Automatic Imitation of Multiple Agents: A Computational Model*

Emiel Cracco
Ghent University

Richard P. Cooper
Birkbeck, University of London

Abstract

There is accumulating evidence that the actions of others are represented in the motor system, leading to automatic imitation. However, whereas early work focused mainly on the effects of observing a single agent, recent studies indicate that the actions of multiple agents can be represented simultaneously. Yet, theorizing has lagged behind. The current study extends the dual-route model of automatic imitation to include multiple agents, and demonstrates, in five simulation studies, that the extended model is able to capture four critical multi-agent effects. Importantly, however, it was necessary to augment the model with a control mechanism regulating response inhibition based on the number of observed actions. Furthermore, additional simulation indicated that this mechanism could be driven by response conflict. Together, our results demonstrate how theories of automatic imitation can be extended from single- to multi-agent settings. As such, they constitute an important step towards a mechanistic understanding of social interaction beyond the dyad.

Keywords: automatic, imitation, multiple agents, simulation, modeling

1. Introduction

There is a well-documented tendency to imitate in humans (Cracco, Bardi, et al., 2018; Heyes, 2011). In social psychology, this is studied by means of *motor mimicry* – the finding that interacting individuals spontaneously imitate each other's facial expression, posture, and mannerisms (Chartrand & Bargh, 1999; Chartrand & Lakin, 2013; Genschow et al., 2017). In cognitive psychology, it is studied by means of *automatic imitation* – the finding that observing congruent actions facilitates, whereas observing incongruent actions impedes, unrelated task responses (Brass et al., 2000; Cracco, Bardi, et al., 2018; Heyes, 2011; Sturmer, Aschersleben, & Prinz, 2000). In one of the first automatic imitation studies, Brass et al. (2000) instructed subjects to lift their index or middle finger in response to the number 1 or 2 while a hand on the screen also lifted its index or middle finger. The results showed that participants responded more quickly when the observed action matched the instructed response (i.e., congruent trial) than when it did not (i.e., incongruent trial). This congruency effect is a measure of imitative response tendencies (Cracco & Brass, 2019), and has now been reliably observed across a wide range of effectors (Cracco, Bardi, et al., 2018), including not only fingers (Butler, Ward, & Ramsey, 2015; Genschow et al., 2017), but also hands (Cracco, Genschow, Radkova, & Brass, 2018; Sturmer et al., 2000), feet (Gillmeister, Catmur, Liepelt, Brass, & Heyes, 2008; Wiggett, Downing, & Tipper, 2013), arms (Kilner, Paulignan, & Blakemore, 2003; Stanley, Gowen, & Miall, 2007), and faces (Leighton & Heyes, 2010; Press, Richardson, & Bird, 2010).

An influential framework to interpret automatic imitation is the dual-route framework (Heyes, 2011), which assumes that the imperative cue and the stimulus movement operate via different routes. In particular, the cue is thought to activate short-term intentional stimulus-response (S-R) associations

* Emiel Cracco: Department of Experimental Psychology, Ghent University. Richard P. Cooper: Department of Psychological Sciences, Birkbeck, University of London.

This work was supported by the Research Foundation Flanders Grant FWO16/ASP_H/050 awarded to the first author. The code for all simulations is available at https://osf.io/kezyt/.

Correspondence concerning this article should be addressed to Emiel Cracco, Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000. E-mail: emiel.cracco@ugent.be

established solely for the purpose of the task, whereas the stimulus movement instead activates long-term automatic S-R associations established as a consequence of genetic disposition or learning experience (Heyes, 2011). When both routes converge on the correct response, motor activation is increased, and response facilitation occurs. In contrast, when the automatic route activates the incorrect response instead, this response has to be inhibited, and response interference occurs.

Following earlier models of S-R compatibility (Zhang, Zhang, & Kornblum, 1999; Zorzi & Umilta, 1995), recent studies have started to use connectionist modeling to test the dual-route model of automatic imitation (Cooper, Catmur, & Heyes, 2013; Cooper, Cook, Dickinson, & Heyes, 2013). These studies revealed that the dual-route model provides an accurate description of the empirical effects (Cooper, Catmur, et al., 2013). Moreover, they showed that automatic imitation – also known as imitative compatibility – and spatial compatibility can be modeled using the same mechanism, and that the observed differences in their time course (Catmur & Heyes, 2011) can be explained by variations in the processing time of both stimulus dimensions (Cooper, Catmur, et al., 2013). Finally, by comparing different mechanisms, computational research has supported the idea that automatic imitation is a consequence of associative learning (Cooper, Cook, et al., 2013). However, current models can accommodate just one agent. In contrast, in daily life, social situations often involve multiple agents acting together. Given that the mechanism linking perception to action has an important social function, by supporting action understanding (Catmur, Thompson, Bairaktari, Lind, & Bird, 2017) and facilitating interpersonal coordination (Colling, Knoblich, & Sebanz, 2013; Wilson & Knoblich, 2005), a key question is how this mechanism contributes to social situations extending beyond the dyad. In other words, is it possible to represent the actions of multiple agents at the same time in the motor system?

Pointing in this direction, research looking at automatic imitation of multiple agents has shown that seeing two hands performing the same action elicits stronger corticospinal excitability (Cracco, De Coster, Andres, & Brass, 2016) and therefore stronger imitation (Cracco & Brass, 2018a; Cracco, De Coster, Andres, & Brass, 2015) than seeing one hand performing a single action. Similarly, research using four stimulus hands has shown that automatic imitation increases proportionally with the number of identical observed actions (Cracco & Brass, 2018c). However, this effect was asymmetrical, with response speed increasing linearly on incongruent trials but decreasing asymptotically on congruent trials. In contrast, when no control was needed, namely in the absence of incongruent trials, the congruent asymptote disappeared (Cracco & Brass, 2018c). This change was accompanied by overall faster responses, suggesting it was not caused by a physical boundary but was instead better explained by dynamic control processes regulating motor inhibition based on the number of observed actions (Cracco & Brass, 2018c). That is, if imitative tendencies increase with group size, then imitative control has to increase accordingly to prevent overt imitation. One way to achieve this is to flexibly lower or raise the response threshold based on the number of observed actions so that response selection becomes slower and more deliberate when more hands perform an action. This slowdown counteracts the decrease in response times on congruent trials but strengthens the corresponding increase on incongruent trials, hence explaining the asymmetrical pattern. In contrast, when there are no incongruent trials, imitative control is no longer needed, and the asymptote disappears (Cracco & Brass, 2018c).

In addition to the above research on automatic imitation of multiple identical actions, two studies also looked at the effects of observing two or more different actions (Cracco & Brass, 2018b; Cracco et al., 2015). The first study found that response times when two hands performed two different actions, one congruent and one incongruent, could not be distinguished from response times when neither hand performed an action. It was argued that seeing one congruent and one incongruent action produced a concurrent facilitation and interference effect that subsequently canceled out each other (Cracco et al., 2015). This interpretation was confirmed by a second study with four stimulus hands, showing weaker imitation when one hand performed a different action than the other three hands, compared with when three or four hands all performed the same action (Cracco & Brass, 2018b). Together, this research indicates that the motor system is able to simultaneously represent multiple observed actions, regardless of whether these actions are identical or different actions.

Nevertheless, it remains an open question whether existing theories of automatic imitation can account for these multi-agent imitation effects. Here, we use connectionist modeling to address this question. More specifically, we extend the dual-route model of automatic imitation to multi-agent

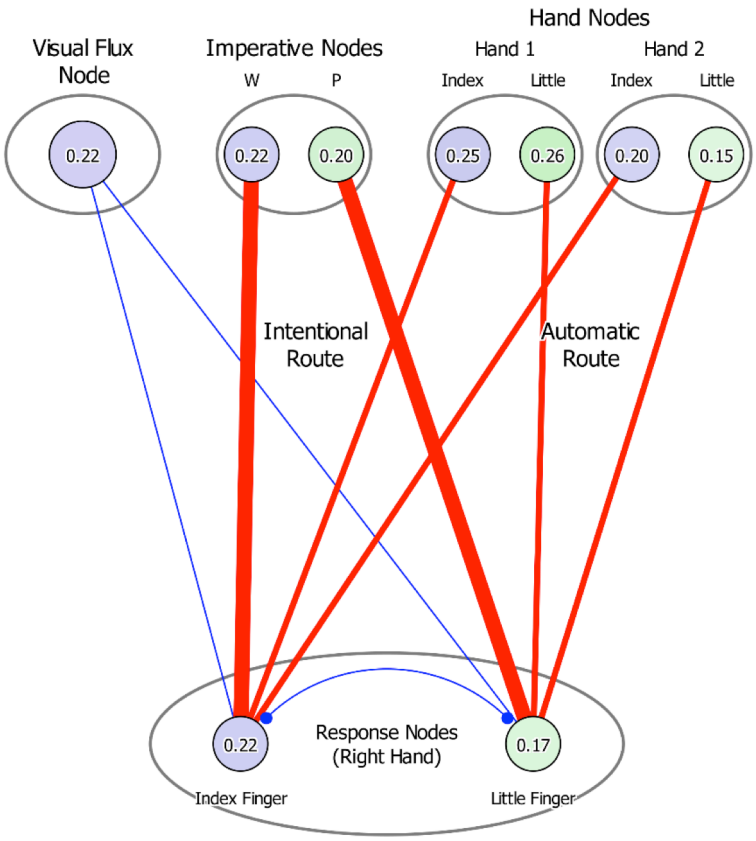


Figure 1. Model A architecture. The visual flux node encodes the amount of visual input. The imperative nodes encode the imperative cue. The hand nodes encode the actions performed by the two stimulus hands. To reflect the hierarchical nature of visual processing, imperative node input starts 20 cycles after visual flux node input and 80 cycles before hand node input. The hand and imperative nodes are connected to the response nodes via excitatory connections (red). The visual flux node is connected to the response nodes via inhibitory connections (blue). The two response nodes inhibit each other through lateral inhibition. The line thickness indicates the strength of the connection.

settings (Cooper, Catmur, et al., 2013; Cooper, Cook, et al., 2013), and test in five simulation studies whether this model can account for the effects previously reported in the literature. Simulations 1-2 focus on the experiments with two stimulus hands. That is, Simulation 1 investigates whether the model can reproduce the stronger automatic imitation for two identical observed actions (Cracco & Brass, 2018a; Cracco et al., 2015) and Simulation 2 whether it can reproduce the absence of automatic imitation for two different observed actions (Cracco et al., 2015). Simulations 3-5 then focus on the experiments with four stimulus hands. In Simulations 3-4, we first investigate how automatic imitation in the model develops as the number of identical observed actions increases. In particular, Simulation 3 explores whether dynamic response inhibition can explain the presence of an asymptote on congruent trials (Cracco & Brass, 2018c) and Simulation 4 evaluates whether such a mechanism could be driven by response conflict (Botvinick, Braver, Barch, Carter, & Cohen, 2001). Finally, in Simulation 5, we investigate whether the model can reproduce the finding that automatic imitation is reduced when three hands perform one action while a fourth hand performs a different action compared with when three or four hands all perform the same action (Cracco & Brass, 2018b).

2. Computational Model

2.1. Model Architecture

As shown in Figure 1 (Model A), the model is an extension of the model by Cooper et al. (2013). Like the original, it is an interactive activation model with a layer of stimulus nodes feeding into a layer of response nodes (McClelland, 1993). In this model, whenever a stimulus is presented, the relevant stimulus node receives input, which is then transformed by an activation function into

activation values that vary over time. Because the stimulus and response nodes are connected, activation of the stimulus nodes leads to excitation or inhibition of the corresponding response nodes, depending on the connection weight. In line with Cooper et al. (2013), the stimulus nodes consist of hand nodes, coding the stimulus movements, and imperative nodes, coding the imperative cue. As in the simulated experiments (Cracco & Brass, 2018c, 2018b; Cracco et al., 2015), the model has to abduct the index finger when the letter W is presented and the little finger when the letter P is presented, while a variable number of stimulus hands abduct either the same or the other finger (Figure 2). Importantly, there is one pair of hand nodes per stimulus hand, meaning that there are two pair of hand nodes in the experiments with two stimulus hands (Cracco et al., 2015) and four pair in the experiments with four stimulus hands (Cracco & Brass, 2018c, 2018b). This formalizes the assumption that different sets of neurons code the different hands depending on their position. Given that the stimulus hands spanned a large portion of the visual display (Figure 2), this is consistent with evidence that even high-level visual neurons are unlikely to code the entire visual field (DiCarlo & Maunsell, 2003; Sayres & Grill-Spector, 2008; Yoshor, Bosking, Ghose, & Maunsell, 2007)[†].

In addition to including more hand nodes, we also added two key features to the Cooper et al. (2013) model. First, we added lateral inhibition between the response nodes to obtain not only facilitation on congruent trials but also interference on incongruent trials. Indeed, it is well-established in the literature that automatic imitation comprises both a facilitation and an interference effect (Brass et al., 2000; Cracco et al., 2015; Genschow et al., 2017). In contrast, Cooper et al. (2013) only considered the facilitation effect. By adding lateral inhibition to the model, activation in one response node is inhibited by activation in the other response node. This then slows down response selection on incongruent trials, resulting in an interference effect (Cooper, Catmur, et al., 2013). Importantly, no lateral inhibition was included at the stimulus level because we assumed conflict to be restricted to the response level, where eventually a single response has to be produced.

Second, we added a visual flux node that codes the amount of visual change and then uses this information to inhibit the response nodes. Because more moving hands equals more visual change, this mechanism regulates response inhibition based on the number of moving hands and as such implements the dynamic inhibition process put forward by Cracco and Brass (2018c). More specifically, imitative tendencies have to be inhibited to prevent overt imitation (Brass, Derrfuss, Cramon, Matthes-von Cramon, & von Cramon, 2003; De Renzi, Cavalleri, & Facchini, 1996; Lhermitte, Pillon, & Serdaru, 1986). However, if motor activation increases with the number of observed actions, motor inhibition has to increase as well. This is achieved by the visual flux node, which, by inhibiting both response nodes, effectively adjusts the response threshold on each trial based on the amount of external input, making the model more cautious when there is high risk of external cues driving behavior (Houghton, Tipper, Weaver, & Shore, 1996; Schuch, Bayliss, Klein, & Tipper, 2010).

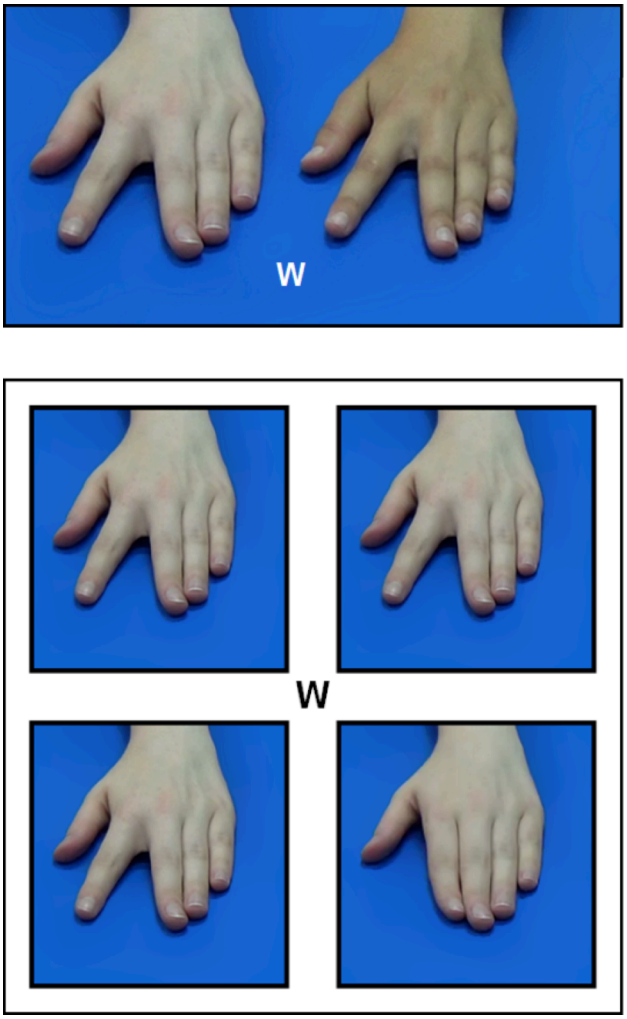
2.2. Mathematical Implementation

The model, data, and processing scripts are available on the Open Science Framework via the following link: https://osf.io/kezyt/. The mathematical implementation of the model is based on Cooper et al. (2013). That is, each node has an activation value between 0 and 1 that changes on each cycle according to the following activation function:

$$a_i(t+1) = \rho\, a_i(t) + (1-\rho)\, \sigma\big(I_i(t)\big) \quad (1)$$

where t is time, ρ is a persistence parameter that determines the persistence of activation from one cycle to the next over time, $I_i(t)$ is the net input to node i at time t, and σ is a sigmoid function that transforms the input to a value between 0 and 1.

[†] One might wonder how this would scale up to more hands. Putting more hands on the screen inevitably puts them closer together and multiple hands may then be coded by the same set of neurons. In this case, the hand nodes may no longer represent single hands but rather groups of hands, consistent with evidence that people rely on ensemble processes to represent larger stimulus quantities (Alvarez, 2011).



Model A:

Simulation 1 (Identical Actions)

Simulation 2 (Different Actions)

Model B:

Simulation 3 (Identical Actions)

Simulation 5 (Different Actions)

Model C:

Simulation 4 (Identical Actions + Response Conflict)

Figure 2. Sample stimuli for the two tasks on which the models are based. In both tasks, participants have to abduct their index finger in response to the letter W (corresponding to "wijsvinger", the Dutch for index finger) and their little finger in response to the letter P (corresponding to "pink", the Dutch for little finger) while the hands perform either a congruent or an incongruent action. The upper panel shows an example of a two-hand stimulus (Simulations 1 and 2). In the example, both hands perform an action that is congruent with the correct response. The lower panel shows an example of a four-hand stimulus (Simulations 3, 4 and 5). In the example, three of the four hands perform an action that is congruent with the correct response. In Simulations 1, 3, and 4, the hands all perform the same (congruent or incongruent) action (Cracco & Brass, 2018c; Cracco et al., 2015). In Simulations 2 and 5, they perform different actions so that one subset of the hands performs a congruent action while another subset performs an incongruent action (Cracco & Brass, 2018b; Cracco et al., 2015). The Model A task includes two hands. The Model B and C tasks include four hands.

The net input $I_i(t)$ to the stimulus nodes (i.e., the imperative, hand, and visual flux nodes) is defined as:

$$I_i(t) = \beta_i + E_i + N(0, \eta^2) \quad (2)$$

where β_i is a bias parameter determining the node's resting activation, E_i is the external input, and η is the standard deviation of the noise added on each cycle. Importantly, while external input to the imperative and hand nodes is a fixed value (Table 1), external input to the visual flux node varies based on the number of stimulus hands performing an action. That is, visual flux node external input is calculated as:

$$E_i = \alpha \frac{N_{mov}}{N_{tot}} \quad (3)$$

where α is a scaling parameter, N_{mov} is the number of moving stimulus hands, and N_{tot} is the total number of stimulus hands. Crucially, this means that the incremental input per additional moving hand decreases as the total number of hands increases. This, in turn, reflects the assumption that each individual stimulus hand provides a more salient visual trigger when there are fewer hands in total, consistent with the fact that the congruency effect produced by a single hand was stronger in the experiments with two hands (Cracco et al., 2015) than in the experiments with four hands (Cracco & Brass, 2018c).

The net input of each response node is determined by the sum of the product of the activation in the stimulus nodes and the strength of the weight connecting those nodes to the response nodes:

$$I_i(t) = \beta_i + \sum_j \left(w_{ji}\, a_j(t-1) \right) + N(0, \eta^2) \quad (4)$$

where w_{ji} is the strength of the weight that connects node j to node i. As shown in Table 1 and Figure 1, the hand and imperative nodes are connected to the response nodes via excitatory connections, whereas the visual flux node is connected to the response nodes via inhibitory connections. Finally, lateral inhibition is implemented by including inhibitory connections between the two response nodes.

2.3. Trial Procedure and Parameter Values

The activation level of all nodes is updated every cycle according to Equations 1-4. At the start of each trial, the model first goes through 500 cycles without input to allow activation in all nodes to settle on their resting activation, determined by β_i. Next, the stimuli are presented, resulting in excitation of the stimulus nodes according to the delay parameters δ_h and δ_f. These delay parameters represent the number of cycles between the presentation of input to the imperative nodes and the presentation of input to the hand (δ_h) and visual flux (δ_f) nodes. In line with Cooper et al. (2013), the δ_h parameter was set to +80 cycles to implement the assumption that observed actions are processed more slowly than letters (Catmur & Heyes, 2011). Similarly, the δ_f parameter was set to -20 cycles to reflect the hierarchical nature of visual processing, with basic features like motion being processed before complex features like the identity of objects or actions (Van Essen & Maunsell, 1983).

Following the stimulus presentation, activation in all nodes is updated each cycle until activation in one of the two response nodes reaches the response threshold τ_r of 0.80. Similar to Cooper et al. (2013), activation in the imperative and visual flux nodes is maintained up to the response, while activation in the hand nodes is transitory. This is implemented by terminating excitation of the hand nodes when activation in these nodes reaches the habituation threshold τ_h of 0.80, leading to a gradual decay of activation.

All parameter values are described in Table 1. As can be seen, most values were taken from Cooper et al. (2013). The only exceptions are the persistence parameter of the hand nodes ρ_h, the lateral inhibition weight w_{rr}, and five new parameters related to the visual flux node, namely the scaling parameter α, the bias parameter β_f, the persistence parameter ρ_f, the flux-to-motor weight w_{rf}, and the delay parameter δ_f. Following Cooper et al. (2013), we chose ρ_h so that the modeled congruency effect was similar in size to the empirical congruency effect. That is, we tuned the model's sensitivity to the hands by making hand node activity more or less transitory. As explained above, δ_f was chosen to reflect the assumption that motion is processed before letters or actions. However, varying this parameter had almost no effect on the reported simulations (Figure S4-S5). The visual flux node parameters β_f and ρ_f were set in accordance with the imperative node parameters β_i and ρ_i. Finally, the last three parameters, w_{rr}, α, and w_{rf}, were set to obtain facilitation effects (i.e., neutral minus congruent) that were approximately the same size as the interference effects (i.e., incongruent minus neutral), based on evidence that these effects are roughly balanced in automatic imitation tasks (Brass et al., 2000; Cracco et al., 2015; Genschow et al., 2017). More generally, this reflects the assumption that performance in S-R compatibility tasks is optimized by balancing the amount of exerted control so that interference on incongruent trials is minimized without eliminating facilitation on congruent trials (see Botvinick et al., 2001 for a similar logic and simulations). The influence of varying these last three parameters is assessed in parameter variation studies.

Table 1.
Description of the Model Parameters

Symbol	Description	Simulations 1-2	Simulations 3-5	Cooper et al. (2013)
E	External input to the stimulus nodes (see Eq. 2)	$E_h = 5.00$	$E_h = 5.00$	$E_h = 5.00$
		$E_i = 5.00$	$E_i = 5.00$	$E_i = 5.00$
α	Scaling factor for visual flux input E_f (see Eq. 3)	$\alpha = 2.00$	$\alpha = 2.00$	N/A
β	Bias added to the input of each node (see Eqs. 2 and 4)	$\beta_h = -2.00$	$\beta_h = -2.00$	$\beta_h = -2.00$
		$\beta_i = -2.00$	$\beta_i = -2.00$	$\beta_i = -2.00$
		$\beta_f = -2.00$	$\beta_f = -2.00$	N/A
		$\beta_r = -6.00$	$\beta_r = -7.50$	$\beta_r = -6.00$
ρ	Persistence of activation between cycles (see Eq. 1)	$\rho_h = 0.945$	$\rho_h = 0.925$	$\rho_h = 0.960$
		$\rho_i = 0.990$	$\rho_i = 0.990$	$\rho_i = 0.990$
		$\rho_f = 0.990$	$\rho_f = 0.990$	N/A
		$\rho_r = 0.990$	$\rho_r = 0.990$	$\rho_r = 0.990$
w	Weight of the connections between the nodes (see Eq. 4)	$w_{rh} = 4.00$	$w_{rh} = 4.00$	$w_{rh} = 4.00$
		$w_{ri} = 8.00$	$w_{ri} = 8.00$	$w_{ri} = 8.00$
		$w_{rf} = -1.00$	$w_{rf} = -1.00$	N/A
		$w_{rr} = -1.00$	$w_{rr} = -1.00$	N/A
η	Standard deviation of the noise added to each node (see Eqs. 2 and 4)	$\eta = 2.00$	$\eta = 2.00$	$\eta = 2.00$
δ	Delay between excitation of the imperative node and the other stimulus nodes	$\delta_h = +80$	$\delta_h = +80$	$\delta_h = +80$
		$\delta_f = -20$	$\delta_f = -20$	N/A
τ	Habituation (hand nodes) and response (response nodes) threshold	$\tau_h = 0.80$	$\tau_h = 0.80$	$\tau_h = 0.80$
		$\tau_r = 0.80$	$\tau_r = 0.80$	$\tau_r = 0.80$

3. Simulation Results

3.1. Simulation 1

Simulation 1 aims to reproduce the first finding of Cracco et al. (2015), namely that two hands performing the same action (2H ID) produce a stronger congruency effect than one hand performing a single action (1H). More specifically, it does so by implementing the hypothesis that identical actions activate the same motor representation (Cracco & Brass, 2018c, 2018a; Cracco et al., 2015, 2016). As shown in Figure 3, the model captures the empirical results, with responses on congruent trials being faster and responses on incongruent trials being slower in the 2H ID condition than in the 1H condition.

To assess the robustness of the model, we then performed a parameter variation study in which the model's performance was evaluated while varying the w_{rr}, α, and w_{rf} parameters from zero to three times their default value. More precisely, the model was run 50 times for each point in parameter space, and on each simulation we calculated the difference between the congruency effect

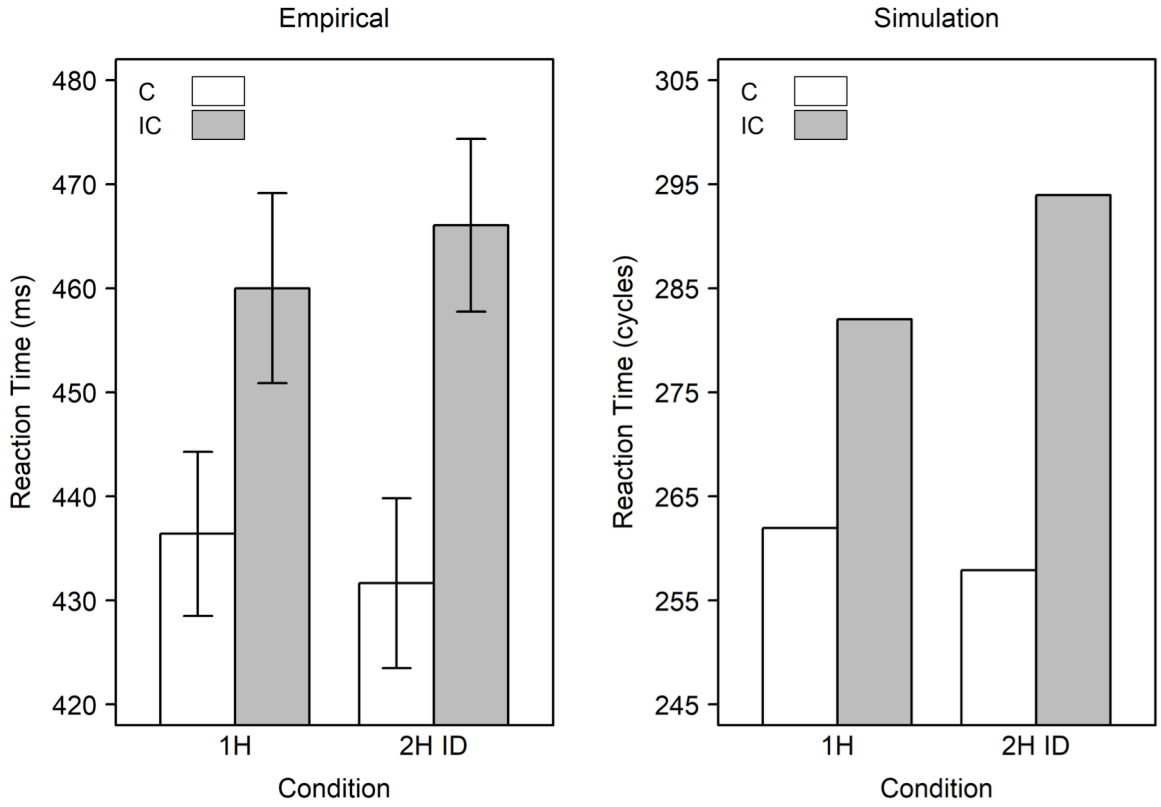


Figure 3. Results of Simulation 1. The congruency effect when one hand performs an action (1H) is compared with the congruency effect when two hands perform the same action (2H ID). The empirical data is based on Experiment 1 of Cracco et al. (2015). The simulation was run using the same number of trials and participants as in the original experiment. Error bars are standard errors of the mean (SEMs).

in the 2H ID condition and the 1H condition. As can be seen in Figure 4, this revealed that the empirical effect was reproduced across almost the entire parameter space, demonstrating robustness.

3.2. Simulation 2

In Simulation 2, the second finding of Cracco et al. (2015) is simulated, namely that RTs in the condition where two hands perform two different actions (2H DIFF) are equal to RTs in the neutral condition where neither hand performs an action (N). Specifically, the model implements the hypothesis that simultaneously seeing one congruent and one incongruent action produces concurrent facilitation and interference effects that then cancel out each other. As demonstrated in Figure 5, the empirical effect was reproduced by the model, with similar RTs in the 2H DIFF and N conditions.

Similar to Simulation 1, the robustness of the model's behavior was evaluated in a parameter variation study. This revealed that the model reproduced the key effect as long as α was proportional to w_{rf}. In other words, stronger input required weaker weights and weaker input required stronger weights (Figure 6). Interestingly, Figure 7 indicates that this pattern could be traced back to the assumed symmetry between facilitation and interference. That is, a comparison with Figure 6 shows that 2H DIFF and N were equal only when facilitation and interference were balanced. This is because the former requires facilitation and interference to cancel out in the 2H DIFF condition, which can only happen if they are similar in size. To sum up, varying parameters in Simulation 2 revealed that the model was able to capture the empirical effect as long as facilitation and interference were balanced. Given that such a balance is well-grounded in the literature (Brass et al., 2000; Cracco et al., 2015; Genschow et al., 2017), this indicates that the model is robust under realistic circumstances.

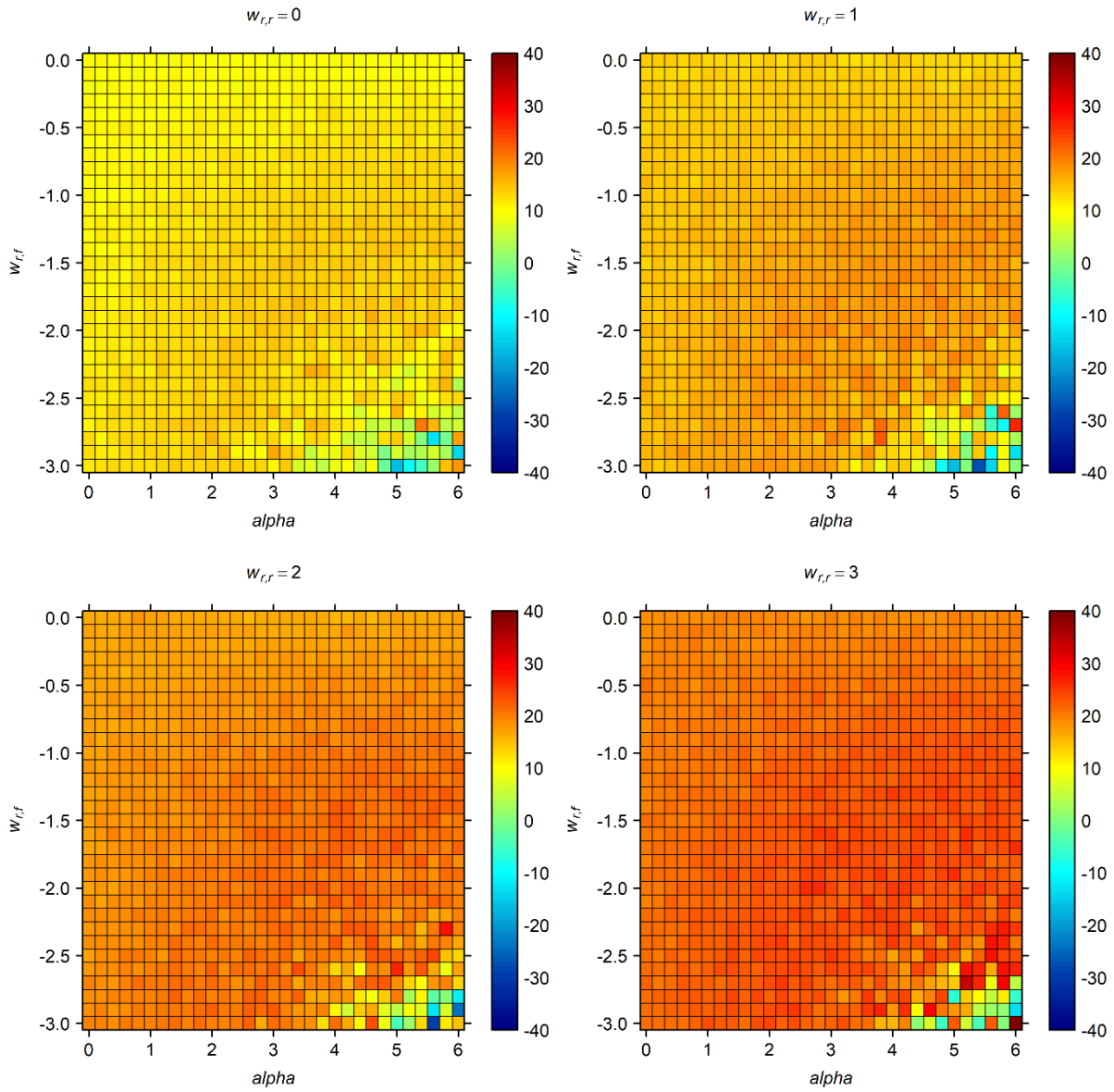


Figure 4. Parameter variation results for Simulation 1. Each cell shows the difference between the congruency effect in the 2H ID condition and the 1H condition for different combinations of w_{rr}, α, and w_{rf}.

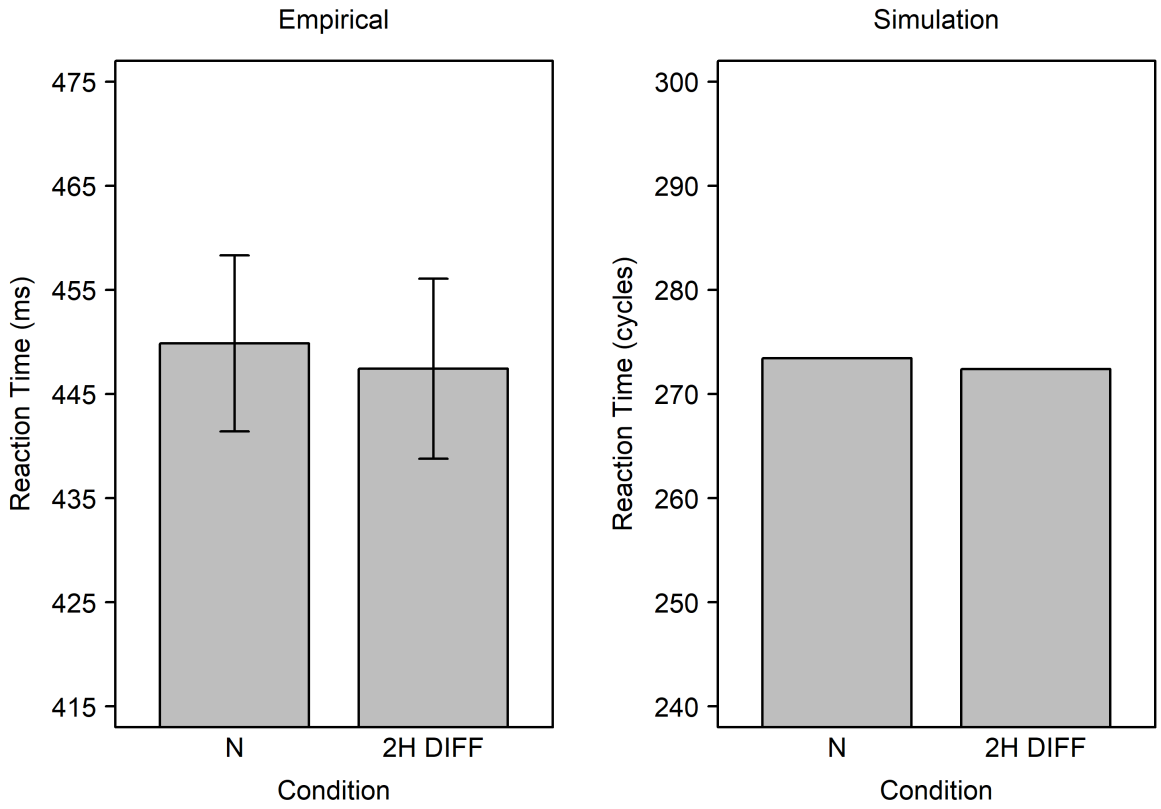


Figure 5. Results of Simulation 2. RTs in the condition where neither hand performs an action (N) are compared with RTs in the condition where the two hands perform different actions (2H DIFF). The empirical data is based on Experiment 1 of Cracco et al. (2015). The simulation was run using the same number of trials and participants as in the original experiment. Error bars are SEMs.

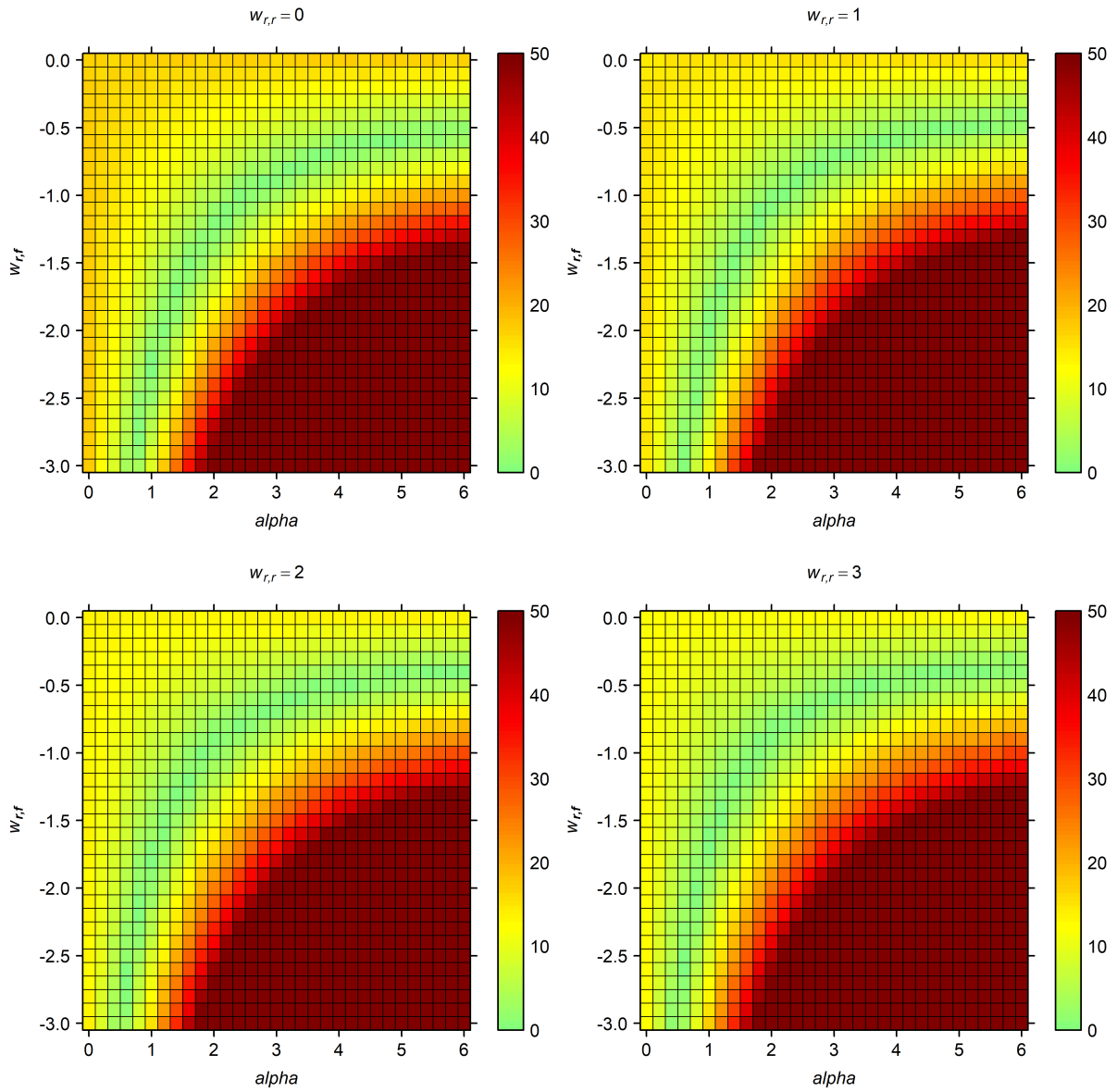


Figure 6. Parameter variation results for Simulation 2. Each cell shows the absolute difference in RTs between the N condition and the 2H DIFF condition for different combinations of w_{rr}, α, and w_{rf}. The color bar was capped at 50.

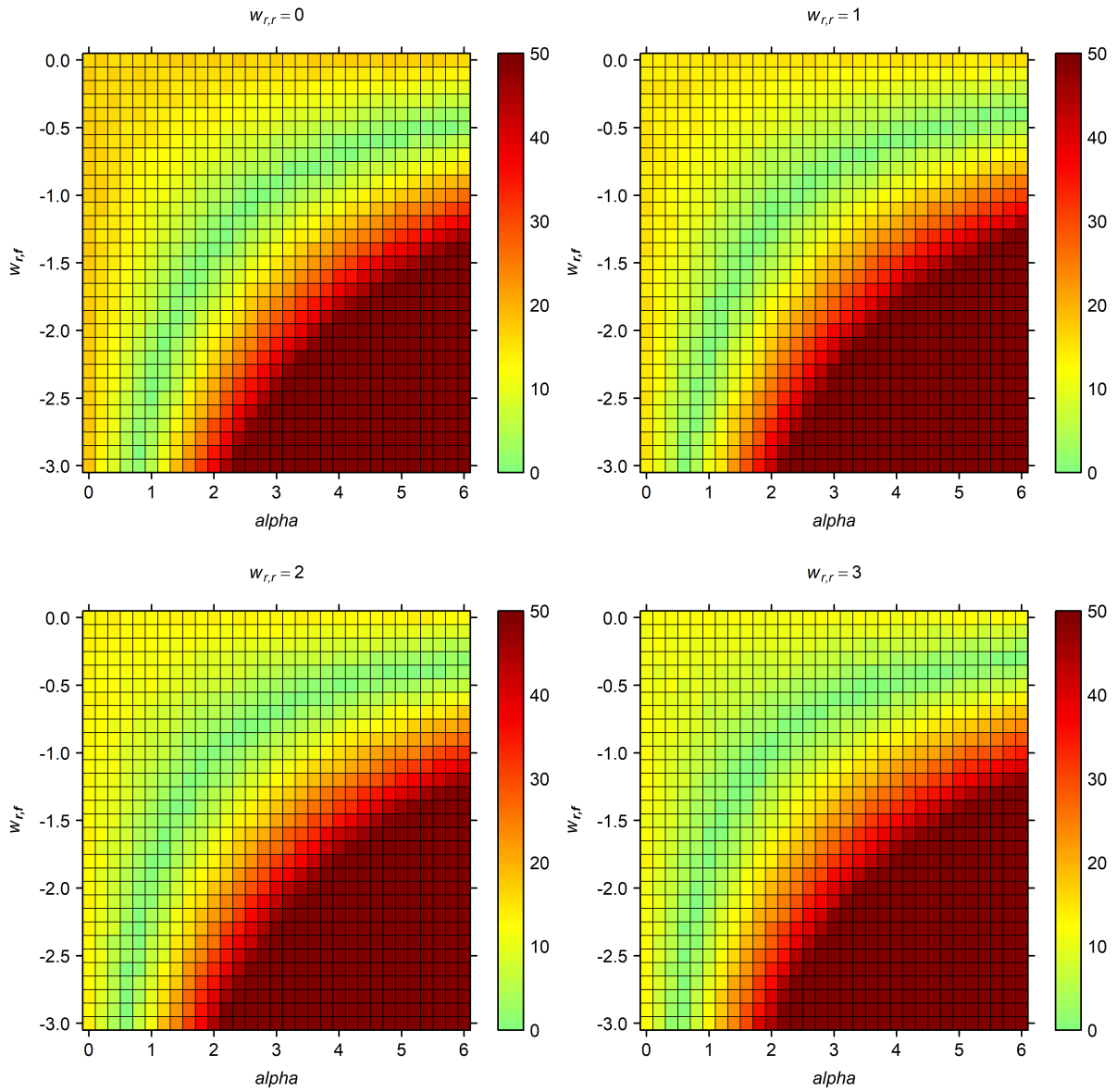


Figure 7. Parameter variation results for Simulation 2. Each cell shows the absolute difference in RTs between the facilitation (neutral minus congruent) and interference (incongruent minus neutral) effects obtained in the 1H condition for different combinations of w_{rr}, α, and w_{rf}. The color bar was capped at 50.

3.3. Simulation 3

Simulation 3 extends Simulations 1-2 by extending the number of stimulus hands from two to four (Model B; Figure 8). Importantly, this required two adjustments. First, β_r was changed from -6.00 to -7.50 to ensure that the resting activation of the Model A and Model B response nodes was comparable[‡]. Second, ρ_h was adjusted from 0.945 to 0.925 to obtain, in line with Model A, congruency effects that were similar in size to the empirical congruency effects (Cooper, Catmur, et al., 2013).

[‡] Changing β_r was necessary because the Model B response nodes receive input from twice the number of stimulus hands as the Model A response nodes. Therefore, keeping β_r constant increased the response node resting activation by a factor of two in Model B, resulting in unreasonably fast RTs. In more general terms, the adjustment of β_r reflects the assumption that different experiments require a different response threshold or, equivalently, different levels of tonic inhibition.

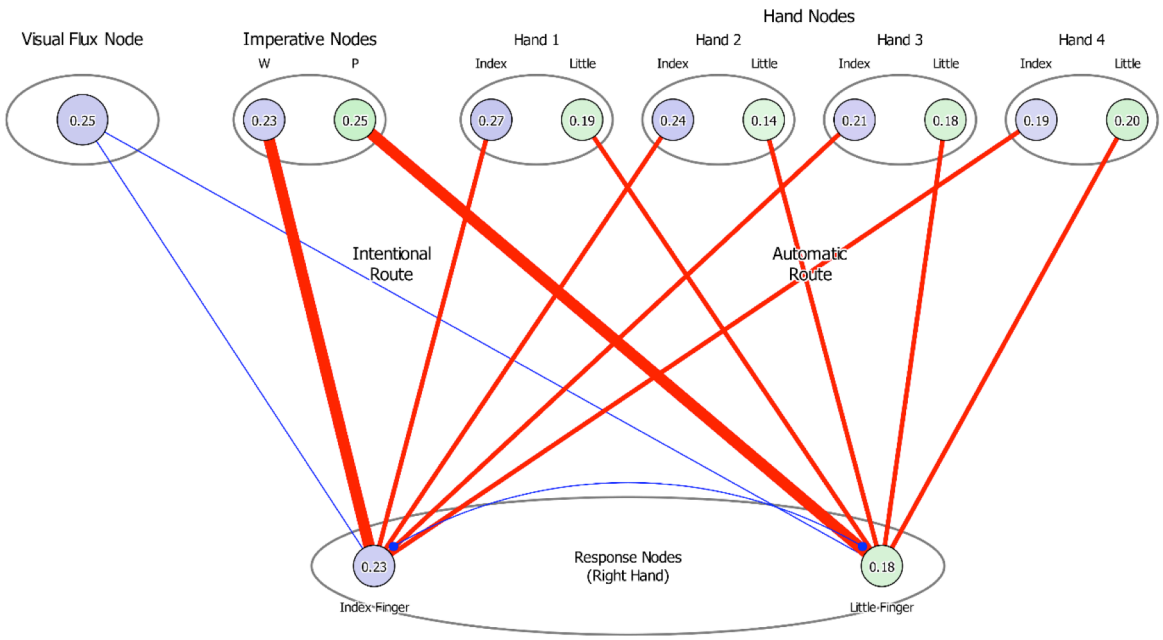


Figure 8. Model B architecture. Model B extends Model A by adding two more sets of hand nodes. Input to the imperative nodes starts 20 cycles after input to the visual flux node and 80 cycles before input to the hand nodes.

Model B was used to investigate the hypothesis that trial-by-trial adjustments in the amount of response inhibition could explain why RTs in Cracco & Brass (2018c) followed an asymptotic curve on congruent trials but a linear curve on incongruent trials. As shown in Figure 9, the model was able to reproduce the empirical pattern. This is because motor inhibition from the visual flux node increases with the number of moving hands, leading to a corresponding RT increase that eliminates the expected RT decrease on congruent trials but strengthens the analogous RT increase on incongruent trials.

However, Cracco & Brass (2018c) also found that the congruent asymptote disappeared when all trials were congruent ("non-mixed procedure") rather than 50% congruent and 50% incongruent ("mixed procedure"). It was argued that response inhibition is sensitive to variations in task context (Bugg & Crump, 2012) and that imitative responses were therefore no longer inhibited when they facilitated task performance (Cracco & Brass, 2018c). To test this hypothesis, we repeated Simulation 3 with the flux-to-response weights set to zero. Mirroring the empirical results, this caused RTs to no longer reach an asymptote but to instead decrease monotonically with the number of moving hands (Figure 10). Importantly, in line with the empirical data, RTs also became faster when the flux-to-response weights were set to zero, indicating that the asymptote could not have been caused by physical restraints preventing participants from speeding up (Cracco & Brass, 2018c).

Finally, we evaluated the model's performance for different combinations of w_{rr}, α, and w_{rf} by delineating (a) the parameter space in which congruent RTs decreased asymptotically using the mixed procedure but linearly using the non-mixed procedure, and (b) the parameter space in which incongruent RTs increased linearly using the mixed procedure. For the first analysis, we first fitted an asymptotic function defined by a vertical intercept parameter, a horizontal asymptote parameter, and a shape parameter to the congruent empirical RTs in the mixed scenario (Figure S1). Next, we extracted the shape parameter and fixed it while fitting an asymptotic function (now with just 2 parameters) to the simulated RTs at varying parameter values. Finally, we fitted a linear function (also with 2 parameters) to the same simulated RTs and compared the Bayesian Information Criterion (BIC) of both functions. For the second analysis, we fitted a linear function to the incongruent simulated RTs and compared the BIC of this function with the BIC of a function including only an intercept.

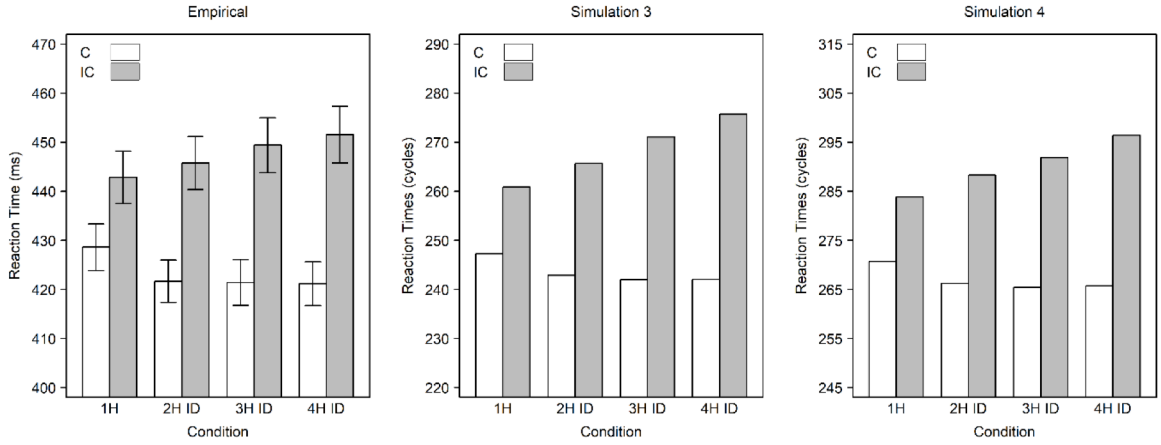


Figure 9. Results of Simulations 3 and 4 using the mixed procedure (i.e., 50% congruent and 50% incongruent trials). The congruency effect when one hand performs an action (1H) is compared with the congruency effect when two hands (2H ID), three hands (3H ID), or four hands (4H ID) perform the same action. The empirical data is based on Experiments 1-2 of Cracco & Brass (2018c). The simulations were run using the same number of trials and participants as in the original experiment. Error bars are SEMs.

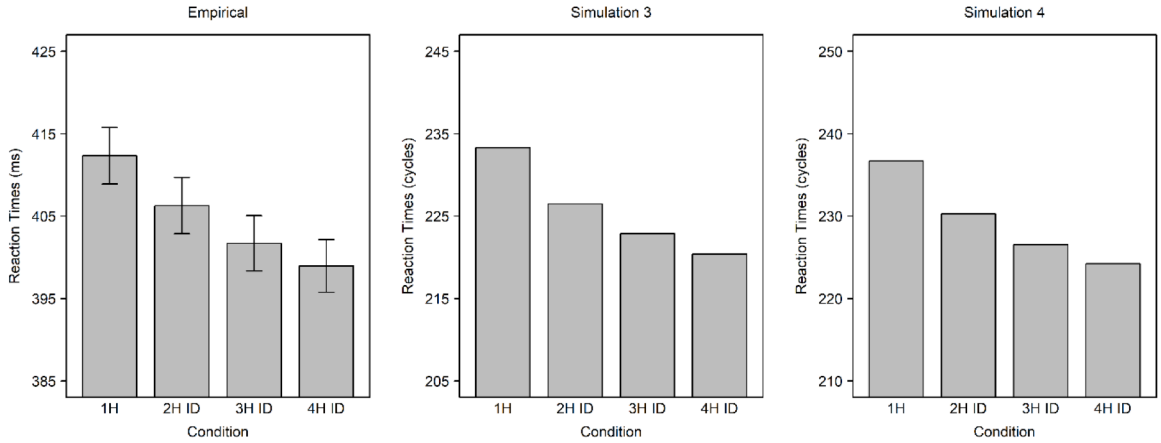


Figure 10. Results of Simulations 3 and 4 using the non-mixed procedure (i.e., 100% congruent trials). RTs in the condition where one hand performs an action (1H) are compared with RTs in the condition where two hands (2H ID), three hands (3H ID), or four hands (4H ID) perform the same action. The empirical data is based on Experiments 3-6 of Cracco & Brass (2018c). The simulations were run using the same number of trials and participants as in the actual experiments. Error bars are SEMs.

The results of the first analysis showed that the model reproduced the empirical asymptote as long as α and w_{rf} were balanced (Figure 11). When either parameter was too weak, RTs were instead better described by a linearly decreasing curve. Consistent with the finding that RTs on congruent trials decreased monotonically when the flux-to-response weights were put to zero, this indicates that inhibition from the visual flux node is needed to produce an asymptote. Conversely, when either α or w_{rf} were too high, RTs were characterized by a linearly increasing curve. That is, when there is too much inhibition, RTs on congruent trials increase rather than decrease with the number of moving hands. Together, these findings highlight the need to balance facilitation and interference (Botvinick et al., 2001; Bugg & Crump, 2012). In support, the region in which the model produced an asymptote (Figure 11) overlapped closely with the region in which the facilitation and interference effects were balanced (Figure 7). Thus, in line with Simulation 2, these findings suggests that the model is robust under realistic circumstances (Brass et al., 2000; Cracco et al., 2015; Genschow et al., 2017).

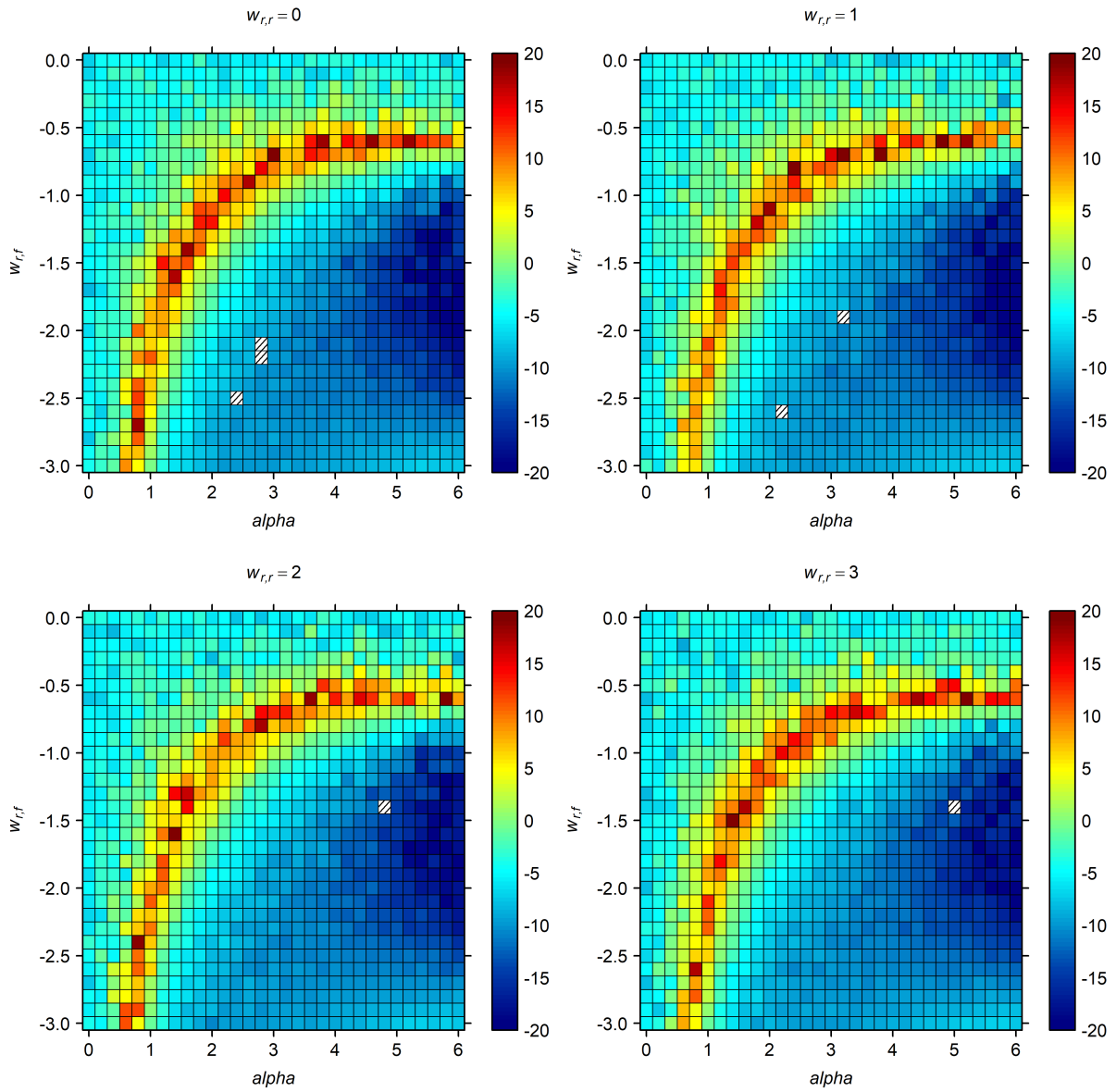


Figure 11. Parameter variation results for Simulation 3, showing model performance on congruent trials. Each cell shows the difference between the BIC of the linear function and the BIC of the asymptotic function for different combinations of w_{rr}, α, and w_{rf}. Positive values indicate better fit for the asymptotic function, meaning that congruent RTs decreased asymptotically as the number of observed movements increased. Shaded cells are cells where the asymptotic function did not converge.

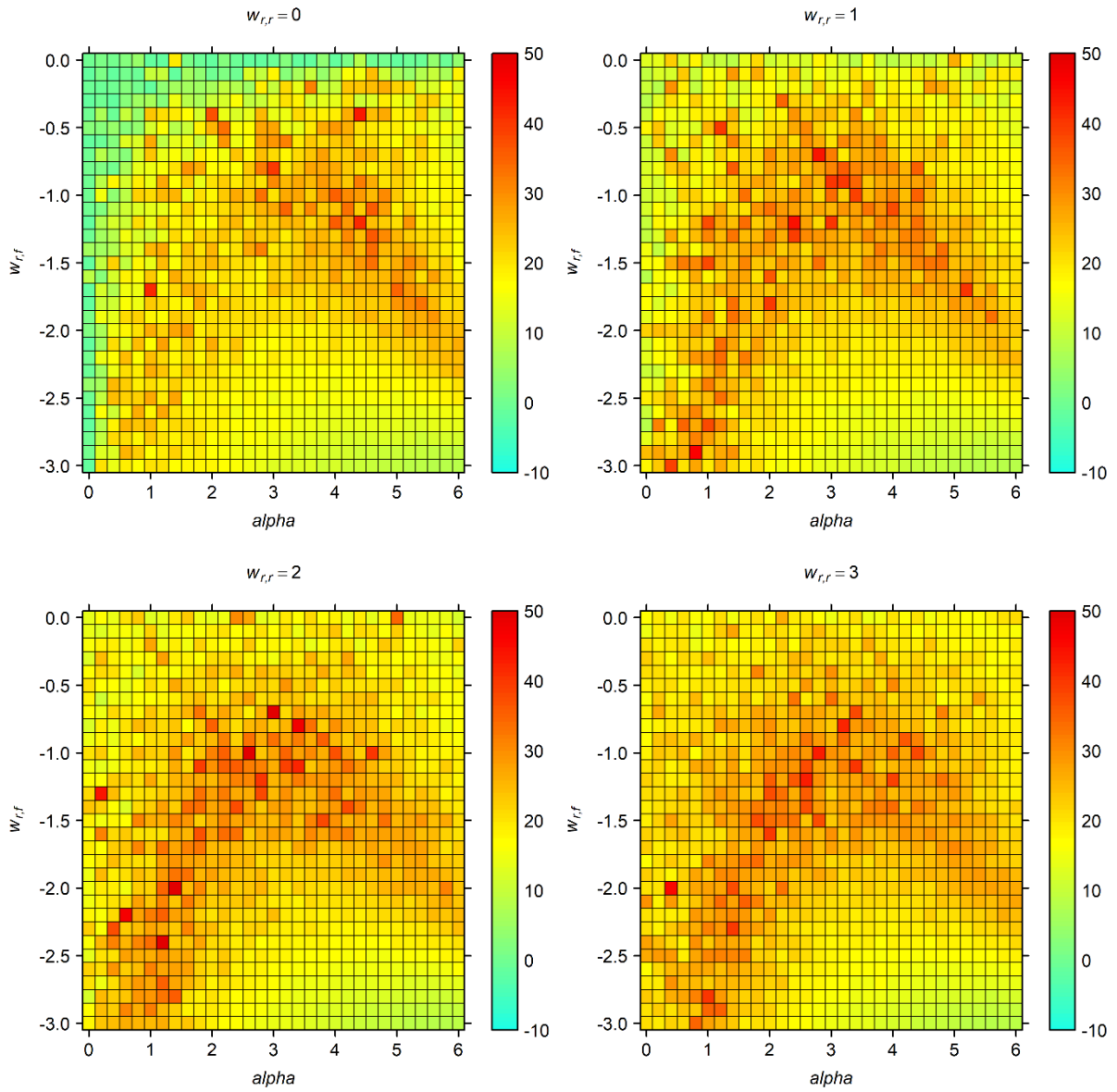


Figure 12. Parameter variation results for Simulation 3, showing model performance on incongruent trials. Each cell shows the difference between the BIC of the intercept function and the BIC of the linear function for different combinations of w_{rr}, α, and w_{rf}. Positive values indicate a better fit for the linear function, meaning that incongruent RTs increased linearly with the number of observed movements.

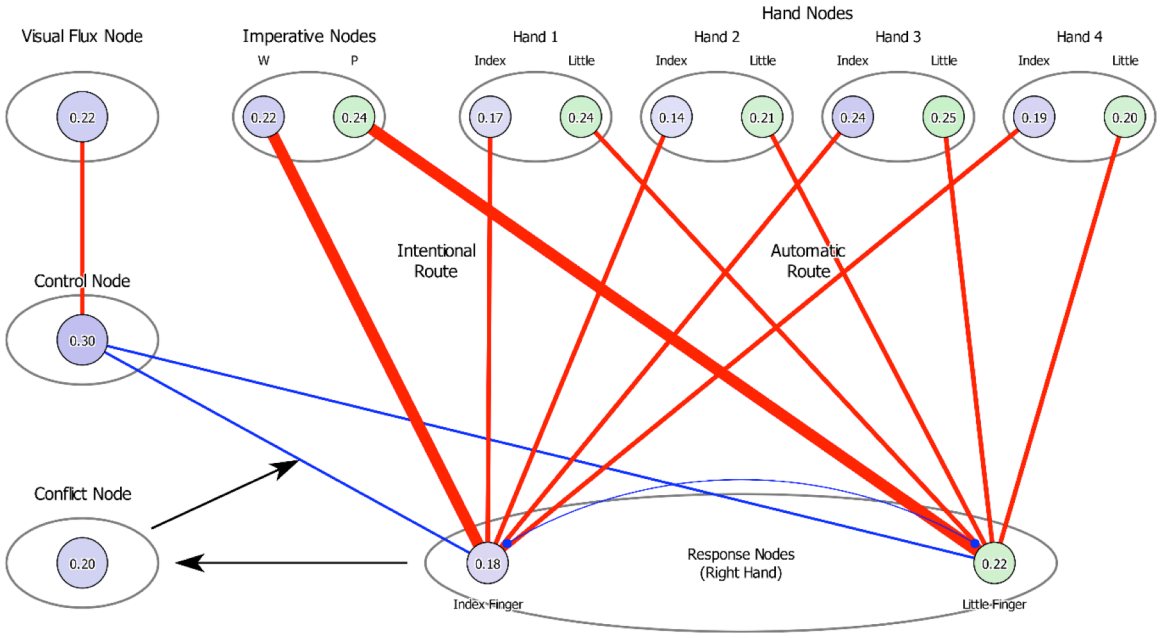


Figure 13. Model C architecture. Model C extends Model B by adding a control and a conflict node. The conflict node encodes the response conflict on each trial. The control-to-response weights are updated on each trial based on the amount of response conflict in the previous trial. Input to the imperative nodes starts 20 cycles after input to the visual flux node and 80 cycles before input to the hand nodes.

The results of the second analysis indicated that incongruent RTs increased linearly across the entire parameter space, except for a small region where lateral and visual flux inhibition were both low, causing the interference effect to be absent (Figure 12). In sum, Simulation 3 shows that flexible response inhibition can explain why RTs reached an asymptote on congruent trials in the mixed scenario, with 50% congruent and 50% incongruent trials, but not in the non-mixed scenario, with 100% congruent trials.

3.4. Simulation 4

Importantly, setting the flux-to-response weights by hand assumes that inhibitory control is fine-tuned based on task instructions or practice trials. However, an alternative hypothesis is that the model gradually learns to optimize response inhibition as the experiment progresses. To test this hypothesis, Simulation 4 regulates cognitive control using response conflict (see also Botvinick et al., 2001; Botvinick, Cohen, & Carter, 2004). To this end, we made two adjustments to Model B (Model C; Figure 13). First, we added an intermediate control node, formalizing the assumption that motor inhibition is implemented not by visual but by prefrontal brain areas (Munakata et al., 2011). Second, we now continuously updated the control-to-response connections by initializing the weights to zero and then updating them based on the response conflict in the previous trial. In line with Botvinick et al. (2001), response conflict was defined as the product of activation in the two response nodes:

$$X(t) = -\,a_i(t)\,a_j(t)\,w_{ji}(t) \tag{5}$$

To update the control-to-response weights, response conflict on the current trial is compared with the median response conflict over the past *n* trials. If response conflict is lower than "usual", the weight becomes more negative. If it is higher than "usual", the weight becomes more positive:

$$w_{ji}(t+1) = \rho\, w_{ji}(t) + \gamma\,(X(t) - X_0) \tag{6}$$

where $w_{ji}(t)$ and $w_{ji}(t+1)$ are the control-to-response weights on the current and next trial, ρ (= 0.95) is a persistence parameter causing the weights to slowly tend towards zero in the absence of conflict, γ (= 7.50) is a scale parameter determining the sensitivity of the weights to variations in conflict, $X(t)$ is the response conflict experienced on the current trial, and X_0 is the median response

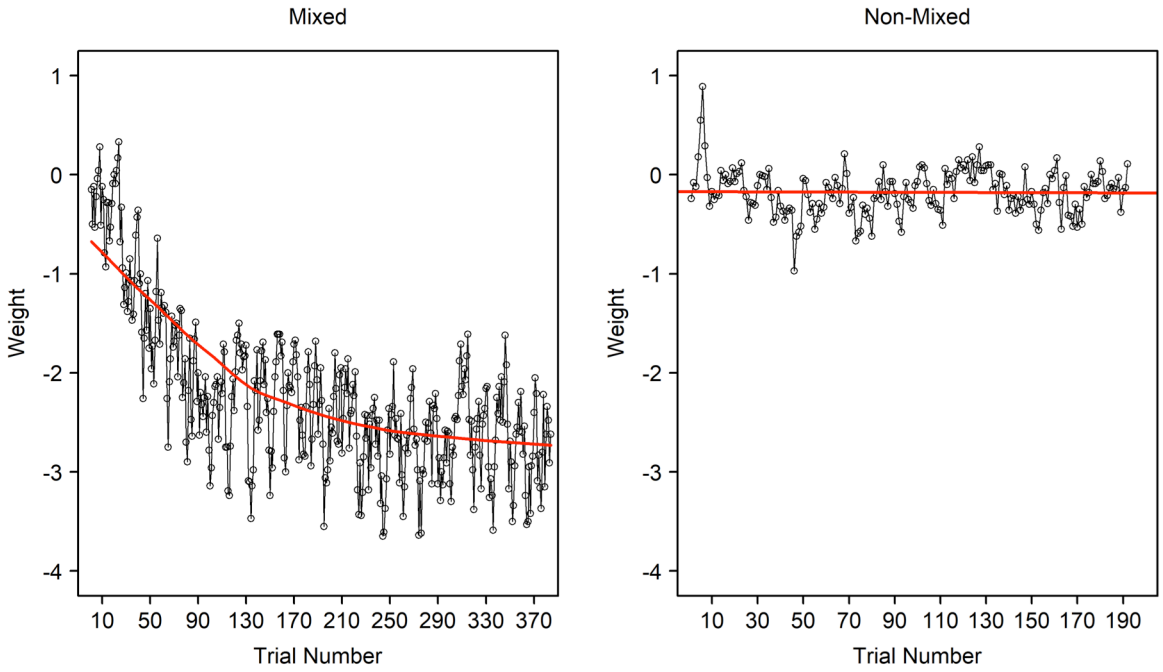


Figure 14. Trial-by-trial variations in the control-to-response weights of Model C using the mixed and non-mixed procedure. The results of a single representative participant are plotted for both scenarios. A loess fit line is shown for the mixed scenario and a regression fit line for the non-mixed scenario.

conflict over the past *n* (= 16) trials[§]. Importantly, because response conflict is the product of activation in the response nodes, it is strong on incompatible trials, where both responses are activated, but weak on compatible trials, where a single response is activated. Therefore, in the mixed scenario, the control-to-response weights become gradually more negative until they reach a state of equilibrium. In contrast, in the non-mixed scenario, the weights fluctuate around zero (Figure 14).

As shown in Figures 9 and 10, Simulation 4 closely reproduces both the empirical results and the results of Simulation 3. This suggests that the dynamic control process proposed in Cracco & Brass (2018c) could be driven by response conflict. In particular, Model C provides a mechanistic account of why congruent responses plateau in the mixed but not in the non-mixed scenario. That is, in the mixed scenario, participants experience response conflict, which causes them to inhibit motor activation based on the number of moving hands. In contrast, in the non-mixed scenarios, participants do not experience response conflict, and no inhibitory control is exerted.

3.5. Simulation 5

Moving away from observing identical actions, Simulation 5 uses Model B to simulate the finding of Cracco & Brass (2018b) that automatic imitation is stronger when three (3H ID) or four (4H ID) hands all perform the same action compared with when three hands perform one action and one hand performs a different action (3H/1H DIFF). Similar to Simulation 2, the model implements the hypothesis that, in the third condition, three hands activate one response, whereas the fourth hand activates a different response. For instance, if three hands perform a congruent action and one hand performs an incongruent action, this should lead to a strong facilitation effect that is partially counteracted by a concurrent but weaker interference effect. Supporting this hypothesis, Figure 15 shows that the empirical effect was closely captured by the model, with smaller congruency effects in the 3H/1H DIFF condition than in the 3H ID and 4H ID conditions. Furthermore, a parameter variation study revealed that this pattern was obtained across almost the entire parameter space,

[§] In Simulation 4, *n* was set to the length of the practice phase (i.e., *n* = 16) for convenience. However, the same results can also be obtained when *n* is set to, for example, 8 or 32.

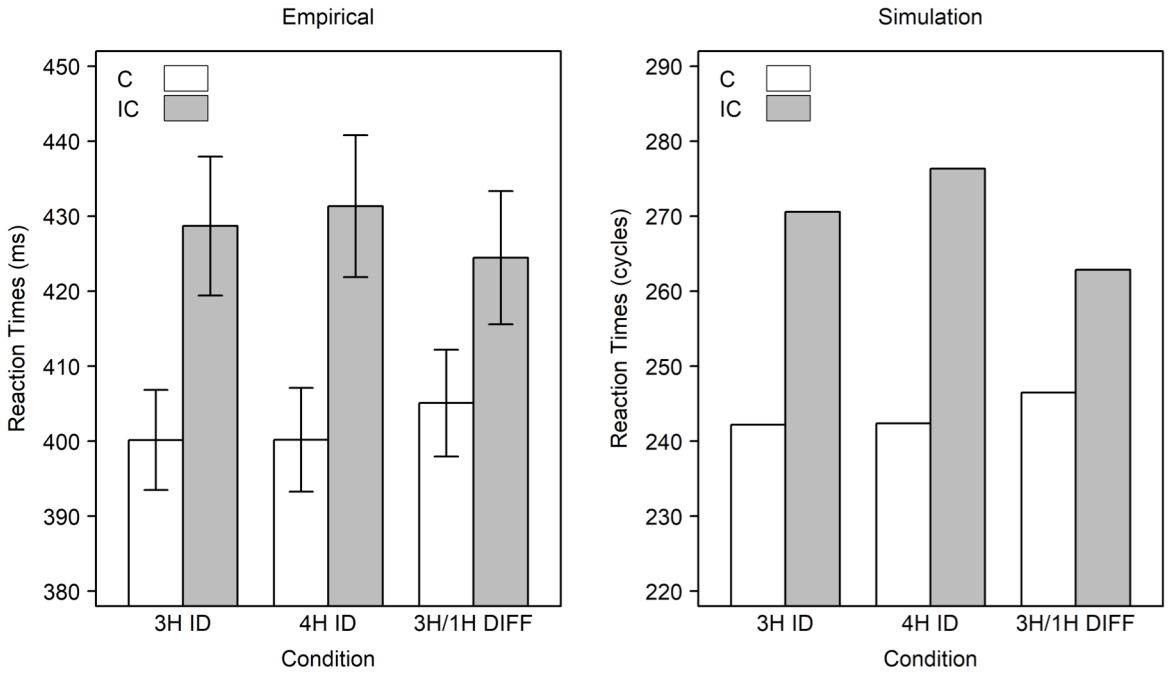


Figure 15. Results of Simulation 5. The congruency effect when three hands perform the same action (3H ID) is compared with the congruency effect when four hands perform the same action (4H ID) and with the congruency effect when three hands perform one action while the fourth hand performs a different action (3H/1H DIFF). The 3H/1H DIFF congruency effect is determined with respect to the majority of the hands. The empirical data is based on Experiment 1 of Cracco et al. (Cracco & Brass, 2018b). The simulation was run using the same number of trials and participants as in the original experiment. Error bars are SEMs.

indicating that the results of Simulation 5 were robust to changes in the w_{rr}, α, and w_{rf} parameters (Figure 16).

4. Discussion

There is now converging evidence that observers represent others' actions in their own motor system (Caspers, Zilles, Laird, & Eickhoff, 2010), and that this leads to automatic imitation (Cracco, Bardi, et al., 2018). Interestingly, while research has traditionally focused on social situations involving a single agent, recent work has started to investigate what happens when there are multiple agents (Cracco & Brass, 2018a, 2018b, 2018c; Cracco et al., 2015, 2016; Cracco, Keysers, Clauwaert, & Brass, 2018; Ramenzoni, Sebanz, & Knoblich, 2014; Tsai, Sebanz, & Knoblich, 2011). Overall, this work shows that the actions of multiple agents can be represented together in the motor system. However, whether these findings can be accounted for within existing theories of action representation remains unclear. Here, we used computational modeling to validate the dual-route model of automatic imitation in multi-agent settings. Across four simulations, the model reproduced the finding that automatic imitation becomes stronger when seeing multiple identical actions (Cracco & Brass, 2018a, 2018c; Cracco et al., 2015) but weaker when seeing multiple different actions (Cracco & Brass, 2018b; Cracco et al., 2015). Importantly, however, to capture the relationship between group size and automatic imitation (Cracco & Brass, 2018c), it was necessary to extend the model with a mechanism that inhibits the motor system based on the amount of visual input. In a fifth simulation, it was shown that this mechanism could be driven by response conflict (Botvinick et al., 2001).

4.1. Theoretical Foundation

The dual-route model is grounded in the associative sequence learning theory (ASL) of automatic imitation (Heyes, 2011). This theory argues that automatic imitation is driven by learned

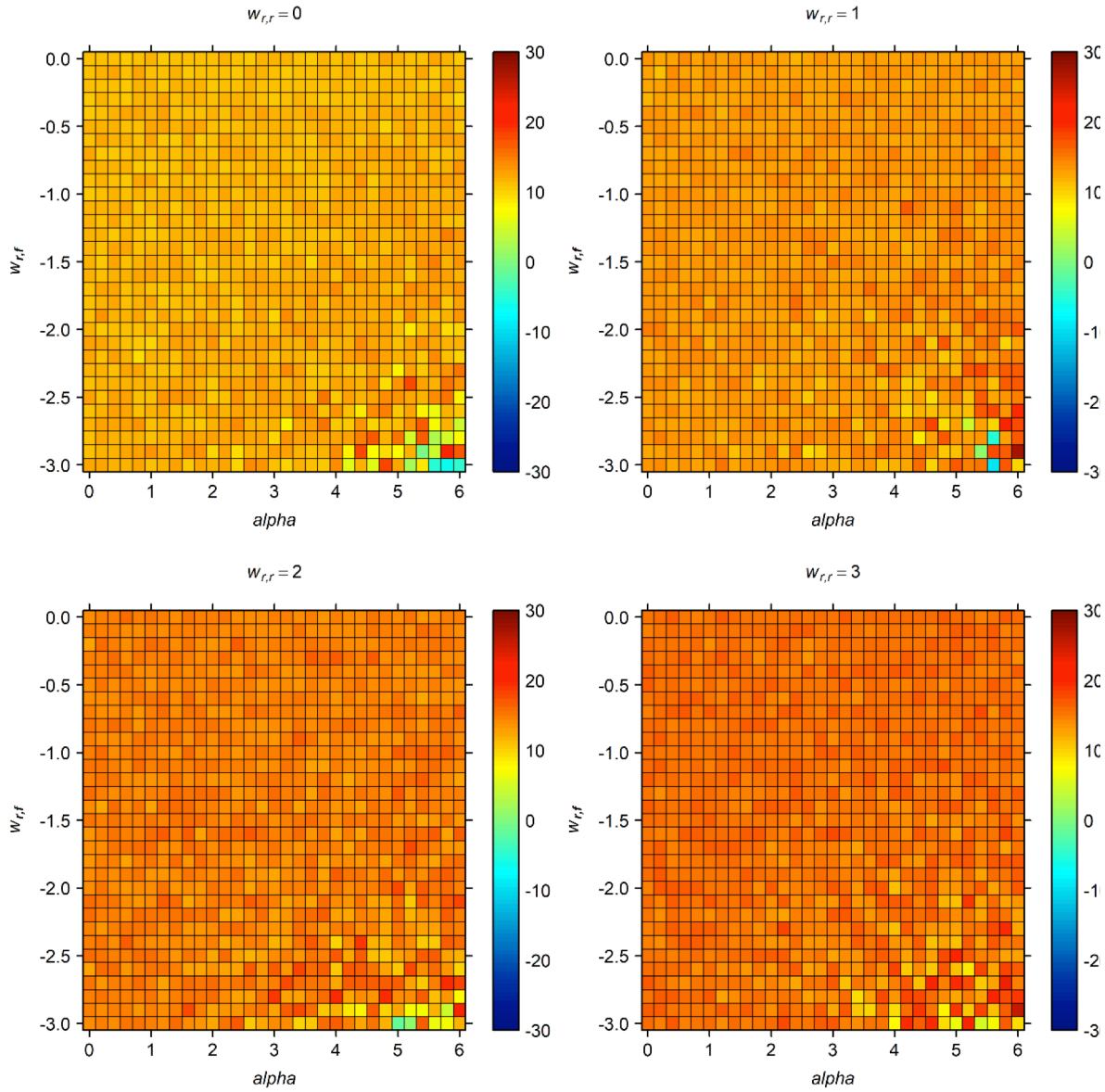


Figure 16. Parameter variation results for Simulation 5. Each cell shows the minimum of the difference between the 3H ID and 3H/1H DIFF and between the 4H ID and 3H/1H DIFF congruency effects for different combinations of w_{rr}, α, and w_{rf}. Positive values indicate that the congruency effect is smaller in the 3H/1H DIFF condition than in both the 3H ID and 4H ID conditions. The difference between the 3H ID and 4H ID congruency effects is shown in Supplementary Material (Figure S9).

associations between the visual and motor representation of an action (Brass & Heyes, 2005; Heyes, 2010). Specifically, it argues that action execution has perceivable sensory consequences that over time become associated with the motor command that produced them. Because people often see the actions they execute, this results in bidirectional connections between perception and action, and these connections are thought to form the basis of automatic imitation (Cracco, Bardi, et al., 2018; Heyes, 2011). Even though the dual-route model is in principle agnostic about the origin of perception-action connections (Cooper, Cook, et al., 2013), it captures the associative sequence learning assumption that these connections are not different from the S-R connections behind non-imitative phenomena such as spatial compatibility (Catmur & Heyes, 2011; Cooper, Catmur, et al., 2013).

However, ASL is not the only theory of automatic imitation. For example, a second prominent theory is ideomotor theory (IM; Brass & Heyes, 2005; Greenwald, 1970). In line with ASL, this theory also posits that perception-action links are acquired through associative learning. However, IM does not consider these links to be the end point. Instead, it proposes that learning leads to the

formation of ideomotor representations that code actions in terms of their anticipated sensory consequences (Greenwald, 1970). Because the visual image of an action is part of its sensory consequences, this implies that action execution and action observation share the same representational format (Prinz, 1997), causing observed actions to trigger automatic imitative responses (Cracco, Bardi, et al., 2018).

Inspired by the idea of shared representations, most computational models of IM assume qualitatively different routes for imitative and non-imitative stimuli, with the former relying on direct matching and the latter on S-R translation (Bertenthal & Scheutz, 2013; Sauser & Billard, 2006; Scheutz & Bertenthal, 2012). Other IM models instead put forward a more general mechanism. For instance, the HiTEC model argues that not only imitative but also non-imitative stimuli trigger ideomotor representations (Haazebroek, Raffone, & Hommel, 2016). To do so, it assumes the existence of *feature nodes* that code stimulus dimensions such as spatial location (e.g., left or right) or finger identity (e.g., index or middle finger) in an a-modal format and act as intermediate nodes connecting *sensory nodes* to *response nodes* (Haazebroek et al., 2016). As such, they can be seen as ideomotor representations controlling motor execution regardless of stimulus format (Haazebroek et al., 2016).

Simulations have shown that not only ASL-based models (Cooper, Catmur, et al., 2013; Cooper, Cook, et al., 2013) but also IM-based models (Bertenthal & Scheutz, 2013; Sauser & Billard, 2006; Scheutz & Bertenthal, 2012) can account for automatic imitation. Applied to multi-agent settings, there is similarly little reason to believe that ASL- and IM-based models would lead to different results. This is not surprising considering that ASL and IM share the same set of core assumptions. Indeed, both theories differ only in whether or not associative learning culminates in ideomotor representations (Cracco, Bardi, et al., 2018) and therefore rarely (if ever) make different predictions (Catmur, Press, Cook, Bird, & Heyes, 2014). Thus, our aim was not to compare ASL and IM but rather to investigate, using the dual-route model, how existing theories can be modified to account for automatic imitation beyond the dyad. As neither ASL nor IM readily include the architecture needed to explain the asymptotic relation between group size and automatic imitation (Cracco & Brass, 2018c), an important contribution of the present work is that it brings theorizing on automatic imitation forward by showing that existing theories have to be extended with a mechanism that inhibits responses based on the amount of visual input in order to account for the full range of results.

4.2. Input-Driven Inhibition

Given that input inhibition is a core feature of the model, it is important to consider its feasibility. Regulating response inhibition in accordance with task demands is widely considered a hallmark of cognitive control (Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010; Botvinick et al., 2001; Bugg & Crump, 2012). However, to inhibit responses based on the number of moving hands, cognitive control has to be a fast process. Supporting this view, research suggests that control processes operate not only across (Botvinick et al., 2001, 2004) but also within trials (Abrahamse, Braem, Notebaert, & Verguts, 2016). For example, a recent study found that reward cues occurring either just before or together with the target can speed up responses by increasing cognitive control (Janssens, De Loof, Pourtois, & Verguts, 2016). Based on this evidence, it seems plausible that response inhibition can be regulated flexibly on each trial based on the amount of visual input (Cracco & Brass, 2018c).

Interestingly, a similar input-dependent inhibition mechanism was also proposed in earlier computational work (Houghton et al., 1996). In particular, Houghton et al. (1996) modeled an inhibitory mechanism that keeps the motor system from producing a response until top-down processes determine whether or not to act on the external input. Importantly, this inhibitory mechanism was designed to mirror the excitatory mechanism so that increases in input cause corresponding increases in inhibition. Supporting this model, and pertinent to automatic imitation, research has shown that motor resonance during action observation produces an inhibitory rebound that mirrors the eliciting motor response in terms of strength (Schuch et al., 2010). This points towards a reactive inhibitory mechanism attempting to prevent external input from gaining control over action. As the external input increases, inhibition has to increase accordingly, and this is visible

in the form of larger inhibitory rebounds following stimulus offset (Houghton et al., 1996; Schuch et al., 2010).

In addition to implementing a similar inhibitory mechanism, the current study also provides a potential explanation for what drives it: the regulation of motor inhibition based on response conflict. Previous computational research has shown that such a conflict adaptation mechanism can account for a wide range of cognitive effects (Botvinick et al., 2001), including the finding that the proportion of incongruent trials modulates S-R compatibility (Bugg & Crump, 2012). Specifically, decreasing the proportion of incongruent trials is thought to reduce response conflict, causing the system to loosen control, and therefore strengthening congruency effects (Botvinick et al., 2001). Simulation 4 indicates that a similar mechanism might also explain why response speed on congruent trials decreased asymptotically when the proportion of congruent to incongruent trials was balanced, but monotonically when congruent trials were presented in isolation (Cracco & Brass, 2018c). That is, response conflict on incongruent trials increases motor inhibition, but when there are no incongruent trials, there is no response conflict, and therefore little inhibition. As such, the present model identifies a potential mechanism through which input-driven inhibition may be established.

4.3. Implications

Given that automatic imitation constitutes a laboratory model of the sensorimotor processes involved in social interaction (Cracco, Bardi, et al., 2018; Cracco & Brass, 2019; Heyes, 2011), our results further have important implications for research on social group phenomena such as social contagion. Specifically, research on this topic has shown that the behavior of others is contagious, and that the degree of contagiousness depends on the number of persons taking part (Darley & Latané, 1968; Latane, 1981; Milgram, Bickman, & Berkowitz, 1969). For example, Milgram et al. (1969) found in a field study that people were increasingly more likely to copy a group of confederates looking up as the group became larger (see also Capozzi, Bayliss, & Ristic, 2018; Gallup et al., 2012; Knowles & Bassett, 1976; Sun, Yu, Zhou, & Shen, 2017). As argued by the authors, a sensible explanation for this result is that larger groups are imitated more often because they are more likely to be looking at something interesting (Milgram et al., 1969). However, based on the current study, an alternative hypothesis might also be that larger groups provide a stronger trigger to the motor system, leading to stronger imitative responses (Cracco & Brass, 2018a, 2018c; Cracco et al., 2015, 2016).

Furthermore, the current model might also explain why social contagion increases asymptotically rather than linearly with group size (Cracco & Brass, 2018c). That is, it is well-known that contagiousness initially increases, but then plateaus as group size reaches three or four members (Bond, 2005; Gallup et al., 2012; Milgram et al., 1969). While this has typically been explained in terms of social cognitive processes (Bond, 2005; Latane, 1981; MacCoun, 2012), it can also be explained by dynamic regulation of the response threshold (Cracco & Brass, 2018c). Indeed, adjusting the response threshold based on the number of agents that look up would likewise lead to an asymptotic curve between group size and imitation (Milgram et al., 1969). As such, by exploring the mechanisms underlying imitation in multi-agent settings, the present study opens up novel hypotheses regarding the interaction of sensorimotor and interpretative processes in social group contagion.

However, from this perspective, a key question is whether the model can be extended beyond the sensorimotor domain. Indeed, research indicates that social contagion is not limited to action but equally occurs for inaction (Darley & Latané, 1968; Fischer et al., 2011) or emotion (Du, Fan, & Feng, 2014). For instance, research on the bystander effect has shown that people become less likely to help someone in need when other bystanders do not intervene, and this effect is known to increase with the number of passive bystanders (Fischer et al., 2011). While it is difficult to see how inaction can be mirrored, research has shown that not only the actions but also the emotions and mental states of others are mirrored (Keysers & Gazzola, 2006). This suggests, in other words, that various forms of social contagion may share the same functional mechanism (Raafat, Chater, & Frith, 2009). Supporting this view, a recent neuroimaging study found that activation in the motor system decreased as the number of passive bystanders increased, indicating that participants embodied the mental state of the bystanders, causing them to inhibit helping responses (Hortensius & De Gelder, 2014).

In sum, the current model sees social contagion as a sensorimotor rather than as an interpretative process. However, this does not mean that interpretation plays no role at all in social contagion. Instead, it may contribute primarily at later stages of processing. For instance, it could act as a gating mechanism that facilitates or inhibits imitation depending on the social context. In support of this view, neuroimaging has revealed that the conscious decision to imitate is associated with the gating of mirror activity (Bien, Roebroeck, Goebel, & Sack, 2009). Similarly, it has been shown that automatic imitation is inhibited when it would result in the execution of a taboo gesture, and that this depends on the degree to which the gesture fits in the social context (Cracco, Genschow, et al., 2018).

4.4. Limitations

The current model also has a number of limitations. A first limitation is that, following most previous computational work, we focused on modeling mean performance and did not model performance variability. More specifically, even though noise was added to the input on each trial, we did not vary the parameter values across participants. This means that no interindividual differences were included, in contrast to real life, where such differences are the norm. The rationale behind this decision is twofold. First, such differences would make the results noisier, therefore making it more difficult to identify the model's typical behavior. Second, we did not have a strong hypothesis regarding which parameters should be sensitive to interindividual differences and how strong the influence of interindividual differences should be for each parameter. Nevertheless, one might wonder whether the model is at all able to reproduce the variability observed in the empirical data. To test this hypothesis, we introduced interindividual variability in the w_{ri} (SD = 0.70), w_{rh} (SD = 0.50), and ρ_h (SD = 0.02) parameters by drawing them for each participant from a normal distribution with the parameter's default value as mean and the listed values as standard deviation. As shown in Supplementary Material (Figures S10-S14), the adjusted model now reproduced not only the mean but also the dispersion of the empirical data. Importantly, this should not be taken as evidence that these parameters are the only three parameters that vary across participants, nor that the model includes all possible sources of variability that could influence performance in real life. However, it does show that the model can be easily adjusted to reproduce not only the central tendency but also the variability of the empirical data.

A second limitation of the model is that it has a relatively large number of free parameters. To address this limitation, most parameter values were taken directly from Cooper et al. (2013), therefore greatly limiting our degrees of freedom. When this was not possible, we either followed the same procedure as Cooper et al. (2013) to set the parameter values (i.e., ρ_{h}) or we explored model performance across parameter space using parameter variation studies (i.e., α, w_{rf}, w_{rr}, and δ_f). This revealed that all simulations were robust to reasonable variation in the parameter values. In other words, by limiting modeler degrees of freedom and through parameter variation studies, we were able to obtain robust results in spite of the difficulties associated with having many free parameters.

Finally, it could be argued that a simpler way of implementing the input-driven inhibition mechanism would have been to inhibit motor activity based on the combined activity in the hand nodes. However, such a mechanism assumes that inhibition is slow in the sense that it has to wait until the hand movements are processed to become active. This seems unlikely, considering that visual processing is hierarchical with basic information like motion being processed before complex information like the identity of objects or actions (Van Essen & Maunsell, 1983). Nevertheless, consistent with the finding that model performance was relatively robust to changes in the δ_f parameter (Figure S4-S5), additional simulations showed that inhibiting responses based on activity in the hand nodes would likewise have been able to account for why congruent RTs decreased asymptotically in the mixed but not in the non-mixed scenario (Figures S6-S8). This suggests that dynamic response inhibition, regardless of how it is implemented, can explain the results of Cracco and Brass (2018c).

4.5. Conclusion

In conclusion, we extended the dual-route model of automatic imitation to multi-agent settings to test whether it can account for four multi-agent imitation effects previously reported in the literature (Cracco & Brass, 2018a, 2018b, 2018c; Cracco et al., 2015). The results revealed that all four effects could be reproduced provided that the model was extended with an inhibitory mechanism regulating the amount of motor inhibition based on the number of observed actions. Furthermore, an additional simulation indicated that this mechanism could be driven by response conflict. The current study critically extends theories of automatic imitation from single- to multi-agent settings and as such constitutes an important step towards the understanding of social interaction beyond the dyad.

References

Abrahamse, E. L., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychological Bulletin*, *142*(7), 693–728. https://doi.org/10.1037/bul0000047

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131. https://doi.org/10.1016/j.tics.2011.01.003

Bertenthal, B. I., & Scheutz, M. (2013). In Praise of a Model but Not Its Conclusions: Commentary on Cooper, Catmur, and Heyes (2012). *Cognitive Science*, *37*(4), 631–641. https://doi.org/10.1111/cogs.12039

Bien, N., Roebroeck, A., Goebel, R., & Sack, A. T. (2009). The brain's Intention to Imitate: The Neurobiology of Intentional versus Automatic Imitation. *Cerebral Cortex*, *19*(10), 2338–2351. https://doi.org/10.1093/cercor/bhn251

Bogacz, R., Wagenmakers, E. J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences*, *33*(1), 10–16. https://doi.org/10.1016/j.tins.2009.09.002

Bond, R. (2005). Group size and conformity. *Group Processes & Intergroup Relations*, *8*(4), 331–354. https://doi.org/10.1177/1368430205056464

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict Monitoring and Cognitive Control. *Psychological Review*, *108*(3), 624–652.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*(12), 539–546. https://doi.org/10.1016/j.tics.2004.10.003

Brass, M., Bekkering, H., Wohlschlager, A., Prinz, W., Wohlschläger, A., & Prinz, W. (2000). Compatibility between observed and executed finger movements: Comparing symbolic, spatial, and imitative cues. *Brain and Cognition*, *44*(2), 124–143. https://doi.org/10.1006/brcg.2000.1225

Brass, M., Derrfuss, J., Cramon, G. M. V, Matthes-von Cramon, G., & von Cramon, D. Y. (2003). Imitative response tendencies in patients with frontal brain lesions. *Neuropsychology*, *17*(2), 265–271. https://doi.org/10.1037/0894-4105.17.2.265

Brass, M., & Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence problem? *Trends In Cognitive Sciences*, *9*(10), 489–495. https://doi.org/10.1016/j.tics.2005.08.007

Bugg, J. M., & Crump, M. J. C. (2012). In support of a distinction between voluntary and stimulus-driven control: A review of the literature on proportion congruent effects. *Frontiers in Psychology*, *3*, 1–16. https://doi.org/10.3389/fpsyg.2012.00367

Butler, E. E., Ward, R., & Ramsey, R. (2015). Investigating the relationship between stable personality characteristics and automatic imitation. *PLoS ONE*, *10*(6), 1–18. https://doi.org/10.1371/journal.pone.0129651

Capozzi, F., Bayliss, A. P., & Ristic, J. (2018). Gaze following in multi-agent contexts: Evidence for a quorum-like principle. *Psychonomic Bulletin & Review*, *Review*. https://doi.org/10.3758/s13423-018-1464-3

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*, *50*(3), 1148–1167. https://doi.org/10.1016/j.neuroimage.2009.12.112

Catmur, C., & Heyes, C. (2011). Time Course Analyses Confirm Independence of Imitative and Spatial Compatibility. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 409–421. https://doi.org/10.1037/a0019325

Catmur, C., Press, C., Cook, R., Bird, G., & Heyes, C. (2014). Mirror neurons: Tests and testability. *Behavioral and Brain Sciences*, *37*(2), 221–241. https://doi.org/10.1017/S0140525X13002793

Catmur, C., Thompson, E. L., Bairaktari, O., Lind, F., & Bird, G. (2017). Sensorimotor Training Alters Action Understanding. *Cognition*.

Chartrand, T. L., & Bargh, J. A. (1999). The Chameleon Effect: The Perception-Behavior Link and Social Interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910.

Chartrand, T. L., & Lakin, J. L. (2013). The Antecedents and Consequences of Human Behavioral Mimicry. *Annual Review of Psychology, Vol 64*, *64*, 285–308. https://doi.org/10.1146/annurev-psych-113011-143754

Colling, L. J., Knoblich, G. G., & Sebanz, N. (2013). How does "mirroring" support joint action? *Cortex*, *49*(10), 2964–2965. https://doi.org/10.1016/j.cortex.2013.06.006

Cooper, R. P., Catmur, C., & Heyes, C. (2013). Are Automatic Imitation and Spatial Compatibility Mediated by Different Processes? *Cognitive Science*, *37*(4), 605–630. https://doi.org/10.1111/j.1551-6709.2012.01252.x

Cooper, R. P., Cook, R., Dickinson, A., & Heyes, C. (2013). Associative (not Hebbian) learning and the mirror neuron system. *Neuroscience Letters*, *540*, 28–36. https://doi.org/10.1016/j.neulet.2012.10.002

Cracco, E., Bardi, L., Desmet, C., Genschow, O., Rigoni, D., De Coster, L., … Brass, M. (2018). Automatic Imitation: A meta-analysis. *Psychological Bulletin*, *144*(5), 453–500. https://doi.org/10.1037/bul0000143

Cracco, E., & Brass, M. (2018a). Automatic imitation of multiple agents: Simultaneous or random representation? *Journal of Experimental Psychology: Human Perception & Performance*, *44*(5), 729–740. https://doi.org/10.1037/xhp0000489

Cracco, E., & Brass, M. (2018b). Motor Simulation of Multiple Observed Actions. *Cognition*, *180*, 200–205. https://doi.org/10.1016/j.cognition.2018.07.007

Cracco, E., & Brass, M. (2018c). The role of sensorimotor processes in social group contagion. *Cognitive Psychology*, *103*, 23–41.

Cracco, E., & Brass, M. (2019). Reaction time indices of automatic imitation measure imitative response tendencies. *Consciousness and Cognition*, *68*, 115–118.

Cracco, E., De Coster, L., Andres, M., & Brass, M. (2015). Motor simulation beyond the dyad: Automatic imitation of multiple actors. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1488–1501. http://dx.doi.org/10.1037/a0039737

Cracco, E., De Coster, L., Andres, M., & Brass, M. (2016). Mirroring multiple agents: Motor resonance during action observation is modulated by the number of agents. *Social Cognitive and Affective Neuroscience*, *11*(9), 1422–1427. https://doi.org/10.1093/scan/nsw059

Cracco, E., Genschow, O., Radkova, I., & Brass, M. (2018). Automatic imitation of pro- and antisocial gestures: Is implicit social behavior censored? *Cognition*, *170C*, 179–189. https://doi.org/10.1016/j.cognition.2017.09.019

Cracco, E., Keysers, C., Clauwaert, A., & Brass, M. (2018). Representing Multiple Observed Actions in the Motor System. *Cerebral Cortex*.

Darley, J. M., & Latané, B. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology*, *10*(3), 215–221. https://doi.org/10.1037/h0026570

De Renzi, E., Cavalleri, F., & Facchini, S. (1996). Imitation and utilisation behaviour. *Journal of Neurology, Neurosurgery, and Psychiatry*, *61*(4), 396–400. https://doi.org/10.1136/jnnp.61.4.396

DiCarlo, J. J., & Maunsell, J. H. R. (2003). Anterior Inferotemporal Neurons of Monkeys Engaged in Object Recognition Can be Highly Sensitive to Object Retinal Position. *Journal of Neurophysiology*, *89*(6), 3264–3278. https://doi.org/10.1152/jn.00358.2002

Du, J., Fan, X., & Feng, T. (2014). Group Emotional Contagion and Complaint Intentions in Group Service Failure: The Role of Group Size and Group Familiarity. *Journal of Service Research*, *17*(3), 326–338. https://doi.org/10.1177/1094670513519290

Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., … Kainbacher, M. (2011). The bystander-effect: a meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin*, *137*(4), 517–537. https://doi.org/10.1037/a0023304

Gallup, A. C., Hale, J. J., Sumpter, D. J. T., Garnier, S., Kacelnik, A., Krebs, J. R., & Couzin, I. D. (2012). Visual attention and the acquisition of information in human crowds. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(19), 7245–7250. https://doi.org/10.1073/pnas.1116141109

Genschow, O., Bossche, S. Van Den, Cracco, E., Bardi, L., Rigoni, D., Brass, M., … Brass, M. (2017). Mimicry and automatic imitation are not correlated. *PLoS ONE*, *12*(9), 1–21. https://doi.org/10.1371/journal.pone.0183784

Gillmeister, H., Catmur, C., Liepelt, R., Brass, M., & Heyes, C. (2008). Experience-based priming of body parts: A study of action imitation. *Brain Research*, *1217*, 157–170. https://doi.org/10.1016/j.brainres.2007.12.076

Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: with special reference to the ideo-motor mechanism. *Psychological Review*, *77*(2), 73–99. https://doi.org/10.1037/h0028689

Haazebroek, P., Raffone, A., & Hommel, B. (2016). HiTEC: a connectionist model of the interaction between perception and action planning. *Psychological Research*, *81*(6), 1–25. https://doi.org/10.1007/s00426-016-0803-0

Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews*, *34*(4), 575–583. https://doi.org/10.1016/j.neubiorev.2009.11.007

Heyes, C. (2011). Automatic imitation. *Psychological Bulletin*, *137*(3), 463–483. https://doi.org/10.1037/a0022288

Hortensius, R., & De Gelder, B. (2014). The neural basis of the bystander effect - The influence of group size on neural activity when witnessing an emergency. *NeuroImage*, *93*(P1), 53–58. https://doi.org/10.1016/j.neuroimage.2014.02.025

Houghton, G., Tipper, S. P., Weaver, B., & Shore, D. I. (1996). Inhibition and Interference in Selective Attention : *Visual Cognition*, *3*(2), 119–164.

Janssens, C., De Loof, E., Pourtois, G., & Verguts, T. (2016). The time course of cognitive control implementation. *Psychonomic Bulletin and Review*, *23*, 1266–1272. https://doi.org/10.3758/s13423-015-0992-3

Keysers, C., & Gazzola, V. (2006). Towards a unifying neural theory of social cognition. In Anders, Ende, Junghöfer, Kissler, & Wildgruber (Eds.), *Progress in Brain Research* (Vol. 156, pp. 379–401). https://doi.org/10.1016/S0079-6123(06)56021-2

Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Current Biology*, *13*(6), 522–525. https://doi.org/10.1016/S0960-9822(03)00165-9

Knowles, E. S., & Bassett, R. L. (1976). Groups and crowds as social entities: Effects of activity, size, and member similarity on nonmembers. *Journal of Personality and Social Psychology*, *34*(5), 837–845. https://doi.org/10.1037/0022-3514.34.5.837

Latane, B. (1981). The Psychology of Social Impact. *American Psychologist*, *36*(4), 1–14. https://doi.org/10.1037/0003-066X.36.4.343

Leighton, J., & Heyes, C. (2010). Hand to Mouth: Automatic Imitation Across Effector Systems. *Journal of Experimental Psychology-Human Perception and Performance*, *36*(5), 1174–1183. https://doi.org/10.1037/a0019953

Lhermitte, F., Pillon, B., & Serdaru, M. (1986). Human autonomy and the frontal lobes. Part I: Imitation and utilization behavior: a neuropsychological study of 75 patients. *Annals of Neurology*, *19*(4), 326–334. https://doi.org/10.1002/ana.410190404

MacCoun, R. J. (2012). The burden of social proof: Shared thresholds and social influence. *Psychological Review*, *119*(2), 345–372. https://doi.org/10.1037/a0027121

McClelland, J. L. (1993). The GRAIN model: A framework for modeling the dynamics of information processing. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 655–688). Hillsdale, NJ: Erlbaum.

Milgram, S., Bickman, L., & Berkowitz, L. (1969). Note on the drawing power of crowds of different size. *Journal of Personality and Social Psychology*, *13*(2), 79–82. https://doi.org/10.1037/h0028070

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences*, *15*(10), 453–459. https://doi.org/10.1016/j.tics.2011.07.011

Press, C., Richardson, D., & Bird, G. (2010). Intact imitation of emotional facial actions in autism spectrum conditions. *Neuropsychologia*, *48*(11), 3291–3297. https://doi.org/10.1016/j.neuropsychologia.2010.07.012

Prinz, W. (1997). Perception and Action Planning. *European Journal of Cognitive Psychology*, *9*(2), 129–154.

Raafat, R. M., Chater, N., & Frith, C. (2009). Herding in humans. *Trends in Cognitive Sciences*, *13*(10), 420–428. https://doi.org/10.1016/j.tics.2009.08.002

Ramenzoni, V. C., Sebanz, N., & Knoblich, G. (2014). Scaling Up Perception-Action Links: Evidence From Synchronization With Individual and Joint Action. *Journal of Experimental Psychology-Human Perception and Performance*, *40*(4), 1551–1565. https://doi.org/10.1037/a0036925

Sauser, E. L., & Billard, A. G. (2006). Parallel and distributed neural models of the ideomotor principle: An investigation of imitative cortical pathways. *Neural Networks*, *19*(3), 285–298. https://doi.org/10.1016/j.neunet.2006.02.003

Sayres, R., & Grill-Spector, K. (2008). Relating Retinotopic and Object-Selective Responses in Human Lateral Occipital Cortex. *Journal of Neurophysiology*, *100*(1), 249–267. https://doi.org/10.1152/jn.01383.2007

Scheutz, M., & Bertenthal, B. I. (2012). A Computational PDP Model for Explaining Automatic Imitation. *Cognitive Science*, *17*, 2288–2293.

Schuch, S., Bayliss, A. P., Klein, C., & Tipper, S. P. (2010). Attention modulates motor system activation during action observation: Evidence for inhibitory rebound. *Experimental Brain Research*, *205*(2), 235–249. https://doi.org/10.1007/s00221-010-2358-4

Stanley, J., Gowen, E., & Miall, R. C. (2007). Effects of agency on movement interference during observation of a moving dot stimulus. *Journal of Experimental Psychology-Human Perception and Performance*, *33*(4), 915–926. https://doi.org/10.1037/0096-1523.33.4.915

Sturmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence effects with manual gestures and postures: A study of imitation. *Journal of Experimental Psychology-Human Perception and Performance*, *26*(6), 1746–1759. https://doi.org/10.1037//0096-1523.26.6.1746

Sun, Z., Yu, W., Zhou, J., & Shen, M. (2017). Perceiving crowd attention: Gaze following in human crowds with conflicting cues. *Attention, Perception, and Psychophysics*, *79*(4), 1039–1049. https://doi.org/10.3758/s13414-017-1303-z

Tsai, J. C. C., Sebanz, N., & Knoblich, G. G. (2011). The GROOP effect: Groups mimic group actions. *Cognition*, *118*(1), 135–140. https://doi.org/10.1016/j.cognition.2010.10.007

Van Essen, D. C., & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, *6*, 370–375. https://doi.org/10.1016/0166-2236(83)90167-4

Wiggett, A. J., Downing, P. E., & Tipper, S. P. (2013). Facilitation and interference in spatial and body reference frames. *Experimental Brain Research*, *225*(1), 119–131. https://doi.org/10.1007/s00221-012-3353-8

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, *131*(3), 460–473. https://doi.org/10.1037/0033-2909.131.3.460

Yoshor, D., Bosking, W. H., Ghose, G. M., & Maunsell, J. H. R. (2007). Receptive fields in human visual cortex mapped with surface electrodes. *Cerebral Cortex*, *17*(10), 2293–2302. https://doi.org/10.1093/cercor/bhl138