical distance and degree of stimulus learning) as additively affecting the strength of the

relationship between the stimuli. So, for example, a recently learned symbol and a close distance to the reference number would, additively, make the relationship weaker, while a well-learned symbol and a far distance from the reference number would make the relationship stronger. Our results, as interpreted through our model, are consistent with this view.

The previous model of PCE and NCE[16] depends on motor self-inhibition assumption and does not show a decline in NCE and overall Reaction Times (RTs) throughout time. It shows a huge PCE eventually at very long SOAs. However, human data shows that NCE decreases and disappears or turns into a very small PCE at very long SOAs. The current model does shows these effects and does not depend on motor self-inhibition, instead it works through attentional modulation that can be affected by conflict. Our model also shows the PCE with strong prime and target and NCE with weak prime and target, and, without any changes in the parameters, shows a PCE at short mask-target SOA and an NCE at long mask-target SOA. It also shows the effect of other factors such as degradation[5], mask density[9], and prime duration[9]. The model is based on previous models that have been used to simulate different tasks such as target detection and simple decisions in monkeys and humans[19-22].

RESULTS

Experimental results

The experiment had a learning phase, where participants mapped symbols to numbers, and a priming phase where participants made numerical decision on symbols and numerals (see Methods). The results of the training phase, the mean RTs and percent errors of participants, showed that they had associated the symbols to numbers successfully. By the third and fourth blocks of the four symbol-number maps, RTs decreased substantially (**Supplementary Fig. 3a** online) and no errors occurred (**Supplementary Fig. 3b** online).

In half of the trials in the priming phase, prime and target were Arabic numerals (henceforth, Number-Number) while in the other half prime was an Arabic numeral and target was a symbol (henceforth, Number-Symbol) (see Methods). Also there were four types of trials based on the distance of prime and target from *five*: far prime and far target (henceforth, Far-Far), far prime and close target (henceforth, Far-Close), close prime and far target (henceforth, Close-Far), and close prime and close target (henceforth, Close-Close).

To analyze the data from the main task, the priming phase, a three-way (2 x 2 x 4) repeated measures ANOVA on RTs was run with three factors: Type (Number-Number vs. Number-Symbol), Congruency (Congruent vs. Incongruent), and Distance (Close-Close, Far-Close, Close-Far, and Far-Far). The results revealed significant main effects of Type, $F_{1, 9} = 131.001$, $p < .001$, Congruency, $F_{1, 9} = 6.519$, $p = .031$, and Distance, $F_{3, 27} = 4.581$, $p = .01$. Thus, participants were faster in Number-Number trials than Number-Symbol trials, showing the difference between numeral and symbol targets. Also, participants were faster in Congruent trials than Incongruent trials, and in the trials with Far target than those with Close target. The two-way significant interactions were Type by Distance, $F_{3, 27} = 9.294$, $p < .001$, and Congruency by Distance, $F_{3, 27} = 6.385$, $p = .002$, and the three-way significant interaction was Type by Congruency by Distance, $F_{3, 27} = 5.995$, $p = .003$. The analysis using t-tests revealed the role of Distance factor, showing a significant PCE at Far-Far in Number-Number (**Fig. 1a**), $t = 4.555$, $p = .001$ ($p=.008$, adjusted for multiple comparisons with Bonferroni method), a significant NCE

at Close-Close in Number–Symbol (**Fig. 1b**), t = 3.692, p = .005 (p=.04, adjusted for multiple comparisons with Bonferroni method), and no significant priming in the other two Distance levels, Close-Far, in Number-Number, t = .980, p = .352 and Number–Symbol, t = .657, p = .527 (**Supplementary Fig. 4a** online), and Far-Close, in Number-Number, t = .441, p = .67 and Number–Symbol (marginally significant), t = 2.039, p = .072 (non-significant when adjusted for multiple comparisons with Bonferroni method, p=.576) (**Supplementary Fig. 4b** online). By excluding Distance factor, PCE was significant in Number-Number, t = 3.737, p = .005 (p=.01, adjusted for multiple comparisons with Bonferroni method), and non-significant in Number-Symbol, t = .665, p = .522 (**Supplementary Fig. 4c** online).

A separate repeated measures ANOVA on errors was run with the same three factors in analysis of RTs: Type (Number-Number vs. Number-Symbol), Congruency (Congruent vs. Incongruent), and Distance (Close-Close, Far-Close, Close-Far, and Far-Far). The main effect of Type was significant, $F_{1, 9}$ = 12.623, p = .006, showing that participants had fewer errors in Number-Number compared with Number-Symbol. Moreover, a Type by Congruency interaction was significant, $F_{1, 9}$ = 5.762, p = .04, showing that participants had less errors in Congruent compared with Incongruent trials, especially in Number-Symbol (**Supplementary Fig. 5** online). The t-test showed no significant PCE for Number-Number, t = .524, p = .613, but a marginally significant PCE for Number-Symbol condition, t = 2.211, p = .054 (non-significant when adjusted for multiple comparisons with Bonferroni method, p=.108).

Figure 1 Main results of the experiment. (a) At Far Prime-Far Target distance (i.e., Far-Far), there was a significant PCE for Number Prime–Number Target (i.e., Number-Number) and a marginally significant PCE for Number Prime–Symbol Target (i.e., Number-Symbol). (b) At Close Prime-Close Target (i.e., Close-Close), there was a non-significant PCE for Number–Number, but a significant NCE for Number–Symbol. Therefore, a PCE was found at one extreme, the Far-Far Number-Number (easy), and at the other extreme, the Close-Close Number-Symbol (hard), an NCE was found (for levels between these two extremes, see **Supplementary Figs. 4a** and **b**).

Modeling results

Simulation 1: strength

This simulation was designed to simulate data from the present experiment to show the role of the representation strength of prime and target in terms of IL-RL weights (see Methods). The main result of this simulation is shown in **Figure 2,** for two opposite conditions: Far-Far Number-Number and Close-Close Number-Symbol. The first condition was simulated with IL-RL weight 3 for prime, 3 for target (i.e., 3-3) and the second condition was simulated with 2.5 for prime, 2 for target (i.e., 2.5-2). For the results of all prime-target weight combinations see **Supplementary Figure 6** (see also **Supplementary Table 1** online). As the weights of prime and target decreased, the priming pattern changed from positive to negative. A stronger weight caused a PCE and a weaker weight caused an NCE. However, this relationship was not linear, as making the weights too weak did not cause larger NCEs (see **Supplementary Fig. 6** online).

In general, the PCE and NCE patterns qualitatively fit the experimental data. The symbol type and close distance were considered as having low strength and Arabic numeral type and far distance were considered as having high strength. When the prime and target representations (here, weights) were strong, a PCE was found, but when the prime and target representations were weak, an NCE was found. This result was found by simply representing the close distance and symbol type in the same way by a weak weight (2), and by adding a small, equal weight (.5) to each of them for far distance and numeral type (see Methods and **Supplementary Table 1** online).

Figure 2 Main results of Simulation 1, strength effect. (a) At Far Prime-Far Target distance (i.e., Far-Far), there was a PCE for Number Prime–Number Target (i.e., Number-Number) and Number Prime–Symbol Target (i.e., Number-Symbol), simulated by prime-target weights 3-3 and 3-2.5. (b) At Close Prime-Close Target (i.e., Close-Close), there was a small PCE for Number–Number and an NCE for Number–Symbol, simulated by prime-target weights 2.5-2.5 and 2.5-2 (the same human data are also shown in **Fig. 1**).

Simulation 2: mask-target SOA

To simulate the data from previous studies i.e., a PCE and an NCE with short and long mask-target SOAs, respectively[1-14] with no changes in the parameters except the mask-target SOA, we used the model in the previous simulation with medium prime and target strength in IL (2.5). Seven intervals of the mask-target SOA (from 71 to 251, with 30 cycles interval) were used to show the effect of SOA on priming pattern. To maintain consistency, the duration of the mask was again 71 cycles, but a longer mask duration has a similar effect (as used in the following simulations).

In previous studies[4,8,14], the NCE has been shown at long SOAs. As shown in **Figure 3a**, here at the first SOA a small PCE occurred (stronger PCE can be found with shorter SOAs), and at longer SOAs an NCE occurred. Then NCE and RTs slowly decreased and finally the effect became slightly positive again. The error results can be found in **Supplementary Figure 4,** showing that in short SOA, errors occurred mainly in the incongruent trials, and in long SOA these errors occurred mainly in the congruent trials (in forms of missing trials, i.e., ML activation did not cross the threshold by the trial deadline). Also, note the decrease in RTs throughout time in **Figure 3a** which is similar to previous experimental data[14], and the U-shaped curve of the RT difference (**Fig. 3b**), both of which result from recovery from the attentional refractory period (see **Supplementary Fig. 10** online). **Figure 3b** shows the same results in **Figure 3a** but using congruency difference (i.e., incongruent – congruent mean RTs). It is similar to the result of attentional blink paradigm, presumably showing a common attentional basis for attentional blink[22] and priming effects in the current study.

Figure 3 The results of Simulation 2, SOA effect. (a) Modelling results at seven levels of mask-target SOA, starting from 71 cycles. Each SOA follows 30 cycles after the previous one, with mask duration of 71 cycles. (b) The same result was shown by the congruency difference

(Incongruent - Congruent) in the seven SOAs. This is similar to the different lags in attentional blink paradigm, showing a similar attentional basis for priming and attentional blink.

Simulation 3: stimulus degradation

A previous study found that degradation of stimuli, by adding small random dots to all stimuli, turns NCE into PCE[5]. Here, the degradation of stimuli was simulated by using lower input activation in IL (for both prime and target) compared to the usual 1 and 0 and increasing the noise of the prime and target in RL. Two levels of degradation were created by using .85 (opposite unit .15) and .75 (opposite unit .25), while 1 (opposite unit 0) was used to encode an intact stimulus. For a better fit between simulation and human data, the noise of the prime and target units in RL was increased from .2 to .3. The IL-RL strength for the prime and target was 2.5 and the mask-target SOA was 125 cycles. The model successfully simulated the human data as shown in **Figure 4a**. With degradation, the NCE turned into PCE and RTs were increased by more degradation.

In another experiment in the same study[5], random dots were added to all stimuli, but the dots did not cover the target (presented above or below the target, randomly). In this case, while degradation turned the NCE into PCE, it did not increase the RTs. For simulating this experiment, a simulation was run identical to the previous one but only the prime was degraded. The result was similar to the human data. As shown in **Figure 4b**, if the target is not degraded the RTs do not increase (because it is stronger and is processed faster).

Figure 4 Results of Simulation 3, degradation effect. (a) Degrading the prime and target with three levels of prime and target inputs in IL: 1 (no degradation), .85 (medium degradation), and .75 (high degradation), as well as an increase in noise. With degraded unit activations NCE turned into PCE and RTs increased. (b) Degrading only the prime turned NCE to PCE but did not increase RTs.

Simulation 4: mask density

It has been shown that the mask needs to be dense enough at a specific rate to cause NCE, and that decreasing the density changes NCE to PCE[10], although beyond that it has no major effects. In this simulation, mask density was simulated by changing the inputs of the mask units to .55 (medium density) and .45 (low density), instead of 1 (very high density, used in other simulations where usual mask was used). The IL-RL strength for the prime and target was 2.5 and the mask-target SOA was 125 cycles. As shown in **Figure 5a**, similar to human data[10] decreasing the mask density from 1 to .55 decreased NCE and then to .45 and 0 turned NCE to PCE (low mask density and no mask are supposed to invoke other types of processes, not discussed here, but see Sohrabi, A. Positive and Negative Congruency Effects in Masked and Unmasked Priming: Match of representation strength, Attention, and Consciousness. PhD dissertation, Carleton University, 2008).

Simulation 5: prime duration

Prime duration has an important role in the priming effect. Stimuli with longer duration have stronger representations and also activate more attentional responses. It has been shown that

increasing the prime duration increases NCE to some extent and turns it to PCE after a specific rate[10]. The current simulation shows the priming effects for three prime durations: 43, 48, and 53 cycles. The IL-RL strength for the prime and target was 2.5 and the mask-target SOA was 125 cycles.

As shown in **Figure 5b**, increasing the prime duration caused larger NCE, but a further increase turned it into PCE. Interestingly, increasing the prime duration does not decrease RTs and even has an opposite effect, similar to human data[10] (longer duration is supposed to invoke other types of processes, not discussed here, but see Sohrabi, A. Positive and Negative Congruency Effects in Masked and Unmasked Priming: Match of representation strength, Attention, and Consciousness. PhD dissertation, Carleton University, 2008).

Figure 5 Results of Simulations 4 (mask density) and 5 (prime duration). (a) Four levels of mask density were employed: 1 (no mask), 2 (low density), 3 (medium density), and 4 (high density), simulated by IL mask unit inputs 0, .45, .55, and 1 compared to masks with ≥ 15, 10, 5, and 0 random lines in human data, respectively**[10]**. (b) Simulation results for three levels of prime duration: 53 cycles (long), 48 cycles (medium), and 43 cycles (short), compared to 64, 32, and 16 ms in human data[10]. Increasing the prime duration increased the NCE but a further increase turned the NCE into PCE.

METHODS

Participants. Participants were ten healthy, right-handed students (4 females and 6 males, age 22-43) with normal or corrected-to-normal vision. Each participant gave informed consent. The experiment was approved by the Carleton University Ethics Committee for Psychological Research.

Experimental Design and procedure. The stimuli were four single digit Arabic numerals (1, 4, 6, and 9) and four symbols used in playing cards: Hearts, Clubs, Diamonds, and Spades (**Fig. 6a**). In the learning phase, participants were asked to associate each symbol to a number (1, 4, 6, or 9), and then to use the mappings in a priming task by deciding if the target is larger or smaller than *five,* then pressing the corresponding right and left button on the response device (i.e., SNARC compatible[18]). In the learning phase, participants were shown the symbols and their corresponding numerals while the instructions were given to them. They had 45s to look at the symbols and their corresponding numerals after the instruction, before the test part of this phase began. The numerals and symbols were presented in two separate rows at the centre of the screen. During this time, the numerals moved away from the symbols downward, in three 15s steps, until they disappeared from the bottom of the screen (**Supplementary Figs. 1** online). Then, the participants were tested to determine if they had memorized the correct mappings between symbols and numbers. Feedbacks on accuracy, and the correct response, were provided after each trial (**Supplementary Fig. 2** online). Because the 45s presentation of the mapping prepared participants to learn the relationships quickly, only four blocks of the four randomly selected symbol-number pairs were presented (randomly in each block),. The assignment of symbols to numbers was randomized for each participant.

At the beginning of each trial, a 500 ms fixation point (# or ∗ with the same size and at the same position as the other stimuli) and 300 ms blank screen were presented. One of the two fixation points was presented randomly each time and the participants were told that the signs are merely a fixation point. Then a 43 ms numeral prime was sandwiched between two 71 ms masks and followed by a 200 ms numeral or symbol target (**Fig. 6b)**. The mask was composed of six symbols that appeared at once as a string but with a random order each time. The symbols in the mask differed from the four symbols used in the task (see **Fig. 6a**). In a congruent trial, both the prime and target were either smaller or larger than *five*, while during an incongruent trial, one was smaller and the other one was larger than *five*. In half of the trials, both prime and target were numerals (i.e., Number-Number) while in the other half prime was a numeral and target was a symbol (i.e., Number-Symbol). Moreover, the distance factor was considered and included in the analysis as another factor (i.e., Distance). One and nine were used as numbers with a far distance from *five* and four and six were used as numbers with a close distance to *five.*
The task was programmed in Microsoft Visual Basic 6 (Microsoft Corp.) with millisecond time precision using Class and Thread Priority procedure[23]. Stimuli were presented with 75 Hz refresh rates on a 14" Sony 1024 by 768 monitor. The stimuli were chosen from the standard Symbol font with size 90 points in Microsoft Windows XP presented at the centre with an approximate distance of 60 cm. In the learning phase, responses were made with the right hand using four buttons on a joystick. In the priming phase, responses were made by the index finger of the right or left hand, using R or L button on the joystick. The participants responded during the target presentation or within the remainder of the trial (2815 ms from the offset of the target); otherwise a missing trial would be recorded. The trials in the priming session were random and consisted of 16 combinations of four numerals as prime and target for Number-Number and 16 combinations of four numerals as prime and four symbols as target for Number-Symbol. Each condition was repeated three times during the task, therefore, the total number of trials was 96, presented randomly.

Figure 6 Stimuli and design in the experiment. (a) The task in the learning phase was to map each symbol (top row) to its corresponding numeral (middle row). In the priming phase the task was to treat symbols as numbers and compare each symbol or numeral target to *five*. The prime was always a numeral. The mask consisted of six symbols (bottom row) in random order each time. (b) The stimuli were presented briefly in order. The prime was sandwiched between two masks and followed by a target. The screen remained blank until the trial deadline.

Modeling methods. The processing elements in the model are a few neurons with self-excitation, lateral inhibition, and accumulative activation that have a strong computational power in simulating basic neural and cognitive processes[19-21,24,25]. It has been demonstrated that these types of reduced models can resemble the neural computation of a large group of neurons[26].
The model (**Fig. 7**) is a multi-layer dynamic neural model that consists of a feed-forward component for perceptuo-motor processing from the Input Layer (IL) to an intermediate layer, called Representation Layer (RL), and from there to the Cognitive Layer (CL) and Motor Layer

(ML, not shown in **Fig. 7**). The stimulus type (numerals and symbols) and distance (close and far) are encoded by their strength of representation, simulated by connection weights between IL and RL. Decreasing the weights is intended to make the representations weaker or less direct (i.e., harder) and increasing the weights is intended to make the representations stronger or more direct (i.e., easier). Another assumption is that the cognitive processing, including the response, is modulated by attention. The Alert Attention layer (AA) simulates attentional modulation, that is supposed to be a model of Locus Coeruleus (LC) that potentiates cortical areas through norepinephrine[27]. The executive attention is only modelled through its effects on AA, using a Cognitive Layer (henceforth, CL) for conflict monitoring. The CL effect on AA simulates direct cortical projections to LC[27]. The CL and ML are affected by both prime and target. The ML is not shown in **Figure 7** for the sake of simplicity, but its architecture is identical to CL, with the exception that it sends no outputs to AA, is slower, and noisier (see **Table 1**).

Each condition in a simulation consists of 20,000 trials (200 independent blocks of 100 trials each, with congruent and incongruent trials counterbalanced randomly within each block). A single trial takes 1100 cycles. Each block starts with 500 cycles without changes in IL to let the units in other layers reach a steady state of activation. Similarly the Inter-Trial Interval (ITI) for each trial is 500 cycles, which allows the activation of units to return to baseline following the responses. The prime is presented by clamping one of the two units in the IL to 1, intended to be smaller or larger than *five,* or left or right in the case of arrows. The mask units in IL are set to 1 at the time of mask presentation and are otherwise set to 0. Therefore, the recognition of the stimuli is implemented with a localized representation, for example, the left unit is turned on when the stimulus is less than five in the case of numerals and symbols, or points left in the case of arrows; otherwise the right unit is turned on. Accordingly, as will be described below, in a congruent trial the two corresponding units (e.g., the left unit of the prime and target in IL) is set to 1 or 0 at the time of stimulus presentation, while in an incongruent trial, one of the two relevant units of the prime or target is set to 1 and the other to 0 (or to real normalized values, e.g., .75 and .25, in some simulations for specific reasons such as stimulus degradation or mask density).

The units in each layer make connections, via excitatory weights, to their corresponding units in other layers. The activations of these units (except IL) are calculated by a sigmoid (logistic) function of the incoming information, and a small amount of random noise. The RL sends excitatory activities to ML and CL continuously but activates AA only if a unit of the prime or target reaches a designated threshold of .62. Similarly, when one of the two units in the ML reaches the same designated threshold it triggers a manual response (i.e., initiating a hand movement). When AA is activated and its activation reaches a threshold, it starts modulating information processing in RL, CL, and ML by making the activation function of their units steeper[35] (**Supplementary Fig. 8** online).

Figure 7 The architecture of the model**.** The IL projects to the RL. The RL excitates AA, CL, and ML (not shown). The AA modulates all other layers except IL. The CL changes the AA response mode in the event of conflict. Different connections are depicted with different arrows:

-♦ modulatory; -● conflict monitoring; ↶ self-excitation and lateral excitation; ●-● Lateral inhibition; -▶ Feed-forward activation.

Modelling details. As shown in **Figure 7**, the IL encodes the prime, the mask, and the target, and projects to RL through excitatory connections. For the sake of simplicity, prime and target, as well as an identical mask for each (shown as a single unit in **Fig. 7,** for the sake of simplicity) were implemented in two separate paths. The IL-RL weights for prime and target were 2.5 in all simulations, except Simulation 1. In Simulation 1, the IL-RL weights were adjusted for stimuli with different representation strength. This can be accomplished by simply representing the close distance and symbol type in the same way with a weak weight (2), and by adding a small, equal weight (.5) for each of them, to represent far distance and numeral type (**Supplementary Table 1** online). All units in RL have a self-excitation connection, intended to simulate mutual excitation among a group of neurons. Connections between mutual units (for prime and target and to the mask) from IL to RL have small cross-talks (see **Table 1**), indicating feature overlaps or similarities among stimuli. The units also have lateral inhibition with neighboring units within the same layer.
The mask units are activated after the prime and before the target for a specific time. They have lateral inhibition with prime and target. The lateral inhibition has been proposed as a good way to simulate masking[28,29]. In addition to lateral inhibition, the model simulates the similarity of the mask to the prime and target through a lateral excitation from mask to the prime and target. It plays a role using this lateral excitation and can affect ML and AA (and CL), indirectly, through its effect on both prime and target. Moreover, the prime and target units, but not the mask, have feed-forward projections into the ML, CL, and AA layers. Therefore, the mask acquires meaning through its relationship with the prime and target. Because it comes right after the prime, it can activate the prime through its excitation. So, it can act partially like the prime and increase the attentional responses to it, forcing it to stay longer, but, on the other hand, its inhibitory effect usually dominates its excitatory effect and interrupts the prime, causing it to decay faster. This interplay depends on the similarity of the mask to the prime and target (Sohrabi, A. Positive and Negative Congruency Effects in Masked and Unmasked Priming: Match of representation strength, Attention, and Consciousness. PhD dissertation, Carleton University, 2008).

The units in all layers (except IL and AA) receive additive Gaussian noise (zero mean and variance σ), intended as general, irrelevant incoming activities. The activations in the model are represented using units with real valued activity levels. The units excite and inhibit each other through weighted connections. Activation propagates through the network when the IL is clamped with input patterns, leading to a final response. As will be described below, the states of units in RL, ML, and CL are adopted in a method similar to a noisy, leaky, integrator algorithm[19-21,25]. These types of models are noisy versions of previous connectionist models[30-32].

In a typical masked trial or epoch, one of the prime units in the IL is turned on and the network is left active for 43 cycles. Then the mask units in IL are turned on for 71 cycles, followed by turning on the target input in IL for 200 cycles. This is similar to a trial in the experiment, except no forward mask is presented, for the sake of simplicity. The prime and target units in the IL are used to represent the stimulus features (i.e., direction or magnitude). However, as mentioned

before, the recognition of the stimuli is not implemented in detail, but is encoded as a binary code. For example, in the case of arrows, 1 is used for the left unit if it points left, or in the case of numerical stimuli, if the number is less than *five*, and 0 is used for the opposite (reciprocal) unit. In the congruent condition, the RL units of the prime and target at the same side (left or right randomly) are turned on (1) or off (0) in each trial at the time of stimulus presentation. By contrast, in the incongruent condition, the two units at the opposite sides are turned on and the other two are left off, with random selection of the two possible cases.
The RL is governed by a modified version of previous models[19-21], which is calculated with discrete integrational time steps using the dynamic equation:

$$X(t+1) = \lambda_x X(t) + (1-\lambda_x)\, f\,[W_{X_iX_i} X_i(t) + W_{X_iI_i} I_i(t) - W_{X_iX_j} X_j(t) - \theta_{X_i} + \xi_{X_i}] \quad (1)$$

Likewise, ML and CL are modelled in a similar way with their inputs coming from RL:

$$Y(t+1) = \lambda_y Y(t) + (1-\lambda_y)\, f\,[W_{Y_iY_i} Y_i(t) + W_{Y_iX_i} X_i(t)) - W_{Y_iY_j} Y_j(t) - \theta_{Y_i} + \xi_{Y_i}] \quad (2)$$

In equations (1) and (2), X and Y denote the activity of units through time t. W is the weight of the connections between units, I is the input, and the subscripts i and j are indexes of the units. The three weight parameters in the brackets correspond to recurrent self-excitation, feed-forward excitation, and lateral inhibition, respectively. However, for the sake of simplicity in equation 1, the lateral excitation from mask units to the prime and target, $W_{X_iX_j} X_j(t)$, and the cross-talk in prime and target to reciprocal units and mask units, $W_{X_iI_j} I_j(t)$, are not present. The term θ is the bias, the term ξ is noise, and f is a sigmoid function (see equation 3 and **Supplementary Fig. 8** online). The term λ represents neural decay[33] which is related to the discrete integrational time steps in the underlying equation[21]. This characterizes neuronal gating with a fast rise followed by a slow decay[31].
The AA modulates other layers by changing their activation from sigmoid toward binary responses[19,22,35]. The activation function, f, transfers the net input, X, of a unit, and modulatory gain, g, to its activity state, implementing the firing rate of a neuron or the mean firing rate of a group of neurons (**Supplementary Fig. 8** online):

$$f(X) = 1/(1+\exp(-Xg)) \quad (3)$$

A conflict-monitoring measurement was employed to take the activations of the units in the CL layer to adjust phasic and tonic response modes of AA. The CL itself may be equivalent to a "convergence zone" [37] that combines different information (here from the prime and target). It may partially involve the temporal lobe, but the conflict is supposed to mostly be involved in the Anterior Cingulate Cortex (ACC) and adjacent prefrontal areas[38]. The activation of the CL units was used to measure the Hopfield energy function between units[39], as used previously[40]. One way to measure conflict is to calculate it as the co-activation of incompatible

representations[40]. So, conflict can be defined as the concurrent activation of the competing units and as the joint effect of both prime and target in CL. Hopfield energy can be calculated as

$$E = -.5\, X^t W X$$
$$= -.5 \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad (4)$$

where E denotes energy, X denotes the activity of a unit, W is the weight of the connection between units, and the subscripts $_1$ and $_2$ are indexes of the two units.

As noted above, CL combines prime and target activations and measures conflict between its two units. When one CL unit is active and the other is inactive, conflict is low. However, when both units are active concurrently, the conflict is high. Instead of directly and dynamically measuring activations in the CL units for measuring conflict[41], those activations are converted to 1 if they are equal to or greater than .5, and to 0 otherwise (i.e., using a threshold function). Also, $E > .5$ is considered as a conflict, otherwise as no conflict. When the activation of a prime or target unit in RL reaches the designated threshold, .62, the AA is activated with a phasic or tonic mode, depending on the absence or presence of conflict in CL. The change in AA response mode usually occurs by the presentation of a target that is incongruent with the prime.

Here the AA is modelled using a reduced or abstracted version of LC neurons in a Willson-Cowan type of system[42] adopted recently[21] (there are similar models and detailed implementations of this type of attention [19-20,22,43]):

$$X(t+1) = \lambda_x X(t) + (1-\lambda_x)\, f\,[c\,(a_x X(t) - bY(t) + I_x(t) - \theta_x)],$$
$$Y(t+1) = \lambda_y Y(t) + (1-\lambda_y)\, f\,[c\,(a_y X(t) - \theta_y)],$$
$$G(t+1) = \lambda_g G(t) + (1-\lambda_g) X(t) \quad (5)$$

where f is again a sigmoid function (as in equation 3), X is the fast variable representing AA activity and Y is a slow auxiliary variable, together simulating excitatory/inhibitory neuron groups in the LC[21]. The X and Y variables have decay parameters λ_x and λ_y, excitatory/inhibitory coefficients, a_x and a_y, as well as thresholds θ_x and θ_y, respectively. The G variable is the output of the AA, which is based on X. The g (used in equation 3) is computed from G: $g = G * K$. The AA modulates other layers when g crosses a threshold, 1. Its activity modes can be phasic or tonic depending on the conflict state, *low* or *high*, respectively. In all conditions the CL can change the AA mode according to the conflict between prime and target (i.e., using within-trial conflict). The phasic and tonic modes of AA responses are implemented using high or low c value (*3* or *1)* (see equation 5). The c value is 3 at the beginning of each trial (for the prime), but it is set to 1 (for the target) if conflict occurs.

The number of computer simulation cycles from the target onset until one of the ML units reached a designated threshold, .62, was considered as RT. A constant, as other sensory and motor processes, could be added to this RT, to increase the match between simulation and human data. The model showed different types of errors including wrong responses (late errors), premature incorrect responses called early errors, premature correct responses called pre-hits

(reaching the threshold before the target presentation), and misses (failing to cross the threshold by the trial deadline). However, to focus on the main idea (i.e., RT result that is most consistent in different studies), the model was set up to not produce these types of error responses frequently (usually about 1%, see Simulation 5, as an example). For all simulations, parameters in **Table 1** were used, and were fixed in all simulations unless otherwise mentioned.

Table 1. *Parameters in the model, fixed for all simulations, unless otherwise mentioned.*

Discussion

The experimental results showed that the distance manipulation on the prime and the target and the target type (well-learned versus recently learned) can have additive effect and change the direction of the priming effects. A numeral primed by another numeral produced a PCE, but only when they both were far from the reference number *five*. Otherwise this did not produce any significant priming effects. By contrast, a symbol, that has been recently learned as number, primed by a numeral produced an NCE when they were both close to the reference number *five*. Previous studies have shown change from PCE to NCE by increasing mask-target SOA[4-14], but in the present study, this change in priming effect direction was found by decreasing the representation strength of the prime and target, still with a fixed and relatively short mask-target SOA.

Consistent with the present human data, the model showed both PCE and NCE by manipulating specific factors. The main factor, the strength of stimulus representation, was implemented simply by changing the weights between input and representation units (i.e., IL to RL); the weights between representation and other layers remained fixed. When the prime and target were easy (as with higher prime and target strength in Simulation 1), they could be processed primarily with the initial activation and attentional response. When they were harder (as with lower prime and target strength in Simulation 1) or when a delay was introduced between them (as with longer mask-target SOA in Simulation 2), the second phase of attention (for the target) was not strong enough to activate the target quickly. This happened because attention showed a phasic response with a refractory period. The conflict was measured based on the incongruency in the stimuli relationship. It decreased the effect of the refractory period by putting the second phase of attention (to the target) in a tonic mode, enhancing the processing of the incongruent trials where conflict occurred. This was not the case in the congruent trials. When both prime and target were hard (i.e., by having low strength weights) the priming pattern was inverted (i.e., NCE). In other cases when prime or target or both were easy (having higher strength weights), the priming was PCE. The NCE found in previous studies[4-14] was simulated by increasing mask-target SOA, with no other changes in the model. A PCE and an NCE were found with short and long mask-target SOA, respectively.

The model also showed the effects of other factors on priming directions such as prime duration, stimulus degradation, and mask density. For example, a prime with longer duration and less degradation has a strong representation that causes a large NCE if the target comes late (and a large PCE if it comes early). The model also showed that decreasing the activation of input units

(e.g., from binary, 1 and 0, to real normalized numbers, .9 and .1, or less, for simulating stimulus degradation) turns NCE into PCE. This supports the idea that the NCE is not caused merely by a decrease in the incoming perceptual information but by a decrease in the representation strength. The involvement of an attentional bottleneck in the decisional rather than perceptional processes has been proposed previously[44]. While the previous model of PCE and NCE[16] showed an effect of strength, it differs from the one here. First, it only dealt with prime strength. Second, the strength was simulated by the amount of sensory information. The only evidence provided for that model is the degradation experiment[5], so the high degradation was treated as low strength. In the current model, degradation was simulated by less incoming information (and higher noise) and the strength was implemented by connection weights between input and representation layers for both prime and target.

The current model, in addition to being more biologically compelling, showed many dynamic effects in RT and error patterns that have not been shown previously (such as the changes in RT and the size of priming effects through time). While in the current model the NCE disappeared and became a very small PCE at very long SOAs, the previous model[16] showed a huge PCE at very long SOAs inconsistent with human data[14]. The present model is similar to some other previous neuro-computational models, especially those employed to simulate the attentional blink[22,45]. In these models blink for the second target occurs at lag 2 (after 100 ms from the first target) and no blink occurs at lag 1 (if the second target is presented during 100 ms after the first target), related to NCE and PCE in the current model, respectively.

The simple way we chose to implement conflict was not meant to represent all executive functions. However, its two-state nature, while simple, is relevant to the binary or rule-based processes of executive functions[46]. The role of conflict has been supported by the previous studies[40,47]. Furthermore, the dynamic response in the model is consistent with the ERP result of the motor cortex[15], that shows an earlier dominance of neural activation in the congruent condition (reflected as PCE) and a later dominance of the incongruent condition (reflected as NCE, see also **Supplementary Fig. 9** online). A change in priming effects has been reported in diseases such as schizophrenia[47] and Parkinson's[48].

1. Marcel, A.J. Conscious and unconscious perception: experiments on visual masking and word recognition. *Cognitive Psychology* **15**, 197-237 (1983).
2. Neumann, O. & Klotz, W. Motor responses to unreportable, masked stimuli: Where is the limit of direct motor specification. In C. Umiltà and M. Moscovitch (Eds.), *Attention and Performance XV: Conscious and non-conscious information processing* (pp. 123-150). Cambridge: MIT Press (1994).
3. Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P. F. & Le Bihan, D. Imaging unconscious semantic priming. *Nature* **395**, 597-600 (1998).
4. Schleghecken, F. & Eimer, M. A central-peripheral asymmetry in masked priming. *Perception & Psychophysics* **62**, 1367–1382 (2000).
5. Schleghecken, F. & Eimer, M. Motor activation with and without inhibition: Evidence for a threshold mechanism in motor control. *Perception & Psychophysics* **64***,* 148–162 (2002).

6. Schleghecken, F. & Eimer, M. Active masks and active inhibition: A comment on Lleras and Enns (2004) and on Verleger, Jaśkowski, P, Aydemir, van der Lube, and Groen (2004). *Journal of Experimental Psychology: General* **135***,* 484–494 (2006).
7. Eimer, M. Facilitatory and inhibitory effects of masked prime stimuli on motor activation and behavioural performance. *Acta Psychologica* **101**, 293–313 (1999).
8. Eimer, M. & Schleghecken, F. Effects of masked stimuli on motor activation: Behavioural and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance* **24**, 1737-1747 (1998).
9. Eimer, M. & Schlaghecken, F. Response facilitation and inhibition in manual, vocal, and oculomotor performance: evidence for a modality-unspecific mechanism. *Journal of Motor Behaviour* **33**, 16–26 (2001).
10. Eimer, M. & Schleghecken, F. Links between conscious awareness and response inhibition: evidence from masked priming. *Psychonomic Bulletin & Review* **9**, 514–520 (2002).
11. Lleras, A. & Enns, J.T. Negative compatibility or object updating? A cautionary tale of mask-dependent priming. *Journal of Experimental Psychology: General* **133***,* 475–493 (2004).
12. Lleras, A. & Enns, J.T. How much like a target can a mask be? Geometric, spatial, and temporal similarity in priming. A reply to Schleghecken & Eimer (2006). *Journal of Experimental Psychology: General* **135**, 495–500 (2006).
13. Verleger, R., Jaśkowski, P., Aydemir, A., van der Lubbe, R. & Groen, M. Qualitative differences between conscious and nonconscious processing? On inverse priming induced by masked arrows. *Journal of Experimental Psychology: General* **133***,* 494–515 (2004).
14. Jaśkowski, P. & Ślósarek, M. How important is a prime's gestalt for subliminal priming? *Consciousness and Cognition* **16**, 485-497 (2007).
15. Eimer, M. & Schleghecken, F. Response facilitation and inhibition in subliminal priming. *Biological Psychology* **64**, 7–26 (2003).
16. Bowman, H., Schleghecken, F. & Eimer, M. A neural network model of inhibitory processes and cognitive control. *Visual Cognition* **13**, 401–480 (2006).
17. Gallistel, C.R. & Gelman, R. Preverbal and verbal counting and computation. *Cognition* **44**, 43–74 (1992).
18. Dehaene, S., Bossini, S. & Giraux, P. The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General* **122**, 371-396 (1993).
19. Gilzenrat, M. S., Holmes, B. D., Rajkowski, J., Aston-Jones, G. & Cohen, J.D. Simplified dynamics in a model of noradrenergic modulation of cognitive performance. *Neural Networks* **15***,* 647–663 (2002).
20. Usher, M., Cohen, J.D., Servan-Schreiber, D., Rajkowski, J. & Aston-Jones, G. The role of locus coeruleus in the regulation of cognitive performance. *Science* **283***,* 549–554 (1999).
21. Usher, M. & Davelaar, E.J. Neuromodulation of decision and response selection. *Neural Networks* **15**, 635–645 (2002).

22. Nieuwenhuis, S., Aston-Jones, G. & Cohen, J.D. Decision making, the P3, and the locus coeruleus–norepinephrine system. *Psychological Bulletin* **131**, 510–532 (2005).
23. Chambers, C. D. and Brown, M. Timing accuracy under Microsoft Windows revealed through external chronometry, *Behavior Research Methods, Instruments, & Computers, **35**, 1, 96-108* (2003).
24. Brown, E. & Holmes, P. Modelling a simple choice task: stochastic dynamics of mutually inhibitory neural groups. *Stochastics and Dynamics* **1**, 2, 159–191 (2001).

25. Usher, M. & McClelland, J.L. The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review* **108**, 550–592 (2001).
26. Wong, K-F. & Wang, X-J. A Recurrent Network Mechanism of Time Integration in Perceptual Decisions, *The Journal of Neuroscience* **26**, 4, 1314–1328 (2006).
27. Aston-Jones, G. & Cohen, J.D. An Integrative Theory of Locus Coeruleus-Norepinephrine Function: Adaptive Gain and Optimal Performance, *Nature Review Neuroscience* **28**, 403-450 (2005).
28. Wiesstein, N. A Rashevsky-Landahl neural net: Simulation of metacontrast. *Psychological Review* **75**, 6, 494-521 (1968).
29. Rolls, E.T. & Tovee, M.J. Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society of London Series B,* **257**, 9-15 (1994).
30. Anderson, J.A., Silverstein, J.W., Ritz, S.A. & Jones, R.S. Distinctive features, categorical perception and probability learning: Some applications of a neural model. *Psychological Review* **84**, 413–451 (1977).
31. McClelland, J.L. On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review* **86**, 287–330 (1979).
32. McClelland, J.L. & Rumelhart, D.E. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review **88**,* 375-407 (1981).
33. Amit, D.J. & Tsodyks, M. Quantitative study of attractor neural network retrieving at low spike rates: I substrate—Spikes, rates and neuronal gain. *Network* **2**, 259–273 (1991).
34. Wang, X-J. Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *Journal Neuroscience* **19**, 9587–9603 (1999).
35. Servan-Schreiber, D., Printz, H. & Cohen, J.D. A network model of catecholamine effects gain signal to noise ratio and behaviour. *Science* **249**,892–895 (1990).
36. Cohen, J.D., Dunbar, K. & McClelland, J.L. On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review* **97**, 332-361 (1990).
37. Damasio, A.R. Synchronous activation in multiple cortical regions: A mechanism for recall. *Seminars in Neurosciences* **2**, 287-296 (1990).
38. Botvinick, M.M., Nystrom, L.E., Fissell, K., Carter, C.S. & Cohen, J.D. Conflict monitoring versus selection for action in anterior cingulate cortex. *Nature* **402**, 179–181 (1999).

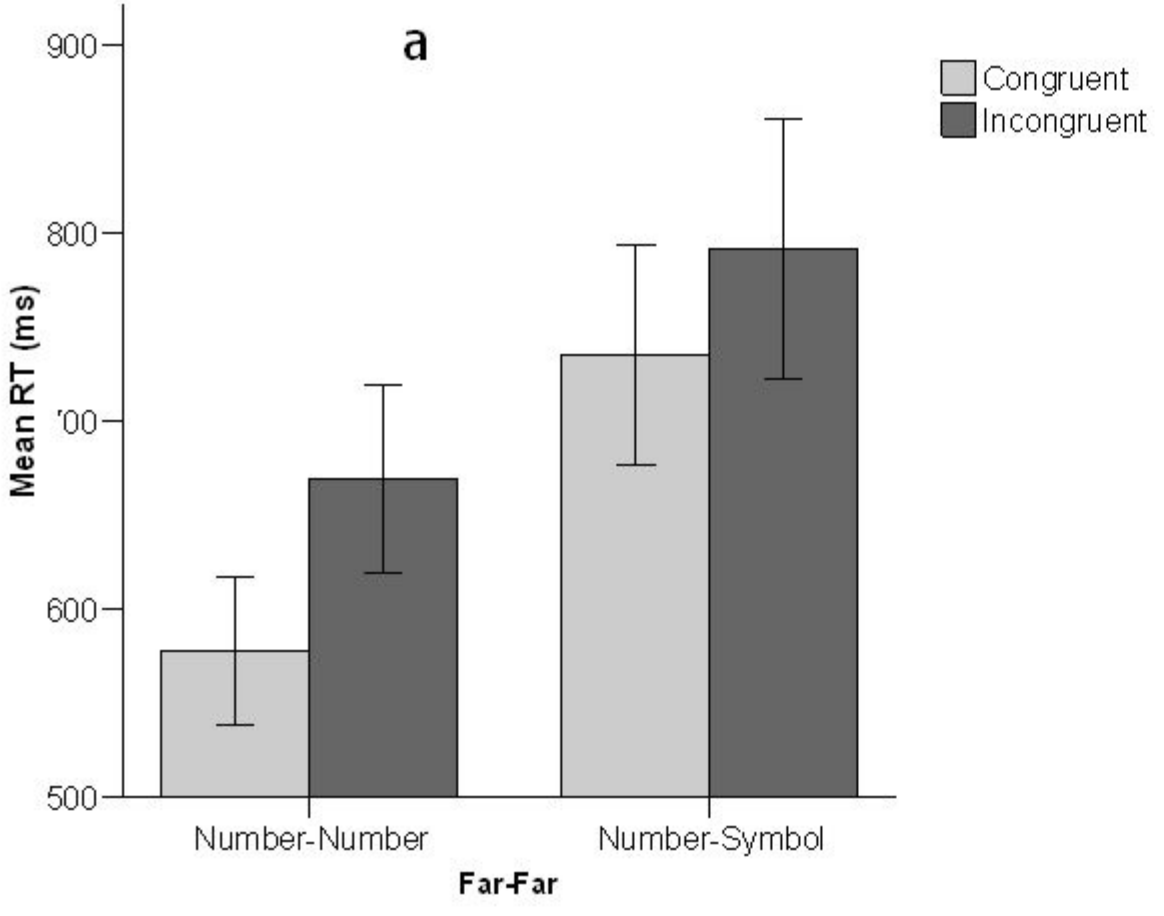
39. Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA* **79***,* 2554-2558 (1982).
40. Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S. & Cohen, J.D. Conflict monitoring and cognitive control, *Psychological Review* **108**, 624-652 (2001).
41. McClure, S.M., Gilzenrat, M.S. & Cohen, J.D. An exploration-exploitation model based on norepinepherine and dopamine activity. *Advances in Neural Information Processing Systems **18*** (2005).
42. Wilson, H. & Cowan, J. Excitatory and inhibitory interactions in localized populations of model neurons. *Biological Cybernetics* **12**, 1–24 (1972).
43. Brown, E., Moehlis, J., Holmes, P., Clayton, E., Rajkowski, J. & Aston-Jones, G. The influence of spike rate and stimulus duration on noradrenergic neurons. *Journal of Computational Neuroscience* **17***,* 13–29 (2004).
44. Sigman, M. & Dehaene, S. Parsing a cognitive task: a characterization of the mind's bottleneck. *PLoS Biology* **3**, 2, e37 (2005).
45. Mathis, W.D. & Mozer, M.C. Conscious and unconscious perception: a computational theory. In: *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates, Hillsdale, N.J., 324–328 (1996).
46. O'Reilly, R.C. Biologically Based Computational Models of High-Level Cognition. *Science* **314***,* 91-94 (2006).
47. Dehaene, S., Artiges, E., Naccache , L., Martelli, C., Viard, A., Schürhoff, F., Recasens, C., Martinot, M.L.P., Leboyer, M. & Martinot, J. Conscious and subliminal conflicts in normal subjects and patients with schizophrenia: the role of the anterior cingulated, *Proceedings of the National Academy of Sciences of the United States of America* **100**, 13722-13727 (2003).
48. Seiss, E. & Praamstra, P. The basal ganglia and inhibitory mechanisms in response selection: Evidence from subliminal priming of motor responses in Parkinson's disease. *Brain* **127**, 330–339 (2004).

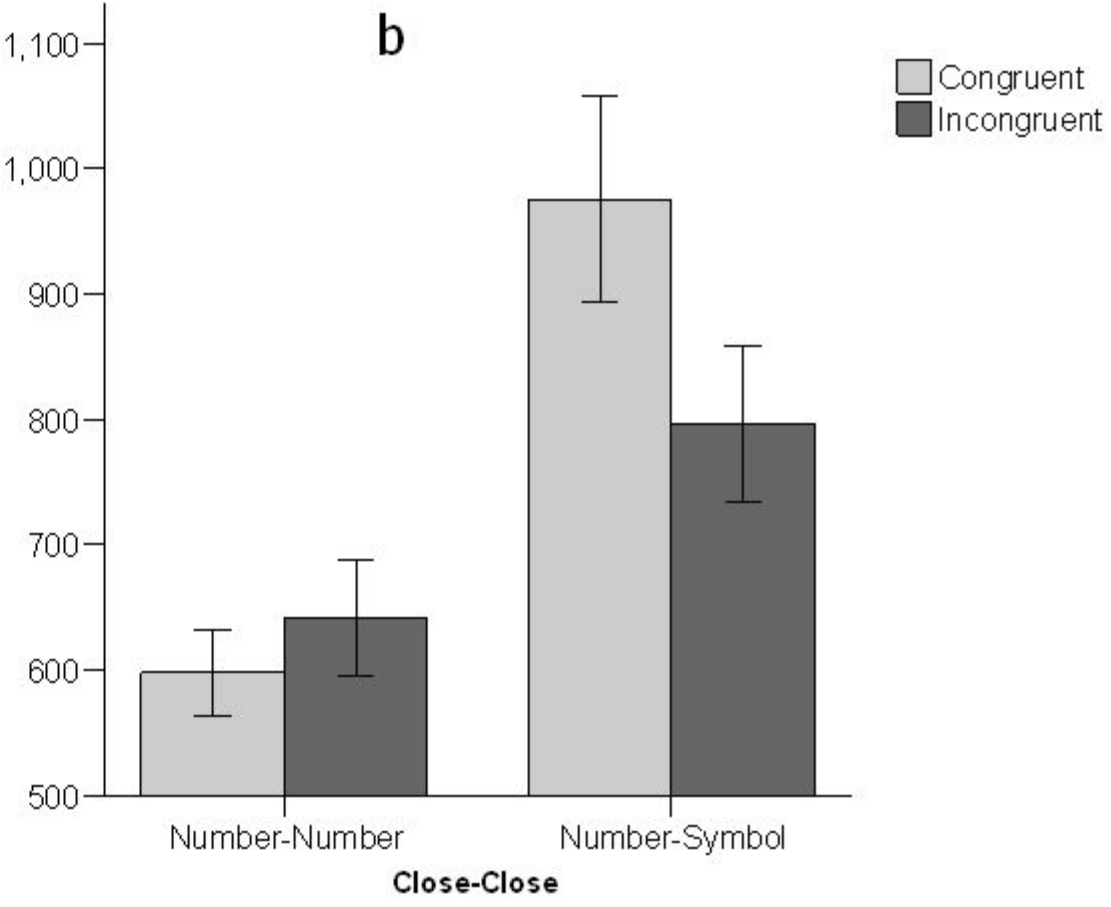
Table 1. *Parameters in the model, fixed for all simulations, unless otherwise mentioned.*

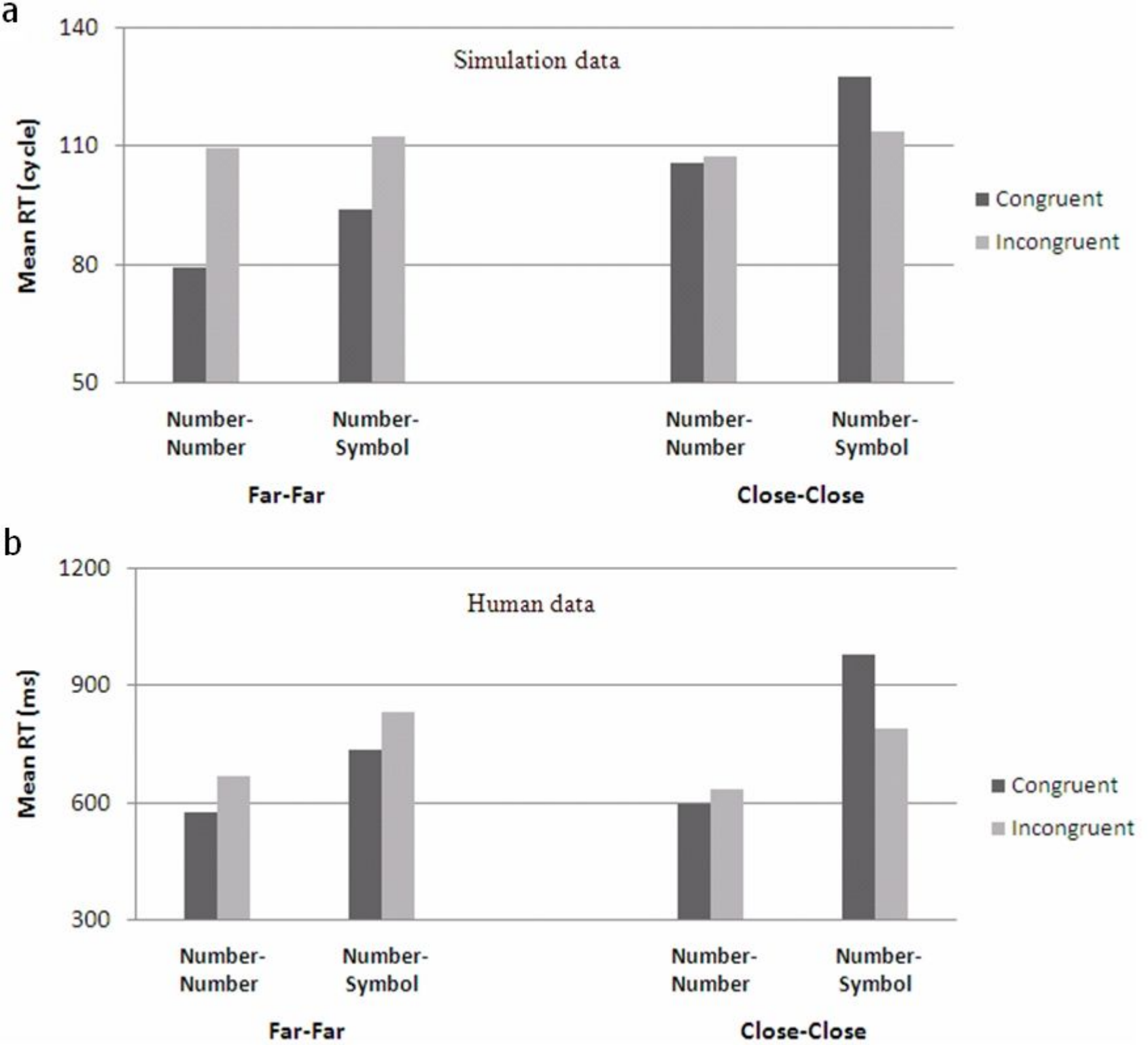
WX_iI_i (IL to RL) [P & T] & WY_iX_i (RL to ML) [P & T]	2-3 & 1.5
WX_iI_i (IL to RL) [M] & WY_iX_i (RL to CL) [P & T]	1.5 & 1
WX_iX_i (RL) [P & T], WX_iX_i (RL) [M], WY_iY_i (CL), & WY_iY_i (ML)	1.5, 1.25, 1, & .9
WX_iX_j (RL) & WY_iY_j (ML & CL)	1 & 1
WX_iX_j (RL) [M to P & T] & WX_iI_j (IL to RL)	.75 & .33
K (AA)	4.52
θ_x, θ_y (AA), θ_x (RL), θ_y (CL), & θ_y (ML)	1.25, 1.5, .5, .85, & 2
b,c, a_x & a_y (AA)	4, 1-3, 2, & 3

λ_x, λ_g, & λ_y (AA)	.92, .98, & .996
λ (CL), λ (ML), & λ (RL)	.75, .925, & .95
σ (CL), σ (RL) [P & T], σ (ML) & σ (RL) [M]	.025, .2, .25, & 1.25

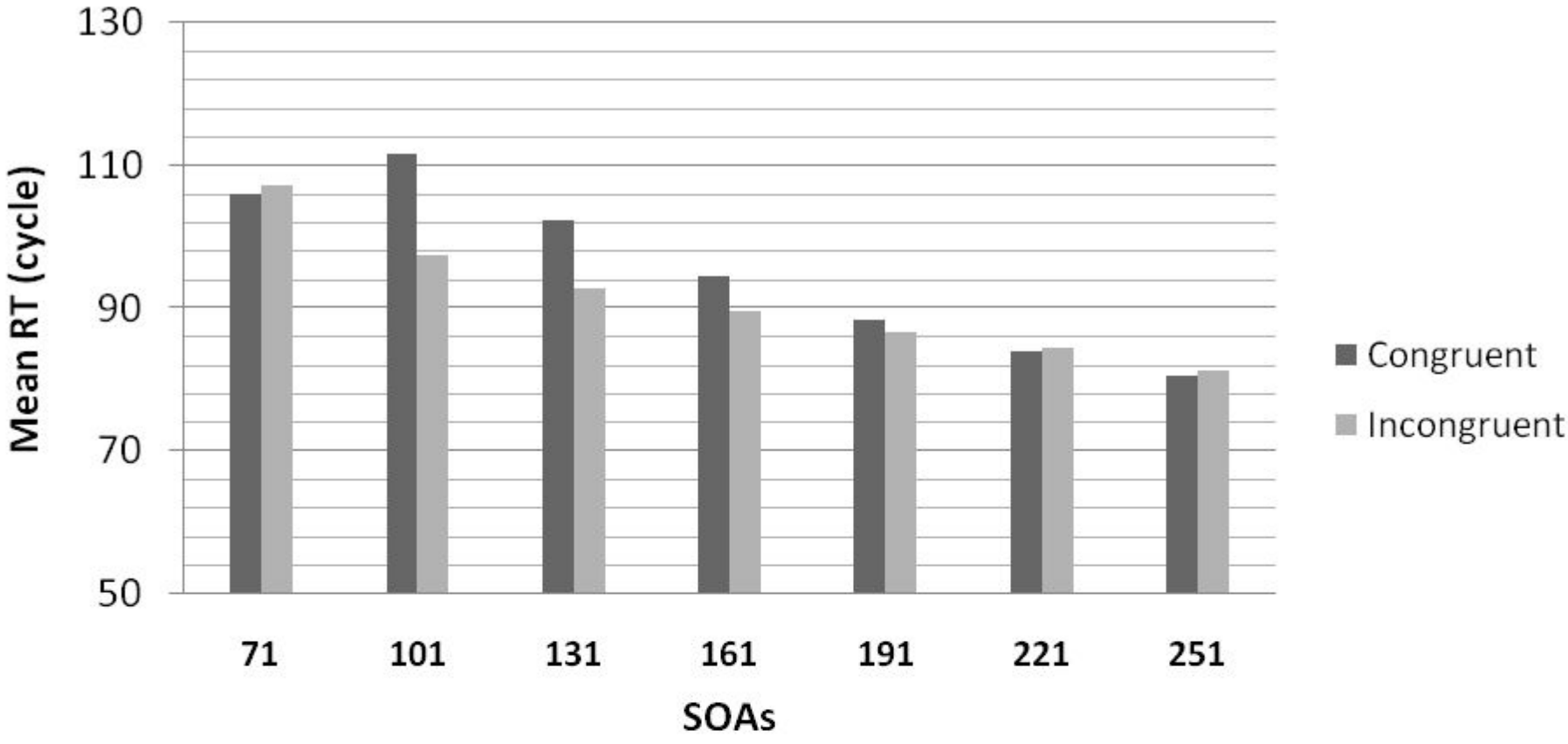
IL=Input Layer; RL=Representation Layer; CL= Cognitive Layer; ML=Motor Layer; AA=Alert Attention; P=Prime; T=Target; M=Mask.


a
900
800
'00
600
500
Mean RT (ms)
Congruent
Incongruent
Number-Number
Number-Symbol
Far-Far
Error Bars: +/- 2 SE

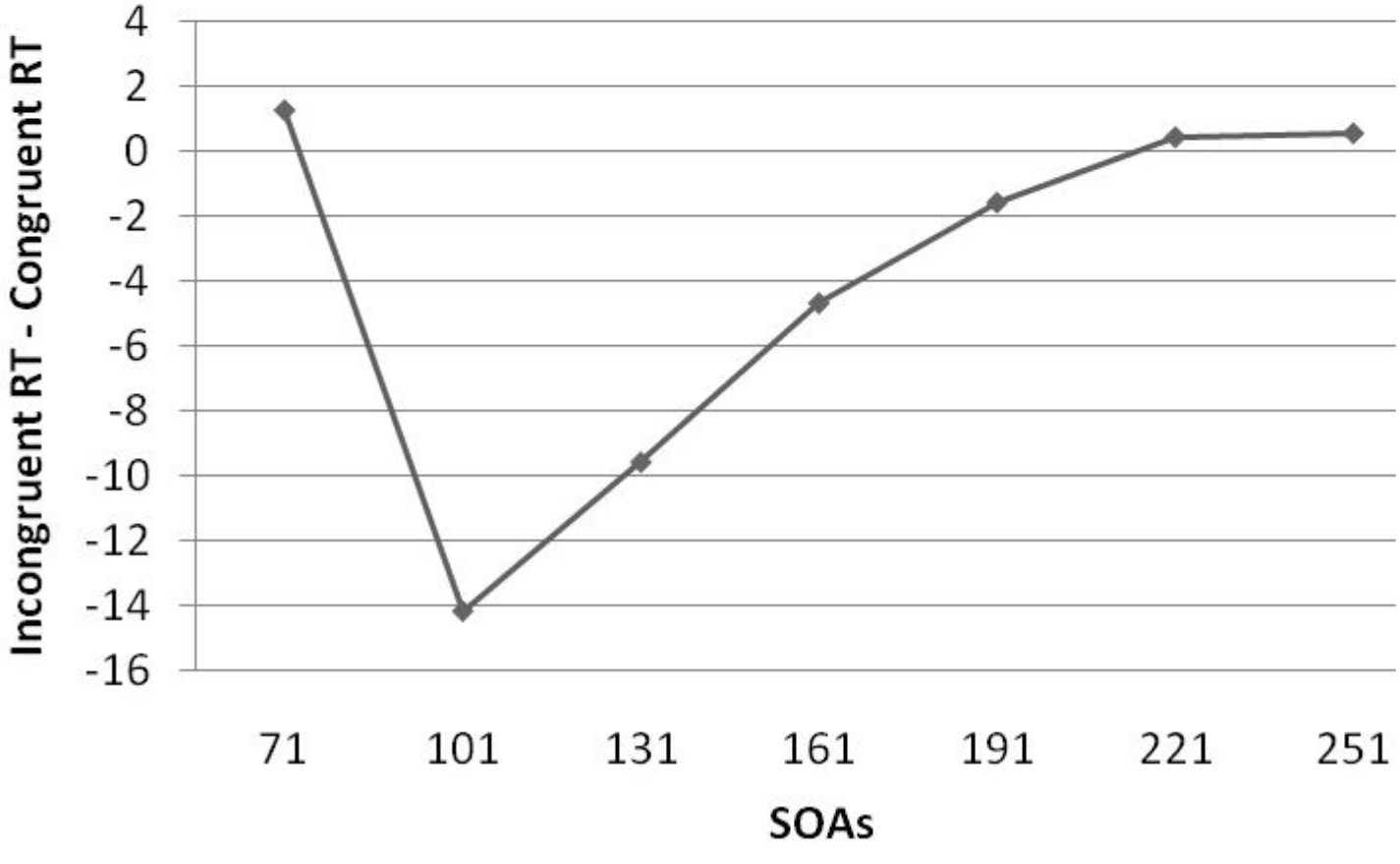

b
1,100
1,000
900
800
700
600
500
Congruent
Incongruent
Number-Number
Number-Symbol
Close-Close
Error Bars: +/- 2 SE

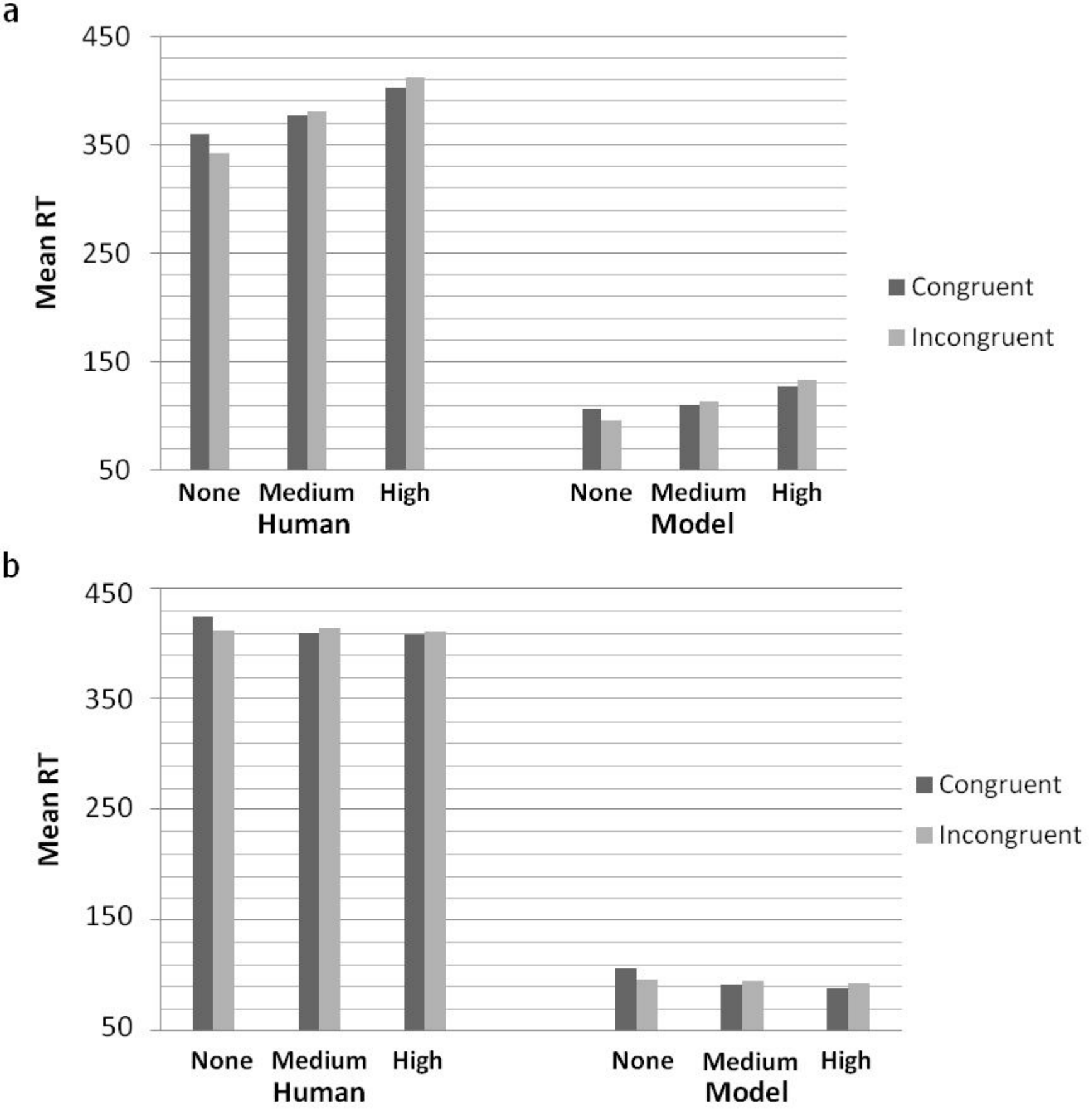

a
Simulation data
Mean RT (cycle)
140
110
80
50
Congruent
Incongruent
Number-Number
Number-Symbol
Number-Number
Number-Symbol
Far-Far
Close-Close
b
Human data
Mean RT (ms)
1200
900
600
300
Congruent
Incongruent
Number-Number
Number-Symbol
Number-Number
Number-Symbol
Far-Far
Close-Close

a

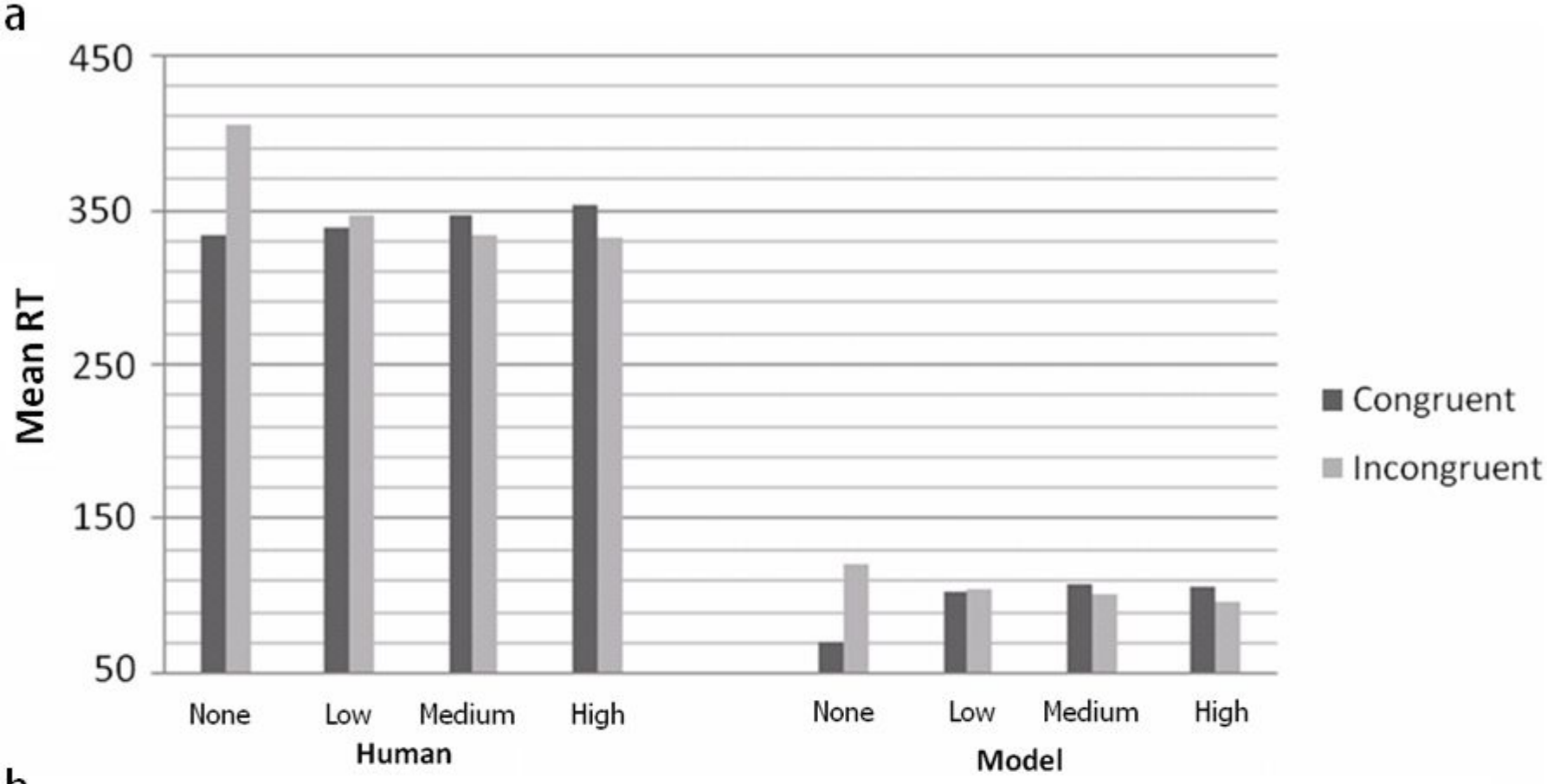

130
110
90
70
50
Mean RT (cycle)
71
101
131
161
191
221
251
SOAs
Congruent
Incongruent

b

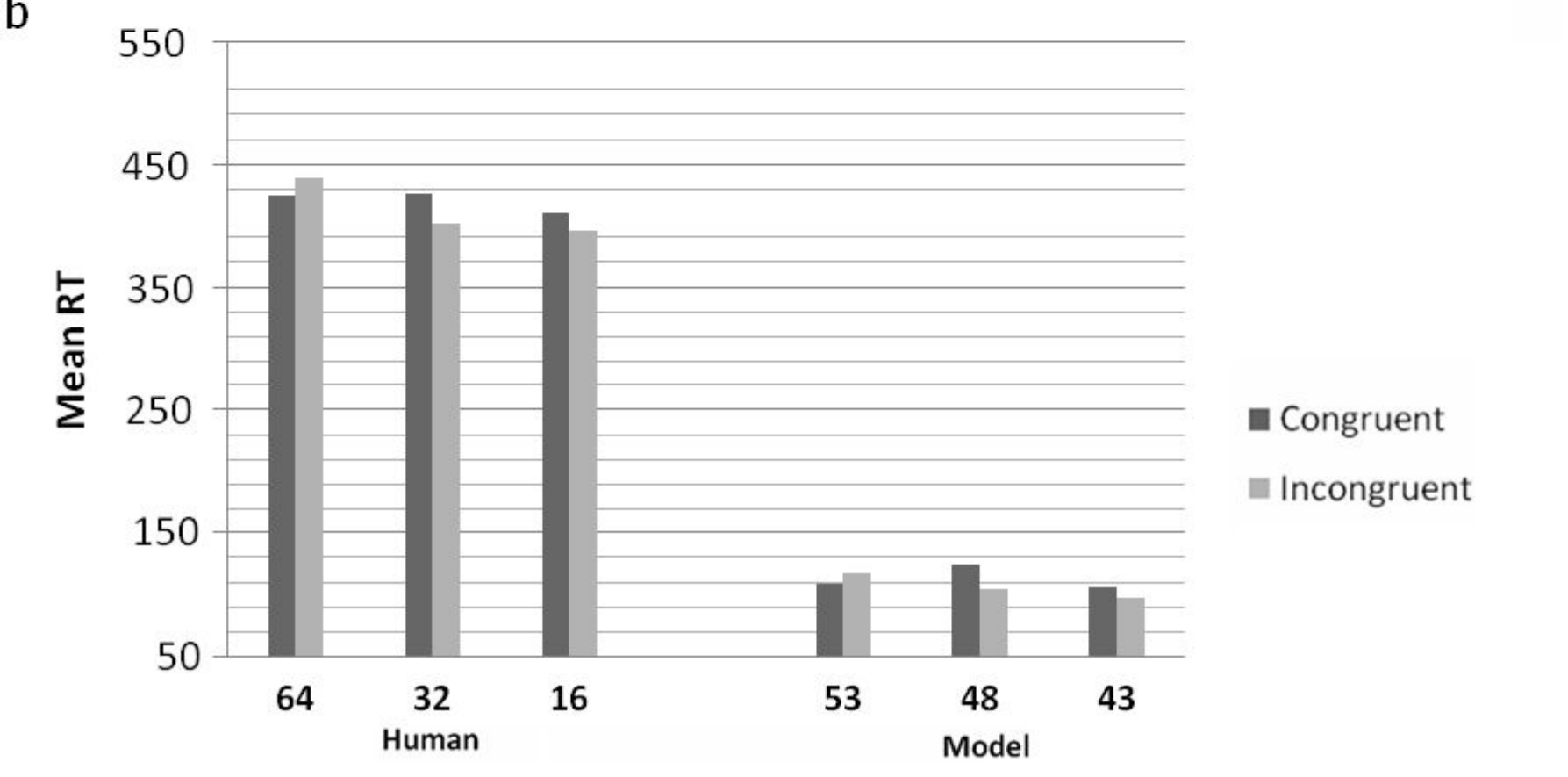

4
2
0
-2
-4
-6
-8
-10
-12
-14
-16
Incongruent RT - Congruent RT
71
101
131
161
191
221
251
SOAs

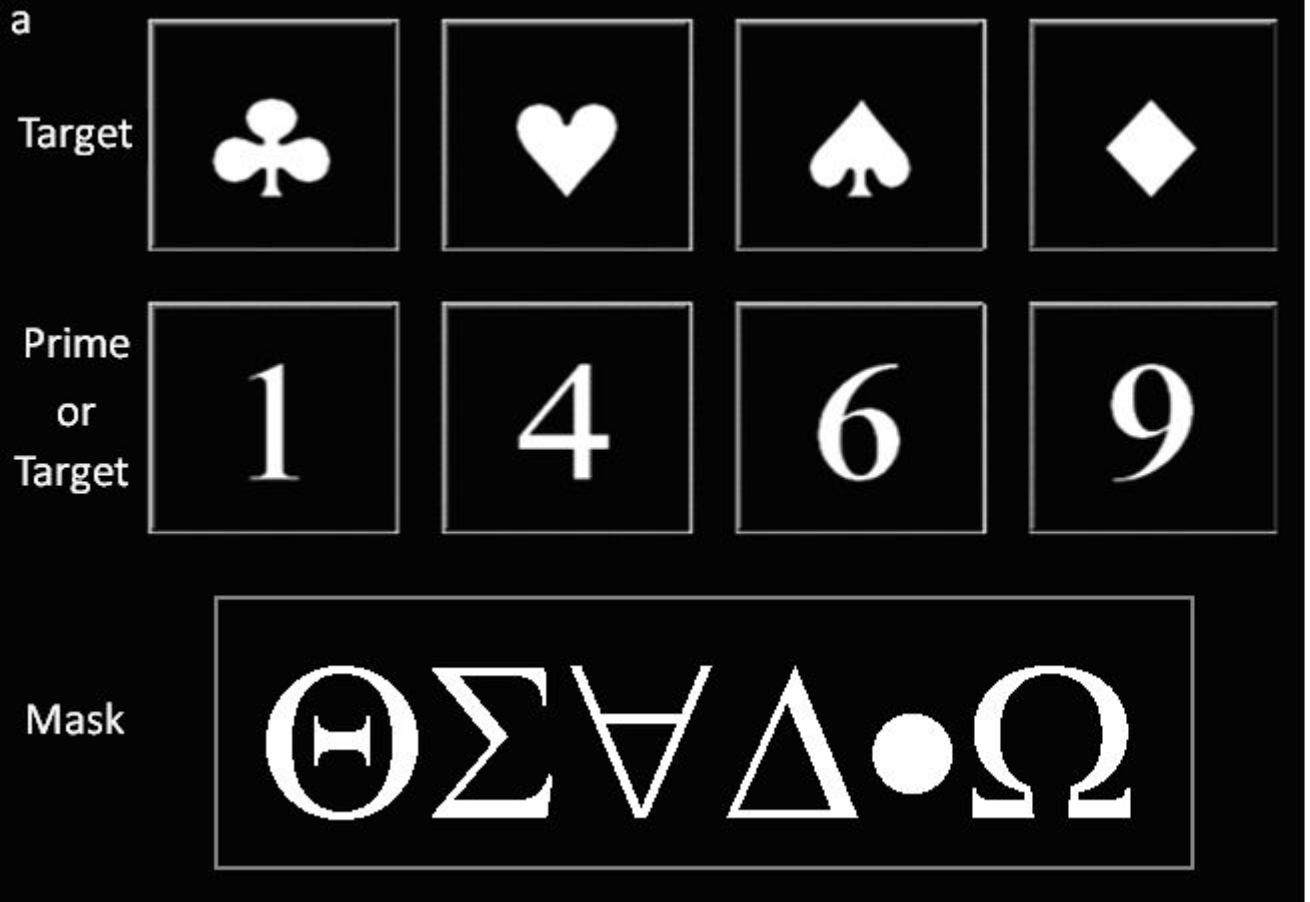

a
450
350
250
150
50
Mean RT
Congruent
Incongruent
None
Medium
High
Human
None
Medium
High
Model
b
450
350
250
150
50
Mean RT
Congruent
Incongruent
None
Medium
High
Human
None
Medium
High
Model

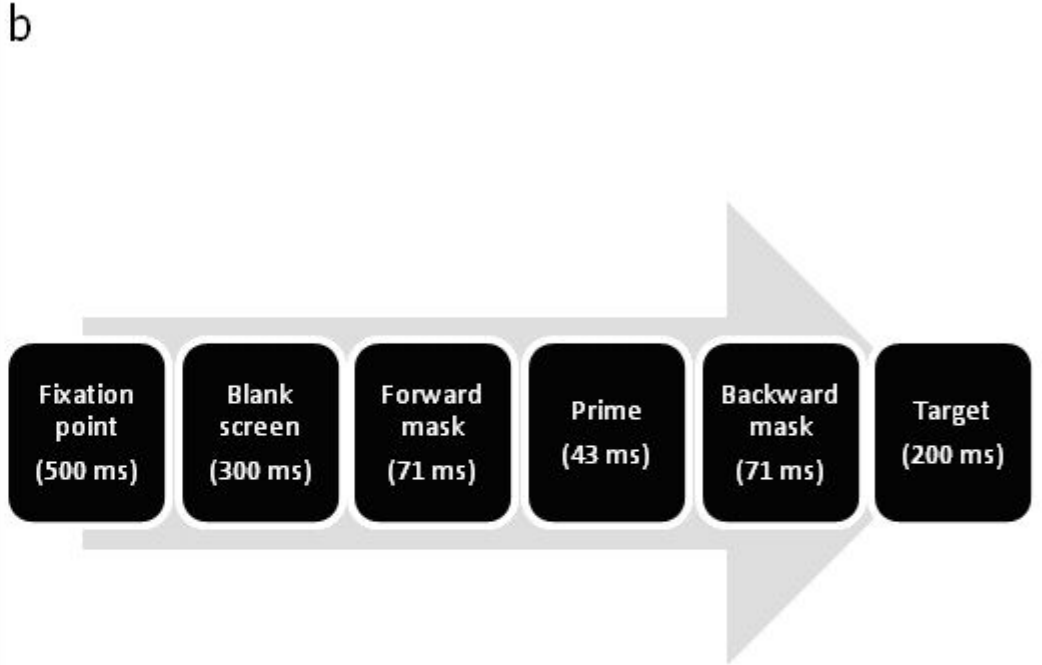
a

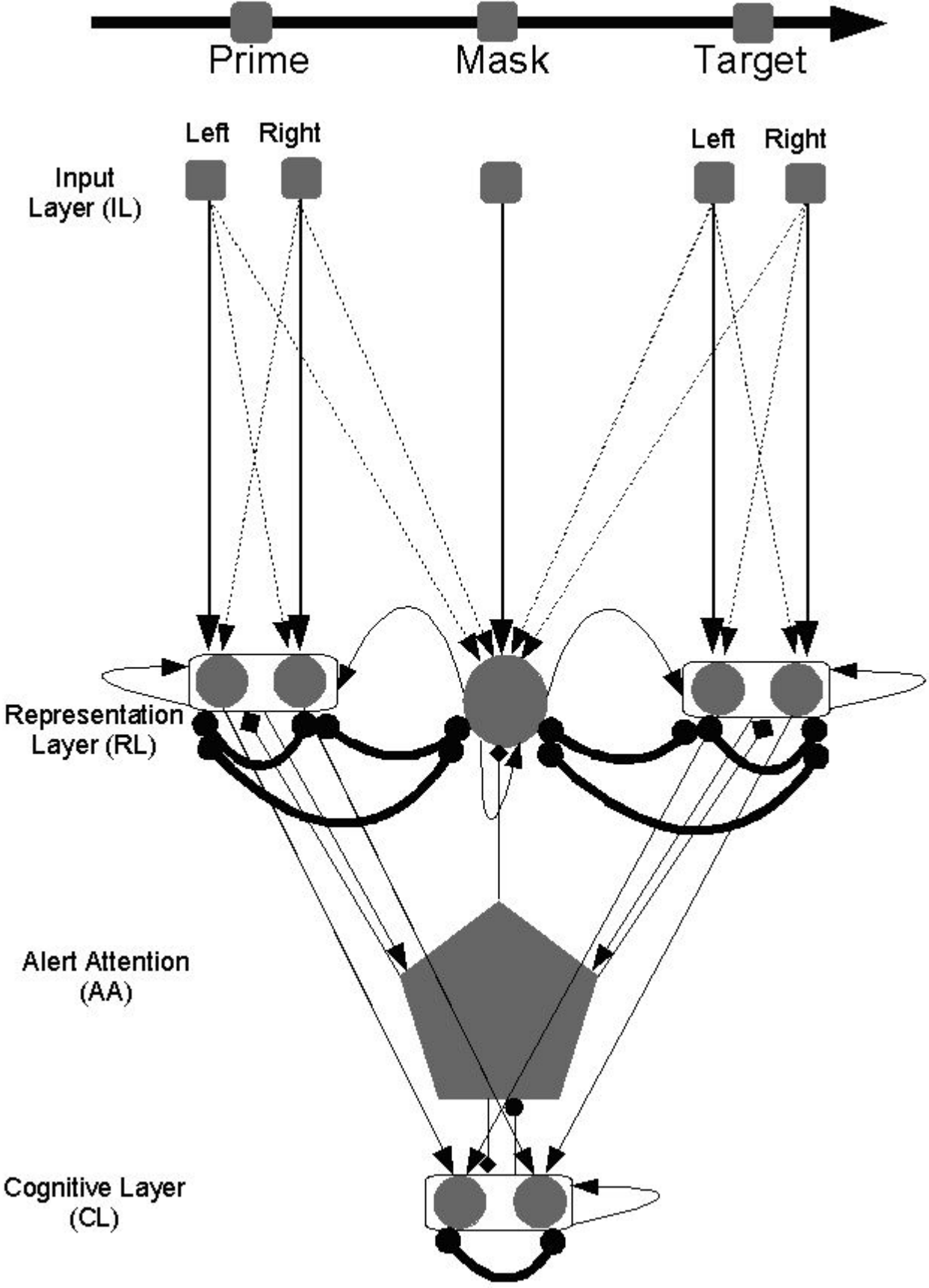

450
350
250
150
50
Mean RT
None
Low
Medium
High
Human
None
Low
Medium
High
Model
Congruent
Incongruent

b


550
450
350
250
150
50
Mean RT
64
32
16
Human
53
48
43
Model
Congruent
Incongruent


a
Target
Prime or Target
1
4
6
9
Mask


b
Fixation point (500 ms)
Blank screen (300 ms)
Forward mask (71 ms)
Prime (43 ms)
Backward mask (71 ms)
Target (200 ms)


Prime
Mask
Target
Left
Right
Left
Right
Input
Layer (IL)
Representation
Layer (RL)
Alert Attention
(AA)
Cognitive Layer
(CL)