Incorporating Social Psychological Theories in the Model Training Regime: How Neural Representations for Social Cognition Emerge from Interactions with Others

Taiji Ueno[1,2] **(taijiueno7@gmail.com),**
Saori Tsukamoto[1,2] **(tsukamoto.saori@h.mbox.nagoya-u.ac.jp)**
Tokika Kurita[1,2] **(kurita.tokika@b.mbox.nagoya-u.ac.jp),**
Minoru Karasawa[1] **(mkarasawa@nagoya-u.jp)**
1. Department of Psychology, Graduate School of Environmental Studies, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya City, Aichi 4648601, JAPAN. 2. Japan Society for the Promotion of Science

Abstract

Social psychology aims to reveal how social behaviors are acquired through interactions with others (i.e., past interpersonal experiences) whereas social neuroscience investigates the neural substrates that correlate with acquired social behaviors. For example, people with greater ingroup bias are known to avoid or have avoided interactions with outgroup members than those with weaker ingroup bias, and their brain activation patterns are more distinct when viewing an ingroup member from an outgroup member. The present study aimed to examine the causal relation of these findings from different disciplines and integrate them within a single framework. A connectionist model was trained with/without the training regime reflecting the interpersonal experiences that were assumed to increase ingroup bias. As a result, if trained with such a training environment, the model's internal representations of ingroup exemplars were more distinct from those of outgroup exemplars. Thus, this model reproduced the dissimilarity structure in the neural representations of ingroup bias. In contrast, training without such a regime alleviated the representation dissimilarities.

Keywords: ingroup bias; connectionist model; contact theory; social psychology; multi-voxel pattern analysis

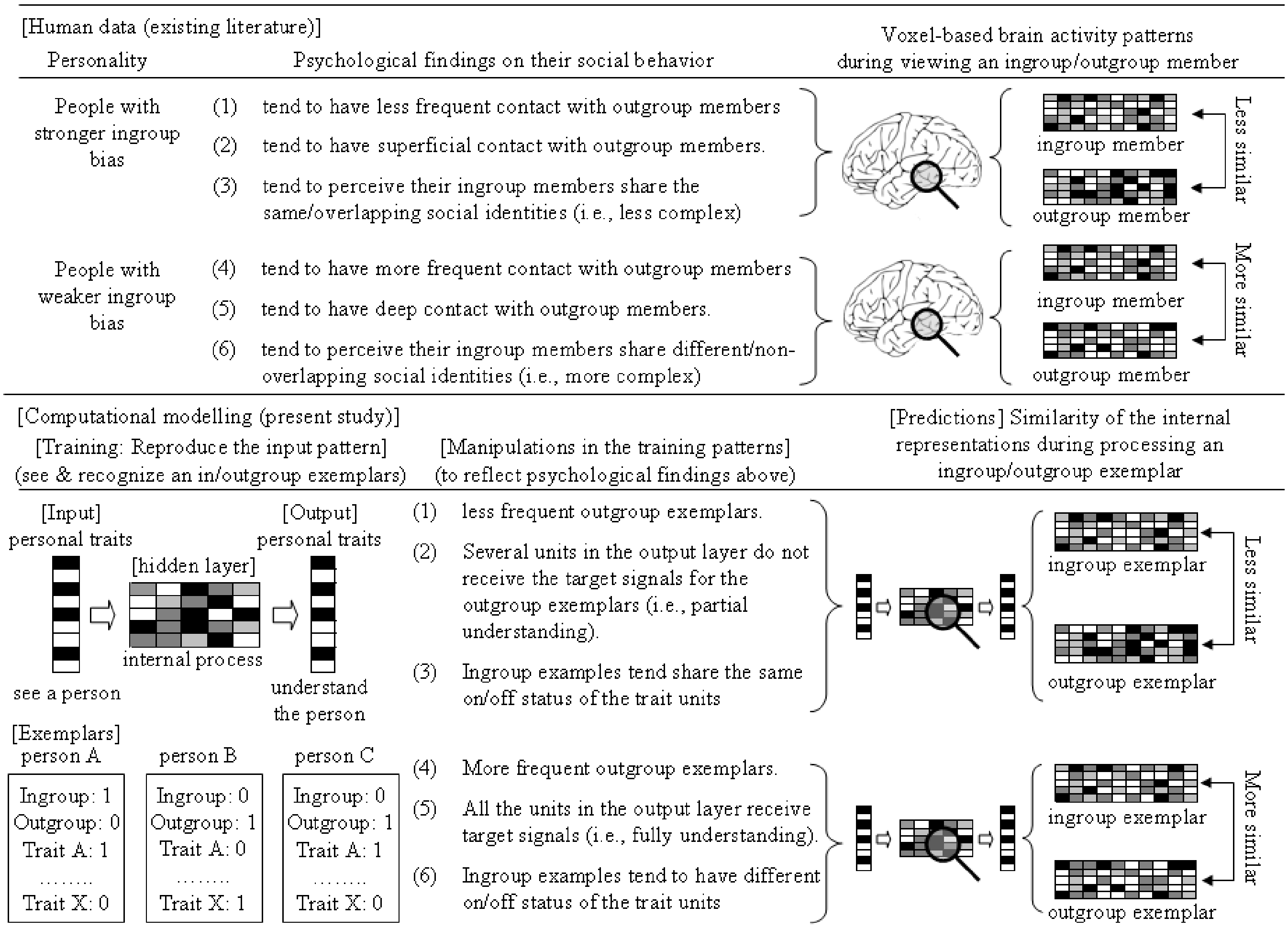
Introduction

Social psychological studies aim to clarify how an individual's behaviors towards others (i.e., social cognition) are acquired and how they change through further interactions with others. For example, one of the widely investigated phenomena in social cognition is ingroup bias (or outgroup prejudice). Seminal works of Gordon Allport hypothesized that intimate contact with outgroup members would reduce outgroup prejudice, known as *contact hypothesis* (Allport, 1954/1979). Supporting evidence for this theory has been documented in the last 60 years of psychology, suggesting that there is a significant negative correlation between the contact frequency and the degree of prejudice towards others (Pettigrew & Tropp, 2008). Furthermore, constructive critics have refined the theory by narrowing the boundary conditions. For example, the nature of contact modulates its effect, such that a superficial contact (e.g., just living nearby) could attenuate its effect (Kanas, Sterkens, & Scheepers, 2013), which Allport (1954/1979) also predicted (top left corner of Table 1).

In parallel to these psychological studies, cognitive neuroscience studies have also revealed the neural correlates of social cognition, including ingroup bias. They revealed not only the brain region involved in social cognition but also how the neural activities change in different social contexts (Molenberghs, 2013). Furthermore, recent advance has allowed to clarify even the *nature of neural representations* that may underlie ingroup bias (Brosch, Bar-David, & Phelps, 2013; Gilbert, Swencionis, & Amodio, 2012). For example, Brosch et al. (2013) employed a multi-voxel pattern analysis methodology and found that stronger implicit ingroup bias (pro-White bias) increases the *dissimilarities in the neural representations* of ingroup and outgroup faces (top-right corner of Table 1).

Whilst the findings from both disciplines are clearly important for understanding ingroup bias, a key question that has yet to be answered is how relevant these findings are. In other words, would neural representations for ingroup and outgroup faces become similar as people interact with outgroup members more frequently? Alternatively, have those with such dissimilar neural representations experienced less frequent contact with outgroup members than those with similar neural representations? To test these possibilities, we would need longitudinal or retrospective data. However, such data cannot exclude potential influence of confounding variables. On this point, a computational model with a learning algorithm (i.e., it can gradually develop) provides us a unique opportunity to overcome such a methodological limitation. It is possible for a modeler to control various extraneous variables and to manipulate a training regime that reflects a social psychological finding (e.g., a model processes outgroup exemplars more/less frequently; a model processes homogeneous/heterogeneous ingroup members, etc.). Then, a modeler can directly investigate the internal representation (i.e., pattern activation in the hidden layer) of the model to test whether such a manipulation modulates the dissimilarities in the representations for ingroup and outgroup exemplars (neuroscience finding). Such an endeavor to simulate fMRI findings in a connectionist model is not very common but is proved to be fruitful in other domains (Cowell & Cottrell, 2013). There are many social psychological findings that have been suggested to affect ingroup bias. The present study aimed to examine the impact of the frequency/nature of contact, identity, and cultural trend (Details will be described in a later section).

Table 1. Human psychological and neuroscience data & the description of the modelling approach and predictions



Method

The bottom half of Table 1 depicts the general modeling approach and the predictions in this study. Various kinds of computational models have been implemented in social psychology to demonstrate the cognitive machinery that can reproduce social cognitive behaviors. These vary in terms of (a) how to implement interactions with others; (b) the explanation of how to acquire social behaviors (c) whether the target cognition is grounded on distributed or localist representations, and so on. From these viewpoints, our model falls into a variant of Tensor-product model by Kashima and his colleagues (Kashima, Woolcock, & Kashima, 2000) and the autoassociative network model by Smith and Decoster (1998). These models focused on the computational operation of an individual, which processes (recognize) the input vectors that represent ingroup and outgroup exemplars. As the model processes the exemplars one by one, it gradually adjusts the connection strength to represent the ingroup and outgroup exemplars as distributed patterns in the internal layer. The modelers analyzed these distributed representations to test if the model reproduced social behaviors or not. For example, the model was trained in a close situation as a psychological experiment of group categorization, and the similarities in the internal activation pattern of one group member from another were taken as a measure of group categorization performance. Thus, importantly, these models were never trained for the target social cognition itself, but for recognizing others. Nevertheless, various social cognitive behaviors came out as *an emergent property* of the interaction with others.

These models have all the characteristics that are necessary for the current study. In this study, a three-layer feedforward network model (bottom-left corner of Table 1) received an input vector that represented an ingroup or outgroup exemplar. These simple vector patterns and the architecture were chosen for simplification (actually, it is a strength of the model to reproduce the target behavior under simplification), but a future target would be to make the model more realistic. The activation spread from left to right (The activity of each unit was a sigmoid function of the summed weighted input from other units). Then, the model was trained for reproducing the input pattern in the output layer. One can say that the model is trained to understand (recognize) the person who meets up. Although sometimes a connectionist model is criticized in terms of its lack of a teacher signal in reality (Baker's paradox), we assumed a target signal is available from the person whom an individual meets up. Nonetheless, we addressed this issue by simulating a situation where this signal was not obvious – thus superficial contact (Simulation 2). The model gradually adjusted its connection strength to reproduce the input pattern in the output layer as closely as possible for all the training patterns. Thus, it is important to emphasize that the model was never trained for anything relevant to ingroup bias. However, a key question we address is that if we train a model in a way social psychologists assume, then does it have an impact on the similarity structure of the acquired internal representations? For example, if the model processes outgroup exemplars more frequently, then the representation patterns for outgroup exemplars become more similar to those for ingroup exemplars? Thus, we incorporated psychological findings into the model training regime (middle column of Table 1), and then tested whether a neural representation for ingroup bias (Brosch et al., 2013) came out as an emergent property of social interactions.

Lens (http://tedlab.mit.edu/~dr/Lens/) was used for all the simulations. Learning rate was set to 0.05. The error derivatives were also scaled down by half for the outgroup exemplars to reflect the less opportunity to interact with them in reality. Weight decay was set to 1E-07. Connection strength was adjusted through back-propagation algorithm after every trial, and adjustment was not made if the target-output difference was below 0.1. Momentum was not used. Gaussian noise (SD = 0.2) was added to the input layer activations to reflect sampling variability (In reality, people do not have 0 or 1 binary values of the traits). An output was scored as correct when the activation in every unit of the output layer was in the correct side of 0.5. Training was terminated when the network reproduced the correct outputs for more than 95% of the training examples.

After training, each ingroup and outgroup exemplar was presented, and the internal representation (hidden layer activation) for each exemplar was recorded. Two measures represented the activation dissimilarities between the ingroup and outgroup exemplars. One was the averaged Euclid distances between all the ingroup-outgroup pairs (higher is more dissimilar). The other involved a cluster analysis on the internal representations. If a hierarchical cluster analysis (Ward) correctly categorized all the exemplars, then a cluster distance was taken as an index of dissimilarity (higher is more dissimilar). If not every pattern was categorized correctly, a non-hierarchical cluster analysis (k-mean) was run, and the entropy index was taken as a dissimilarity index (A smaller entropy means more dissimilar). In each Study, randomly initialized 100 simulations were run, and the results were averaged.

Results

Simulation 1: Contact frequency

The first test case was the effect of contact with outgroup members to reduce outgroup prejudice (Pettigrew & Tropp, 2008). Twenty-two units in the input and output layers were connected via 10 units in the hidden layer. Five ingroup exemplars and five outgroup exemplars (Table 2) were presented to a model. A key manipulation involved how frequently the model processed the outgroup exemplars. In the more frequent model, five ingroup exemplars and five outgroup exemplars were presented alternately. In contrast, in the less frequent model, each of five ingroup exemplars was presented four times before the model encountered each of five outgroup exemplars once.

As a consequence, the mean Euclid distance in the internal representations between the ingroup-outgroup exemplars was 2.544 (SE = 0.007) for the less frequent model and was 2.472 (SE = 0.007) for the more frequent model, $t_{(198)} = 6.637$, $p < .001$. The cluster distance between the groups was 9.84 (SE = 0.12) for the less frequent model and was 9.47 (SE = 0.11) for the more frequent model, $t_{(198)} = 2.204$, $p < .05$. Thus, more frequent contact with outgroup exemplars reduced the internal representation dissimilarities.

Simulation 2: Nature of Contact

Next, not all the contacts were known to be effective, but a superficial contact was predicted to have the opposite effect (Allport, 1954/1979) or has been shown to have a smaller effect (Kanas et al., 2013). To incorporate this theory in the training regime, some of the units in the output layer did not receive a target signal (i.e., zero error derivatives) for the outgroup exemplars (Table 2). This means that the network was not forced to recognize some aspects of the outgroup exemplars. We framed this as the superficial contact model, and compared it to the deep contact model, which was trained to recognize all the aspects of the outgroup exemplar. Eleven units in the input and output layers were connected via eight units in the hidden layer in this model.

As a result, the mean Euclid distance between the ingroup-outgroup exemplars was larger for the superficial contact model, mean = 1.691 (SE = 0.003) than for the deep contact model, mean = 1.604 (SE = 0.005), $t_{(198)} = 13.978$, $p < .001$. Entropy value as a result of a k-mean cluster analysis (k = 3, one cluster for the ingroup, and two clusters for the two outgroups) was 0.62 (SE = 0.03) for the superficial model, and was 0.77 (SE = 0.02) for the deep contact model, $t_{(198)} = 3.476$, $p < .001$. Therefore, superficial contact attenuated the contact effect to reduce the representational dissimilarities between the groups.

Simulation 3: Social identity complexity

In addition to the nature of contact with outgroup, we argue that the nature of contact within ingroup members also matters. Schmid, Hewstonm Tausch, Cairns, and Huges (2009) found a positive correlation between the outgroup contact frequency and the degree of *social identity complexity* (Roccas & Brewer, 2002). Social identity complexity refers to the perceived correlation of one category to another within the ingroup members. Those with low social identity complexity perceive that their ingroup memberships are highly overlapping whereas those with high social identity complexity perceive memberships of their various ingroups are not overlapping. Schmid et al.

(2009) found that people with low identity complexity have a stronger outgroup prejudice than those with high complexity. Our working assumption here is that low in identity complexity means that people have had frequent experiences of meeting ingroup members with the same identities. In contrast, high in identity similarity complexity means that they have frequently encountered with ingroup members with various identities. In order to incorporate our assumption on such past experiences with ingroup members, a key manipulation was made in the training patterns for the ingroup exemplars (Table 2). A high complexity model (NB. less outgroup prejudice) was trained with the same ingroup-outgroup training set as Simulation 1 (thus, exemplars within a group did not share the same on/off status of the units). In contrast, a low complexity model was trained with the different ingroup exemplars. As Table 2 shows, four exemplars shared the same on/off status of the five units.

As a result, the mean Euclid distance in the internal representations between the ingroup-outgroup exemplars was 2.454 (*SE* = 0.009) for low complexity model (higher overlapping categories); and was 2.358 (*SE* = .008) for the high complexity model (fewer overlapping categories), $t_{(198)} = 7.436$, $p < .001$. The cluster distance between the groups was 10.76 (*SE* = 0.15) for the low complexity model, and was 9.01 (*SE* = 0.12) for the similarity complexity model, $t_{(198)} = 8.797$, $p < .001$. Therefore, the internal representation dissimilarity between the ingroup-outgroup members was greater when the model had more frequent experiences to encounter ingroup members with the same social identities. One may argue that it is not empirically supported but just our assumption that those with low social identity complexity have more frequently encountered ingroup members with the same social identities. However, in this way, modelling can provide a possible explanation about why social identity complexity and outgroup prejudice correlate with each other, and provides an explicit question that can be empirically tested in a social psychological study.

Simulation 4: Cultural Context to Follow Others

Finally, even though it is not directly relevant to ingroup bias, it should be desirable to test the generalizability of the current approach to understand other social cognitive neuroscience data. There is another test case for an effect of past interpersonal interactions (to incorporate into a model training regime) on neuroscience data. Specifically, Zhu, Zhang, Fan, and Han (2007) found the neural activity in medial prefrontal cortex was more similar when thinking about self and mothers in Asian culture than Western one. Mayer et al. (2013) conducted a follow-up study and found a deep encoding (e.g., empathy) of other close friends also recruited this area in Asian participants (Meyer et al., 2013), whereas the activation patterns for strangers were different. Then a question here is why such neural representation dissimilarities from strangers (and similarities among close people) differ across cultures? Markus and Kitayama (1991) assimilated various psychological data across continents and argued an effect of culture on cognition. Specifically, one can safely assume that there is in general a cultural trend to follow others in Asia whereas that to self-assert in Western cultures. Then, a testable question is, if a model is trained in a similar environment as Asian cultures (e.g., people follow others), then would neural representations of close people be more distinct from those of strangers?

These different cultural trends were incorporated into the training regime in the following way. First, nine ingroup and outgroup exemplars were created, respectively (Table 2). Each exemplar was presented with one of the five behavior units ‘on’ (i.e., in total 18*5 = 90 training patterns). A key manipulation involved the temporal order of the to-be-‘on’ behavior unit. In the Asian cultural trend model, a *context* to follow others was implemented as a *temporal constraint in the sequence* of the training set. Specifically, if one exemplar was presented with Behavior 1 ‘on’, then following 17 exemplars were presented with the same Behavior unit ‘on’. Then, another behavior unit was ‘on’ for the 19th exemplar, and the following 17 (20th-36th) exemplars appeared with the same behavior unit ‘on’. In contrast, such a temporal constraint was not made for the Western cultural model. In summary, in Simulation 4, the manipulation was not made in the training patterns themselves but in the sequence of the training patterns. Sixteen input and output layers were connected via nine units in the hidden layer. In order for the temporal sequence effect to come out, the activities in the hidden and output layers were feedback to the hidden layer in the next trial through the (self-) recurrent connections. During the test, the hidden layer activations for the 18 exemplars were measured with all the Behavior units ‘off’.

As a consequence, the mean Euclid distance between the ingroup-outgroup exemplars was larger for the Asian cultural model, mean = 1.813 (*SE* = 0.011) than for the Western cultural model, mean = 1.757 (*SE* = 0.009), $t_{(198)} = 3.728$, $p < .001$. The cluster distance between the groups was 17.16 (*SE* = 0.40) for the Asian cultural model, and was 15.21 (*SE* = 0.25) for the Western cultural trend model, $t_{(198)} = 4.086$, $p < .001$. Therefore, training in a temporal context to follow others increased the representation dissimilarities between close exemplars from others.

Discussion

Since the seminal work of Allport (1954/1979), social psychologists have found the crucial interpersonal experiences that correlate with ingroup bias or outgroup prejudice (Pettigrew & Tropp, 2008). In parallel to these works, various computational models have been implemented to understand the cognitive mechanism to reproduce ingroup bias and other social behaviors (Kashima et al., 2000; Smith & DeCoster, 1998). Our model is clearly a descendant of these models. A key difference, however, was that our model was implemented to explain how neural representations for social cognition emerge from interactions with others. Recently, cognitive neuroscientists have clarified that those with greater ingroup bias show more distinct neural representations for ingroup faces from outgroup faces (Brosch et al., 2013; Gilbert et al., 2012). We tested if this neuroscience finding was relevant to social psychological findings. Specifically, the training regime of the computational model reflected the past interpersonal experiences that social psychologists have found to correlate with ingroup bias. From Simulations 1 to 3, we demonstrated that these manipulations actually had an effect to reduce the representation dissimilarities between groups. Thus in the present study, we succeeded in integrating findings from different disciplines within a single framework, and therefore demonstrated a plausibility of ingroup bias being learned, which is difficult to demonstrate by a cross-sectional or a retrospective survey.

A meta-analysis of the experimental and survey data (Pettigrew & Tropp, 2008) also suggested the mediators (e.g., increased knowledge) by which contact reduces outgroup prejudice. Our proposed model also demonstrated that increased knowledge about outgroup members (Simulation 2) reduces representation dissimilarities. Thus, a modelling is a promising approach to demonstrate the mechanism by which past interpersonal experiences (according to psychological theory) affect social cognition. Indeed, we demonstrated the generalizability of this approach (Simulation 4) by showing the effect of a temporal context to follow others on the neural dissimilarities between close others and strangers (Meyer et al., 2013).

In addition, our model was never instructed to acquire or reduce ingroup bias itself in any way. Moreover, any component (e.g., unit and layer) of our model was not specialised for ingroup bias itself. Rather, the model had only a mechanism to recognize an input vector that represented an ingroup or outgroup exemplar, a task that humans do in daily lives. Nevertheless, the internal representations that the model acquired for the task captured the nature of the neural representations for ingroup bias. Thus, one possibility is that ingroup bias (and other social cognitions we expect) is an emergent property of interactions with others. In other words, we have demonstrated the plausibility of ingroup bias being learned, without hardwiring a distinct mechanism tailored for the sake of group bias. Related to this, current simulation contributes to the understanding of fMRI data (Cowell & Cottrell, 2013). Our original target was the data from Brosch et al. (2013), which found that the implicit race bias measure was correlated with the classification performance of the neural activities only in the fusiform face area, not other areas. From this pattern, the authors argued the "role of independently identified regions of the face-processing network" for race decoding, rather than a distributed pattern. The current simulation suggests group bias (more specifically, the neural representations for group bias) would not need a distinct, modular mechanism for group bias itself. A future study would be required to incorporate other phenomena relevant to ingroup bias within a single framework, but we hope this would be an initial step to bridge social psychology and social neuroscience by computational modelling.

Acknowledgments

This study was supported by Grant-in-Aid for JSPS (Japan Society for the Promotion of Science) Fellows to T. Ueno.

References

Allport, G. W. (1954/1979). *The nature of prejudice: 25th anniversary edition.*: Perseus Book Publishing, L.L.C.

Brosch, T., Bar-David, E., & Phelps, E. A. (2013). Implicit Race Bias Decreases the Similarity of Neural Representations of Black and White Faces. *Psychological Science, 24*(2), 160-166.

Cowell, R. A., & Cottrell, G. W. (2013). What evidence supports special processing for faces? A cautionary tale for fMRI interpretation. *J Cogn Neurosci, 25*, 1777-1793.

Gilbert, S. J., Swencionis, J. K., & Amodio, D. M. (2012). Evaluative vs. trait representation in intergroup social judgments: distinct roles of anterior temporal lobe and prefrontal cortex. *Neuropsychologia, 50*(14), 3600-3611.

Kanas, A., Sterkens, C., & Scheepers, P. (2013). *Interreligious Contacts and Attitudes towards the Religious Out-group: Testing Contact Hypothesis among Christian and Muslim Students in Indonesia and the Philippines*. Paper presented at the International Society of Political Psychology Annual Meeting, Herzliya, Israel.

Kashima, Y., Woolcock, J., & Kashima, E. S. (2000). Group impressions as dynamic configurations: The tensor product model of group impression formation and change. *Psychological Review, 107*(4), 914-942.

Markus, H. R., & Kitayama, S. (1991). Culture and the self - Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224-253.

Meyer, M. L., Masten, C. L., Ma, Y. N., Wang, C. B., Shi, Z. H., Eisenberger, N. I., & Han, S. H. (2013). Empathy for the social suffering of friends and strangers recruits distinct patterns of brain activation. *Social Cognitive and Affective Neuroscience, 8*(4), 446-454.

Molenberghs, P. (2013). The neuroscience of in-group bias. *Neuroscience and Biobehavioral Reviews, 37*, 1530-1536.

Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology, 38*, 922-934.

Roccas, S., & Brewer, M. (2002). Social identity complexity. *Personality and Social Psychology Review, 6*(2), 88-106.

Schmid, K., Hewstone, M., Tausch, N., Cairns, E., & Hughes, J. (2009). Antecedents and consequences of social identity complexity: intergroup contact, distinctiveness threat, and outgroup attitudes. *Pers Soc Psychol Bull, 35*(8), 1085-1098.

Smith, E. R., & DeCoster, J. (1998). Knowledge acquisition, accessibility, and use in person perception and stereotyping: Simulation with a recurrent connectionist network. *Journal of Personality and Social Psychology, 74*(1), 21-35.

Zhu, Y., Zhang, L., Fan, J., & Han, S. (2007). Neural basis of cultural influence on self-representation. *Neuroimage, 34*(3), 1310-1316.

Table 2. Input & target vector patterns used for training in Simulations 1-4.

Simulation 1 (10 exemplars)	Group index		Traits																			
	Ingroup	Outgroup	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Ingroup exemplar 1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ingroup exemplar 2	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Ingroup exemplar 3	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Ingroup exemplar 4	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Ingroup exemplar 5	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Outgroup exemplar 1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
Outgroup exemplar 2	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
Outgroup exemplar 3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
Outgroup exemplar 4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
Outgroup exemplar 5	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1

Simulation 2 (54 exemplars)	Group index			Traits								
	Ingroup	Outgroup 1	Outgroup 2	Trait A			Trait B			Trait C[1]		
Ingroup exemplar 1	1	0	0	1	0	0	1	0	0	1	0	0
Ingroup exemplar 2	1	0	0	1	0	0	1	0	0	0	1	0
Ingroup exemplar 3	1	0	0	1	0	0	1	0	0	0	0	1
…		…		in total 18 exemplars, formed by crossing 3 * 3* 3 localist patterns								
Ingroup exemplar 18	1	0	0	0	0	1	0	0	1	0	0	1
Outgroup1 exemplar 1	0	1	0	The same localist patterns as 18 ingroup exemplars								
…		…										
Outgroup1 exemplar 18	0	1	0									
Outgroup2 exemplar 1	0	0	1	The same localist patterns as 18 ingroup exemplars								
…		…										
Outgroup2 exemplar 18	0	0	1									

Simulation 3 (10 exemplars)	Group index		Traits																			
	Ingroup	Outgroup	A	B	C	D	E	F[2]	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Low identity complexity ingroup (high overlap)																						
Ingroup exemplar 1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ingroup exemplar 2	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ingroup exemplar 3	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ingroup exemplar 4	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ingroup exemplar 5	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
High identity complexity ingroup (low overlap)																						
Ingroup exemplars 1~5	1	0	Exactly the same as the 5 ingroup exemplars in Simulation 1																			
Outgroup exemplars 1~5	0	1	Exactly the same as the 5 outgroup exemplars in Simulation 1																			

Simulation 4 (18 exemplars * 5 behaviors)	Group index		Traits								Behaviors[3]				
	Ingroup	Outgroup	Trait A		Trait B			Trait C			a	b	c	d	e
Ingroup exemplar 1	1	0	1	0	1	0	0	1	0	0	1	0	0	0	0
Ingroup exemplar 2	1	0	1	0	1	0	0	0	1	0	1	0	0	0	0
Ingroup exemplar 3	1	0	1	0	1	0	0	0	0	1	1	0	0	0	0
…	…		in total 9 exemplars, formed by crossing 2 * 3* 3 localist patterns								Every pattern can take one of the 5 behaviors				
Ingroup exemplar 9	1	0	0	1	0	0	1	0	0	1	1	0	0	0	0
Outgroup1 exemplar 1	0	1	The same localist patterns as 9 ingroup exemplars								Every pattern can take one of the 5 behaviors				
…	…														
Outgroup1 exemplar 9	0	1													

Notes . 1. These three units (Trait C) in the output layer did not receive a target signal (error derivertive was zero) in a superficial contact condition.

2. Trait F represents the highly overlapping category in the low complexity ingroup.

3. In the Asian cultural trend model, one of the five behavior units was randomly selected' per 18 trials (18 succesive exemplars) whereas in the Western cultural trend model, one unit was randomly selected per every trial/exemplar.