Freeman et al. Supplementary Material
DOI: 10.1177/0956797615627418

Supplementary Material

Model and Simulation Method

To illustrate why more unstable White–Black category activation dynamics are predicted for low-exposure participants when they confront racial ambiguity, we conducted a simulation using a new instantiation of the Dynamic Interactive (DI) Model of social categorization (Freeman & Ambady, 2011). The model has a recurrent connectionist architecture with stochastic interactive activation (McClelland, 1991; Rumelhart, Hinton, & McClelland, 1986). Depicted in Fig. S1, this new instantiation of the model provides an approximation of the kind of processing that might take place in a human brain (Rogers & McClelland, 2004; Rumelhart et al., 1986; Smolensky, 1989; Spivey, 2007), specifically in the context of perceiving a face's race and depending on one's interracial exposure.

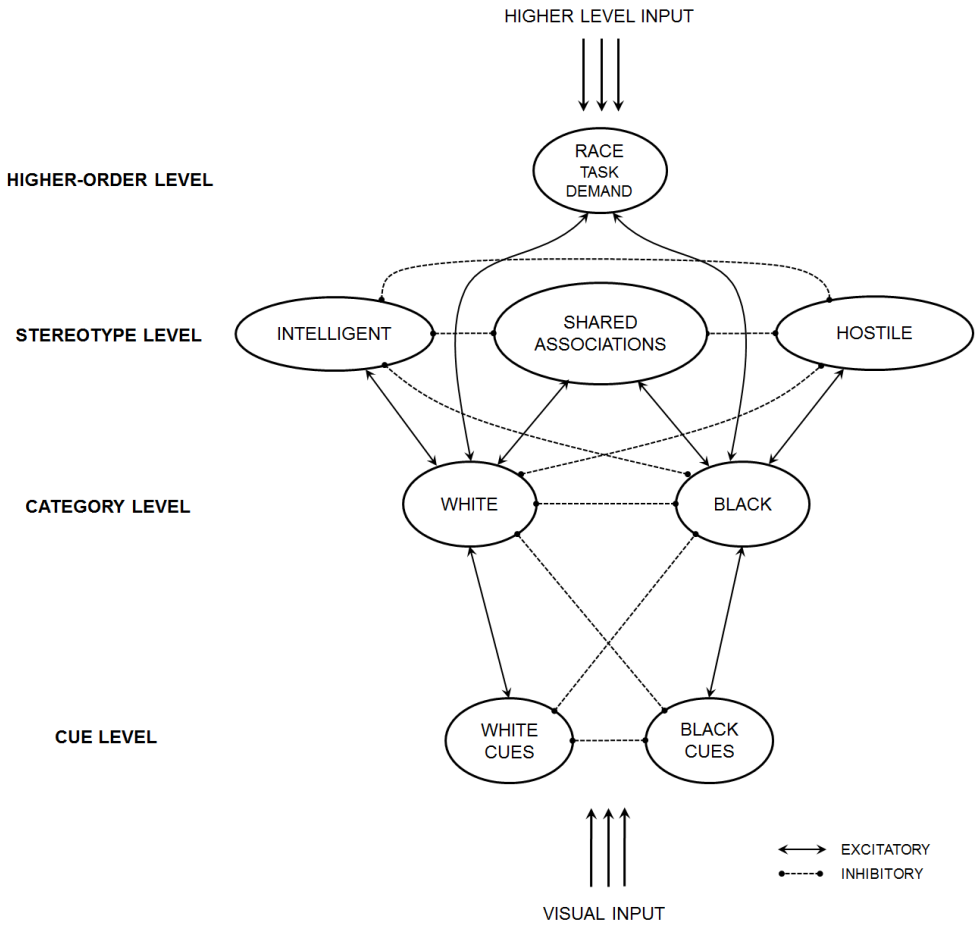

Figure S1. A new instantiation of the Dynamic Interative (DI) Model of social categorization.

DOI: 10.1177/0956797615627418

Initially, the network is stimulated simultaneously by both visual input and higher-level input. Visual input originates from the visual system, which processing an incoming face stimulus, and higher-level input in this case originates from a top-down attentional system, which directs attention toward particular categories based on the current task (to categorize race). The network contains a collection of nodes, and each node is associated with a transient level of activation at each moment in time. This activation corresponds with the strength of a tentative hypothesis that the node is represented in the input. The nodes in this network are organized into four levels of processing: cue nodes (visual detectors for facial features), category nodes (social category representations), stereotype nodes (stereotype attributes and social-conceptual knowledge), and higher-order nodes (task demands). Once the network is initially stimulated, activation flows among all nodes as a function of their connection weights. Because connections between nodes are bidirectional, this flow results in a dynamic back-and-forth of activation between all nodes in the system. As such, nodes in the system continually readjust each other's activation and mutually constrain one another to find an overall pattern of activation that best fits the given inputs. Gradually, the flows of activation lead the network to converge on a stable, steady state, where the activation of each node reaches an asymptote (i.e., an attractor; see Fig. 1 of the main text). This final steady state corresponds to an ultimate race categorization.

A critical aspect of the model is that, as facial cues automatically activate categories, those categories in turn automatically activate knowledge structures (stereotypes). Those knowledge structures in the stereotype level, in turn, provide an immediate constraint on category activation, providing top-down feedback to the category level (Freeman & Ambady, 2011; Freeman et al., 2011). Thus, conceptual knowledge and stereotypes are spontaneously activated in a manner that, in turn, shapes the race categorization process. Accordingly, there are two simultaneous forces at play that drive race categorization (category level): bottom-up visual processing (from the cue level) and top-down conceptual processes (from the stereotype level).

In response to a mixed-race face, the network seeks to settle into the stable, lower-energy attractors (White or Black) due to learned conceptual knowledge. At the same time, dynamic visual processing of mixed-race cues creates bottom-up pressure that works against the natural descent of the system into the attractors (pushing White and Black category activations closer together and driving the system back up toward the ridge between the two attractors). Because we predict the White and Black attractors to be more differentiated in low-exposure perceivers

(deeper and farther apart, creating a steeper descent), the system will experience a stronger "pull" to settle down into the White and Black categories, while visual processing of mixed-race cues "pushes" against the attraction, causing more unstable dynamics for the low-exposure network. In other words, a mixed-race face creates instability for the perceptual system of a low-exposure perceiver because bottom-up visual processing attempts to bring together two race categories that top-down conceptual knowledge is trying to strongly pull apart. The force of this "pulling apart" due to conceptual knowledge is stronger in low-exposure perceivers, as the category representations are more distinct. We predict that this bottom-up "pushing" of race category activations together (due to racial ambiguity) and top-down "pulling" of race categories apart (forcing a race categorization in line with learned conceptual knowledge) will create uniquely unstable dynamics in low-exposure perceivers. Thus, a more unstable experience is expected when the low-exposure neural network encounters racial ambiguity, even though there need not be any increase in the overall amount of competition.

Structure of the DI Model

In the DI Model, before the presentation of a face stimulus, activations of all nodes in the network are set equal to a resting activation value (zero), and external inputs are presented to certain nodes for processing. Processing occurs over a number of iterations. On each iteration, each node computes its net input from the nodes connected to it based on their latest activation (excitation and inhibition are summed together), as well as any external input into the node. Because the model is stochastic, the input is also altered by normally distributed noise. Specifically, the net input to node *i* is:

$$net_i = \sum_j w_{ij} o_j + ext_i + \varepsilon_\sigma$$

where w_{ij} is the connection weight to node *i* from node *j*, o_j is the greater of 0 and the activation of node *j*, ext_i is any external input to node *i*, and ε_σ is a small amount of normally distributed random noise with mean 0 and standard deviation σ. Once the net input into all nodes has been computed, the activation of node *i* is updated as:

DOI: 10.1177/0956797615627418

$$\text{If } net_i > 0:$$
$$\Delta a_i = I(M - a_i)net_i - D(a_i - r)$$
$$\text{If } net_i \leq 0:$$
$$\Delta a_i = I(a_i - m)net_i - D(a_i - r)$$

such that *M* is the maximum activation, *m* is the minimum activation, *r* is the resting activation level, *I* is a constant that scales the influence of external inputs on a node, and *D* is a constant that scales a node's tendency to decay back to rest. The parameters are as follows: $M = 1$, $m = -0.2$, $r = 0$, $I = 0.4$, $D = 0.1$, and $\sigma = 0.01$. Network parameters, connection weights, and input values were set based on our prior studies, intuitions regarding stimulus and task features, and previous simulations. See Freeman and Ambady (2011) for complete details on the DI Model.

Modeling interracial exposure

We developed two variants of the model instantiation (low-exposure and high-exposure networks); connection weights for both networks are provided in Table S1. What distinguishes the low-exposure and high-exposure network is the connections between race categories and conceptual knowledge (stereotypes), i.e., category–stereotype connections. Specifically, in the high-exposure network, the WHITE and BLACK categories were modeled as conceptually more similar, as high exposure leads such categories to be represented with greater similarity and overlapping characteristics (Allport, 1954; Dovidio, Gaertner, & Kawakami, 2003). To model this overlap, we included three stereotype nodes: one to represent exclusively White-related stereotype attributes (e.g., INTELLIGENT), having an excitatory connection with WHITE and inhibitory connection with BLACK; one to represent exclusively Black-related attributes (e.g., HOSTILE), having an excitatory connection with BLACK and inhibitory connection with WHITE; and critically, one to represent overlapping attributes (SHARED ASSOCIATIONS), having mutual excitatory connections with both race categories (see Fig. S1). The low-exposure and high-exposure network differ only in that the presence of the SHARED ASSOCIATIONS node was stronger in the high-exposure network, such that its connections were stronger (i.e., stronger excitatory connections with race categories, and stronger inhibitory connections with non-overlapping stereotypes; see Table S1). This represents the effects of high interracial exposure, with stronger conceptual overlap and shared attributes between the White and Black categories.

DOI: 10.1177/0956797615627418

Table S1. Connection weights for the low-exposure and high-exposure networks.
Note: All connections are bidirectional and symmetrical. *Shared* stereotypes refer to the SHARED ASSOCIATIONS node (positively related to both WHITE and BLACK), and *non-shared* stereotypes refer to the INTELLIGENT node (positively related to WHITE and negatively related to BLACK) and the HOSTILE node (positively related to BLACK and negatively related to WHITE).

Connection	*Low-exposure*	*High-exposure*
Category–Stereotype (*shared*) excitation	.4	.6
Stereotype (*shared*)–Stereotype (*non-shared*) inhibition	–.2	–.4
Category–Stereotype (*non-shared*) excitation	.8	.8
Category–Stereotype (*non-shared*) inhibition	–.4	–.4
Stereotype (*non-shared*)–Stereotype (*non-shared*) inhibition	–.4	–.4
Cue–Cue inhibition	–.1	–.1
Cue–Category excitation	.25	.25
Cue–Category inhibition	–.25	–.25
Higher-order–Category excitation	.8	.8
Category–Category inhibition	–.7	–.7

As theoretically predicted (see Fig. 1 of the main text), a stronger presence (high-exposure) vs. weaker presence (low-exposure) of shared associations influenced the shape of the attractors to which the networks gravitated, i.e., the White and Black category representations. To visualize the attractors, we estimated energy landscapes associated with the two networks. Specifically, at a given state of the network, the network's energy was calculated as:

$$E = -\frac{1}{2}\sum_{i}\sum_{j} w_{ij} a_i a_j$$

where w_{ij} is the connection weight to node i from node j, and a_i and a_j are the respective node's activation. Energy landscapes associated with the two networks are depicted in Fig. S2, showing that the White and Black category attractors (indicated by energy minima) were more distinguished, including farther apart and deeper, in the low-exposure (distance: 0.76, relative depth: 0.24) than high-exposure (distance: 0.64, relative depth: 0.17) network, consistent with our predictions.

DOI: 10.1177/0956797615627418

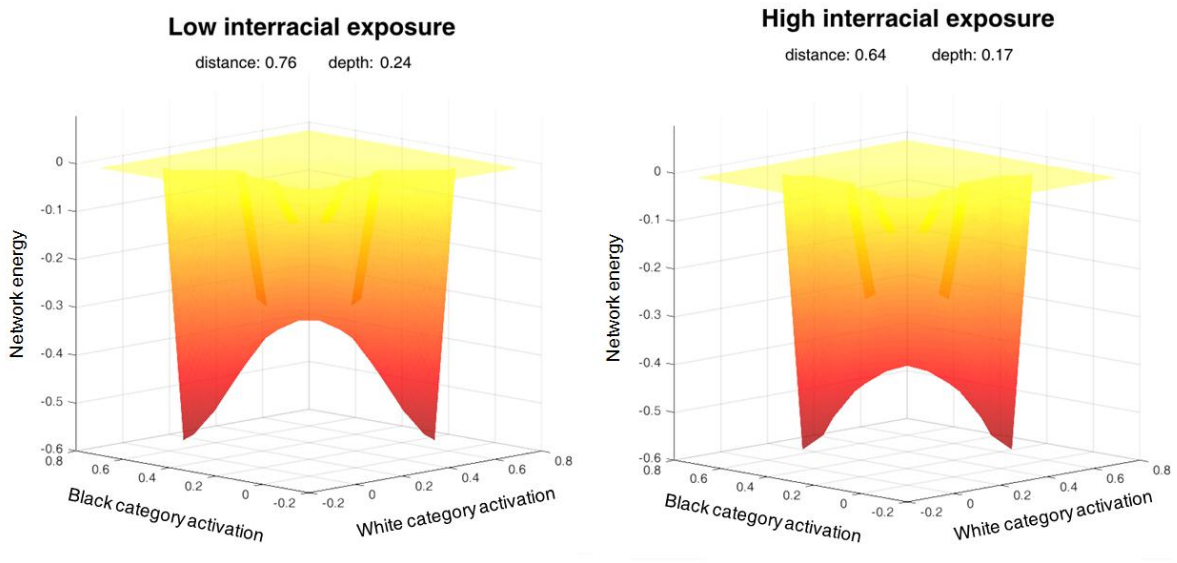

Figure S2. Energy landscapes associated with the low-exposure and high-exposure networks, where the network's energy is plotted as a function of the WHITE and BLACK nodes' activation. The White and Black attractors (energy minima) are more distinct, with increased distance and depth, in the low-exposure network.

Simulations

For both the low-exposure and high-exposure network, we conducted 3 simulation types (White face, ambiguous face, Black face) and ran each type 1,000 times. Thus, we ran a total of 6,000 simulations (akin to 6,000 human trials). For all simulation types, input into the RACE TASK DEMAND node was set at .6, simulating the task demand that requires attention on race. For the White face condition, we set input into the WHITE CUES node at 1 and into the BLACK CUES node at 0, and vice-versa for the Black face condition. For the ambiguous face condition, we set input into both nodes at .5. This rendered visual input equibiased with respect to a face's race in the ambiguous condition, thereby causing both categories to co-activate early in the categorization process until small amounts of noise in the system eventually led one category to win out and the other to decay. After 300 iterations, we selected the race-category node with the highest activation as the network's categorization response.

For each simulation trial, we estimated MD as the unselected category node's maximum activation level across all 300 iterations, divided by the maximum activation level of the selected category node's activation. As such, this reflected the extent to which the unselected category

was partially activated in parallel, thereby providing an index of overall competition. For each simulation trial, we provided an estimate of *x*-flips using the same equation described in the main text, except measuring changes in the relative direction between the WHITE and BLACK category nodes' activations rather than *x*-coordinates from mouse-tracking. For example, if from $t - 1$ to t the WHITE category is gaining activation and the BLACK category is losing activation, but from $t - 2$ to $t - 1$ the WHITE category was losing activation and the BLACK category was gaining activation, this would increase the *x*-flip count by 1. Specifically, if the difference in WHITE and BLACK category nodes' activation at time t is $d_t = w_t - b_t$, then:

$$x\text{-flips} = \sum H[-(d_t - d_{t-1})(d_{t-1} - d_{t-2})]$$

As with the actual *x*-flips measure in mouse-tracking, we first smoothed the two activation time series using a sliding average in order to detect larger-scale directional changes in category activation rather than small, random perturbations. For the actual *x*-flips mouse-tracking measure, we had applied 5% smoothing (a sliding average across a window of 5 time bins in a total time series of 100 time bins). Accordingly, for our simulations we applied 5% smoothing as well (a sliding average across a window of 15 iterations in a total time series of 300 iterations).

Simulation Results

For analyses, we aggregated simulations of White and Black faces to permit a comparison between typical and ambiguous faces. To approximate the *x*-flips counts observed with human participants, each trial's *x*-flips count was divided by a constant of 6.5. Predictably, a network exposure (low-exposure, high-exposure) × racial ambiguity (ambiguous, typical) ANOVA on MD indicated a large effect of ambiguity, $F(1, 5996) = 85372.24$, $p < .0001$, with higher MD for ambiguous trials. Specifically, the MD ambiguity effect in the low-exposure network [$M = 0.697$, $SE = 0.004$; $t(2998) = 199.37$, $p < .0001$] was highly significant, as was the MD ambiguity effect in the high-exposure network [$M = 0.692$, $SE = 0.003$; $t(2998) = 214.82$, $p < .0001$]. Critically, however, these effects did not differ across low- and high-exposure networks, as the racial ambiguity × exposure interaction was not significant, $F(1, 5996) = 0.98$, $p = .321$. A network exposure (low-exposure, high-exposure) × racial ambiguity (ambiguous, typical) ANOVA on *x*-flips also indicated a large effect of racial ambiguity, $F(1, 5996) = 1794.98$, $p < .0001$, but more importantly, a significant racial ambiguity × exposure interaction,

$F(1, 5996) = 8.977$, $p = .003$. This was because the *x*-flips ambiguity effect (higher *x*-flips for ambiguous relative to typical trials) in the low-exposure network [$M = 0.893$, $SE = 0.027$; $t(2998) = 32.62$, $p < .0001$] was considerably stronger than in the high-exposure network [$M = 0.775$, $SE = 0.028$; $t(2998) = 27.40$, $p < .0001$]. These results therefore converge with the data from the mouse-tracking tasks with human participants of Studies 1 and 2 (described in the main text), where low-exposure relative to high-exposure participants exhibited more abrupt category shifts (*x*-flips) but not overall category competition (MD) for racially-ambiguous faces.

References

Allport, G. W. (1954). *The nature of prejudice*. Oxford: Addison-Wesley.

Dovidio, J. F., Gaertner, S. L., & Kawakami, K. (2003). Intergroup contact: The past, present, and the future. *Group Processes & Intergroup Relations, 6*(1), 5-21.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review, 118*, 247-279.

Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: Social status cues shape race perception. *PLoS ONE, 6*, e25107.

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology, 23*, 1-44.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach*. Boston: Bradford Books.

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). *A general framework for parallel distributed processing*. Cambridge, MA: MIT Press.

Smolensky, P. (1989). Connectionist modeling: Neural computation/mental connections. In L. Nadel, A. Cooper, P. Culicover & R. M. Harnish (Eds.), *Neural connections, mental computations*. Cambridge, MA: MIT Press.

Spivey, M. J. (2007). *The continuity of mind*. New York: Oxford University Press.