Social Cognition, Vol. 36, No. 2, 2018, pp. 221–246

SOCIAL CATEGORIZATION IN CONNECTIONIST MODELS: A CONCEPTUAL INTEGRATION

André Klapper
Radboud University, Donders Institute

Ron Dotsch
Utrecht University; Radboud University, Behavioral Science Institute

Iris van Rooij
Radboud University, Donders Institute for Brain, Cognition, and Behavior

Daniël H.J. Wigboldus
Radboud University, Behavioral Science Institute

We present a conceptual integration of two major types of social perception models. First, according to social categorization models, perceivers can employ two processes: they either treat other people as individuals (individuation) or as members of social groups (social categorization). Second, according to connectionist models, person perception is driven by a single process of spreading activation between mental representations in a learned associative network. We suggest that social categorization and individuation can be conceptualized as different types of inputs to a single (connectionist) process. Furthermore, we implement this idea in computer simulations and show that it can account for an empirical dissociation between social categorization and individuation despite being a single process model. Overall, this work aims to contribute to the coherence and integration of the theoretical and empirical literature on social cognition.

Keywords: social categorization, connectionism, person perception, memory

This research was supported by an NWO grant 464-11-036 awarded to Ron Dotsch and Daniël Wigboldus.

Correspondence concerning this article should be addressed to André Klapper, Donders Institute for Brain, Cognition, and Behavior, Radboud University, Postbus 9101, 6500 HB Nijmegen, Netherlands; E-mail: a.klapper@donders.ru.nl.

When people process information about other people, they do not always treat them as individuals (individuation) but frequently treat them as members of social groups (social categorization; Allport, 1954; Hugenberg, Young, Bernstein, & Sacco, 2010; Macrae & Bodenhausen, 2000, 2001; Tajfel, 1969). Treating other people as group members may be a useful cognitive strategy to reduce the information load faced by social perceivers (Macrae & Bodenhausen, 2000, 2001; Quinn & Macrae, 2005). However, social categorization may also lead to systematic errors such as confusions between people within social groups (Klauer & Wegener, 1998; Taylor, Fiske, Etcoff, & Ruderman, 1978). These are key insights from *social categorization models,* which have been influential in the literature (Brewer, 1988; Fiske & Neuberg, 1990; Hugenberg et al., 2010; Macrae & Bodenhausen, 2000, 2001).

An important source of support for social categorization models comes from person memory research. It is a highly robust and widespread phenomenon that people tend to confuse other people more often within groups (e.g., men with other men) than people between groups (e.g., men with women; Blanz, 1999; Gawronski, Ehrenberg, Banse, Zukova, & Klauer, 2003; Klauer, Hölzenbein, Calanchini, & Sherman, 2014; Klauer & Wegener, 1998; Taylor, Fiske, Etcoff, & Ruderman, 1978) during memory retrieval. This effect gets larger to the extent that the social group is salient, which suggests that the effect is not purely driven by properties of the stimuli but also by the cognitive strategies through which people encode other people, such as categorization and individuation (Blanz, 1999; Klauer & Wegener, 1998; Van Twuyver & van Knippenberg, 1995). In addition, multinomial processing tree (MPT) modeling has been applied to these findings (Klauer & Wegener, 1998; see also Gawronski et al., 2003; Klauer et al., 2014). In line with the core notions of social categorization models, the results showed evidence of two independent cognitive components: one where speakers are treated as equivalent group members (social categorization) and one where speakers are treated as different individuals (individuation).

In addition to the findings above, it was shown that people are more likely to falsely indicate that they have seen a person before (i.e., a false positive in recognition) if that person is from another race than the perceiver (Hugenberg et al., 2010). This has aroused considerable debate given that such false positives can have serious consequences in witness testimony, causing an innocent person to be sentenced for a crime that the person did not commit (for a review see Hugenberg et al., 2010). Importantly, this bias toward false positives for other-race people is reduced (or even eliminated) when people are motivated to individuate (Hugenberg, Miller, & Claypool, 2007; Hugenberg et al., 2010; Young & Hugenberg, 2011). This finding suggests that memory confusions (and their potentially devastating consequences) may in part be caused by reliance on social categorization rather than individuation.

Roughly at the same time, connectionist models emerged, which do not make an explicit distinction between social categorization and individuation (Ehret, Monroe, & Read, 2014; Freeman & Ambady, 2011; Kunda & Thagard, 1996; Smith & DeCoster, 1998; Van Overwalle & Labiouse, 2004; Van Rooy, Van Overwalle, Vanhoomissen, Labiouse, & French, 2003). Instead, these models assume that person

perception is driven by interactions between mental representations that are all subject to the same processing rules. For this reason, connectionist models have often been seen as single process models (Ehret et al., 2014; Kunda & Thagard, 1996; Thagard & Verbeurgt, 1998). More specifically, in connectionist models each mental representation can be activated by an observed stimulus and this activation spreads via learned associative links to other representations. Connectionist models have been successful in explaining a multitude of social phenomena (e.g., assimilation effects, contrast effects, illusory correlations, polarization effects, etc.) and their general notions (e.g., that activation spreads via associations between mental representations) have become ubiquitous in the general social cognition literature (Dalege et al., 2016; Freeman & Ambady, 2011; Kunda & Thagard, 1996; Smith & DeCoster, 1998; Van Overwalle & Labiouse, 2004; Van Rooy et al., 2003; Zebrowitz, Fellous, Mignault, & Adreoletti, 2003). In addition, connectionist models have successfully been used to unify various dual process notions into a single model such different language processes (Seidenberg & Plaut, 2014; Smith, 2009), bottom-up and top-down processes in social perception (Freeman & Ambady, 2011), and learning and use of exemplars and prototypes (McClelland, McNaughton, & O'Reilly, 1995; Smith & DeCoster, 1998, 2000).

However, while both social categorization and connectionist models have been used to potentially explain a multitude of empirical findings, the conceptual relationship between social categorization and connectionist models has remained relatively unclear. That is, while some researchers have adopted the viewpoint that connectionist models are a competing alternative to social categorization models (e.g., Cox & Devine, 2015; Kunda & Thagard, 1996), others adopted the viewpoint that these models may be compatible (e.g., Freeman & Ambady, 2011). Moreover, if they are compatible, an account is lacking of how these two types of models can be reconciled (although precursors exist: McClelland, McNaughton, & O'Reilly, 1995; Smith & DeCoster, 1998, 2000). In particular, it remains relatively unclear how the distinction between social categorization and individuation can be reconciled with the idea that all social information is processed by one underlying connectionist mechanism. In addition, it remains unclear how (single process) connectionist models fit to evidence of a cognitive dissociation between two seemingly independent cognitive components (Klauer & Wegener, 1998).

THE PRESENT ARTICLE

The present article aims to contribute to the coherence and integration of the literature by resolving seeming conceptual contradictions. Specifically, while it is clear that the notion of social categorization is consistent with evidence of a cognitive dissociation, two other conceptual relationships remain unclear (see Figure 1). First, it remains unclear how (dual process) social categorization models can be synthesized with (single process) connectionist models. This raises the question: can they both be true? If social categorization and connectionist models are incompatible, either social categorization or connectionist models must be false

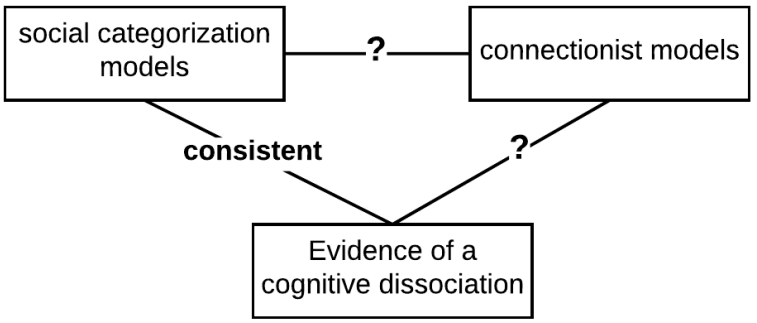

FIGURE 1. An illustration of the intended contribution of our framework. While it is clear that (dual process) social categorization models are consistent with evidence of a cognitive dissociation, it is unclear (1) how they fit to (single process) connectionist models, and (2) how (single process) connectionist models fit to evidence of a cognitive dissociation. Our framework aims to provide a coherent integration of these three parts of the literature.

regardless of the empirical evidence and simply based on the logical constraint that: if one type of model is true, the other must logically be false. As such, clarifying whether and how they are compatible is essential for the progression of these models. For this purpose, we aim to provide a framework, which shows that and explains how the categorization-individuation distinction of social categorization models can be integrated into connectionist models. The key idea of our framework is that categorization and individuation may be seen as two different types of inputs to a (single process) connectionist model. This is worked out to the level of a formal model to give a proof of concept for the general idea.

The second conceptual relationship that has remained unclear is the relationship between (single process) connectionist models and the finding of a cognitive dissociation between individuation and social categorization in person memory. This raises the question: how can a single process model be true given empirical findings that seem to suggest dual processes? This is particularly important for the framework and formal model we aim to provide, as they will provide a single process interpretation of social categorization models. We will argue that situating the categorization-individuation distinction at the input level rather than process level is consistent with existing evidence of a cognitive dissociation (Klauer & Wegener, 1998) and support this argument through computer simulations. Specifically, we will show in computer simulations that we can reproduce the cognitive dissociations purely by varying the inputs to a single connectionist process.

Overall, our work aims to (1) demonstrate the compatibility of social categorization and connectionist model and (2) demonstrate that (and how) a single process connectionist model can account for existing evidence of a cognitive dissociation. In the following sections, we will introduce our framework, present a more specific formal implementation of the framework, and finally demonstrate through computer simulations that a single connectionist model is consistent with the finding of a cognitive dissociation.

THE FRAMEWORK

Marr (1982) proposed an influential distinction between three levels at which a cognitive mechanism can be described. At the computational level, one describes the *input-output mapping* that the mechanism performs (i.e., *what* the mechanism does). At the algorithmic (or process) level, one describes the *processing steps* by which the input is transformed into the output (i.e., *how* the mechanism does it). Finally, at the implementational level one describes the physical implementation of the mechanism. The history of (social) cognition research has shown that distinguishing between these levels is crucial when comparing cognitive models because sometimes seemingly antagonistic models can turn out to be descriptions of the same theoretical mechanism at different levels (De Houwer & Moors, 2015). Here, we employ Marr's distinction between computational and algorithmic level to resolve seeming inconsistencies in the person perception literature. Interested readers can find more information on Marr's levels in Appendix A.

Consider a simplified analogy to illustrate how our framework synthesizes the literature. A coffee machine may take a coffee capsule, some water, and a cup as input and returns a cup of coffee as output (computational level). The process by which this input-output mapping is achieved is to run the water through the coffee capsule and into the cup (algorithmic level). First, notice that although the coffee machine utilizes a single process (running the water through the coffee capsule and into the cup), one can nevertheless make distinctions between different types of inputs (e.g., normal coffee cups and caramel coffee cups). As such, the coffee machine can be seen simultaneously as a "dual" model at the input (or computational) level but as a "single" model at the process (or algorithmic) level. In other words, the model can be both "dual" and "single" as long as these labels refer to different levels of the machine. Second, notice that the coffee machine can also produce dissociable outputs: if different types of coffee capsules are employed (inputs) the coffee machine will produce dissociable types of coffee (outputs). Therefore, a researcher who applies cognitive dissociation analyses to the output may arrive at the conclusion that there is a dissociation between two underlying cognitive components. Importantly, notice that such a dissociation based on outputs does not necessarily reflect a dissociation between processes (algorithmic level) but may instead reflect a dissociation between different types of inputs (computational level).

Our framework follows the same logic. We propose that the distinction between social categorization and individuation can be conceptualized as a distinction between different types of connectionist inputs, making it consistent with a single process connectionist model. In a connectionist model, each node receives an *external input*, which reflects the degree to which the node is excited by currently observed stimuli. We propose (see Figure 2) that some nodes can receive positive external input from any member of a social group (category nodes) while other nodes can receive positive external input only from specific individuals (individual nodes). For example, we may call the node *man* a category node, and excitation of this node "social categorization," because the node *man* can receive positive

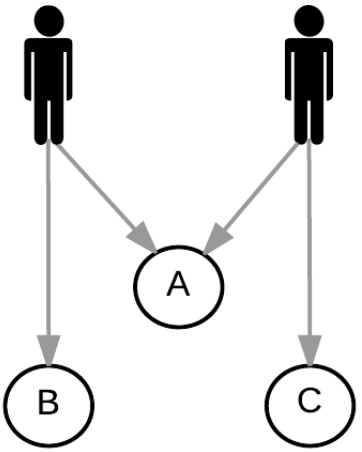

FIGURE 2. An illustration of our distinction between social categorization and individuation in the input of a connectionist model. While node *A* can be excited by the observation of any member of a social group (social categorization), nodes *B* and *C* can be excited exclusively by the observation of specific individuals (individuation).

external input from several observed people (any man). Conversely, we may call the node *Brad* an individual node, and exciting this node "individuation" because the node *Brad* can receive positive external input exclusively from the perception of a specific individual (i.e., Brad).

Importantly, while the external inputs to these types of nodes can be different, the processes that operate on them can be the same. In connectionist models all nodes are subject to the processing rules to increase activation based on external inputs, spread activation via associative links, and decay activation over time (Freeman & Ambady, 2011; McClelland & Rumelhart, 1989). Given that all nodes employ those same processing rules, connectionist models can be seen as single process models (Ehret et al., 2014; Kunda & Thagard, 1996; Thagard & Verbeurgt, 1998). Our framework is consistent with these assumptions because it situates the distinction between categorization and individuation in the input (computational level) without affecting the single process assumptions (algorithmic level) of connectionist models. As such, our framework is consistent with the key notions of both connectionist and social categorization models. This reasoning suggests that the answer to our first question is yes: social categorization and connectionist models can be true at the same time (see Figure 1).

However, would such a single process unification be compatible with evidence of a cognitive dissociation (Klauer & Wegener, 1998)? We suggest that the answer is yes. Although there is only a single process operating in connectionist models, we suggest that a cognitive dissociation can nevertheless occur because the process is applied to two different types of inputs (external inputs to category and individual nodes). In the following, we will describe a more detailed implementation of the framework. Subsequently, we will demonstrate using computer simulations that the framework can account for empirical evidence of two dissociable cognitive components in person memory.

THE FORMAL IMPLEMENTATION OF THE FRAMEWORK

We will proceed by describing an implementation of the framework in terms of computer simulations. This serves several purposes. First, computer simulations force us to fill in more details of the framework, thereby ensuring that our integration of social categorization and connectionist assumptions is internally consistent. Hence, the formal implementation in our computer simulations can be seen as a proof of concept for the more general ideas presented in our framework. Second, computer simulations also enable us to test whether the single process unification is compatible with existing evidence of a cognitive dissociation.

Connectionist models of person perception are usually decomposed into two sub-mechanisms: *learning* and *perception*. Importantly, these are not two "dual" mechanisms in the sense that the cognitive system must select between them. Instead, they constitute different sub-mechanisms of person perception that serve different sub-functions. Namely, while *learning* serves the purpose to generate stored knowledge, *perception* utilizes this knowledge to make inferences from observed stimuli.

In the *learning* mechanism, the cognitive system generates associative links between nodes based on observation. A commonly assumed input-output mapping (computational level) is that the learning mechanism takes as input the degree to which nodes (e.g., *beard*, *professor*, *man*) are currently observed as present (positive external input) or absent (negative external input), and returns as output associative links that reflect the correlations between these external inputs to nodes (McClelland & Rumelhart, 1989; Thagard & Verbeurgt, 1998). A possible *process* (algorithmic level) by which this input-output mapping may be achieved is to strengthen associations at moments where two nodes are observed as present (or absent) and weaken associations at moments where one node is observed as present and the other as absent. This is also known as Hebbian learning (McClelland & Rumelhart, 1989) and will be the learning mechanism we employ in the computer simulations in the subsequent section (for formal details see Appendix B). It is worth noting that this is one of several learning mechanisms that have been applied in the literature (Ehret et al., 2014; McClelland & Rumelhart, 1989; Van Overwalle & Labiouse, 2004; Van Rooy et al., 2003). Importantly, the goal here is not to argue for Hebbian learning but to provide a proof of concept for the general framework: that is, we aim to show that our (single process) synthesis of connectionist and social categorization models can account for evidence of a cognitive dissociation (using some arbitrary learning mechanism).

In the *perception* mechanism, an activation pattern that reflects a perception output (e.g., a memory retrieval result) of another person is generated based on observation (external input) and internal knowledge (associative links). An overview is given in Figure 3. The *input* (first part of the computational level) of the *perception* mechanism refers to the starting state of the network, which consists of a set of nodes with starting activations (usually zero), the degree to which each node is observed as present or absent in the perceived world (external inputs), and a set of weighted associative links between the nodes (derived from the learn-

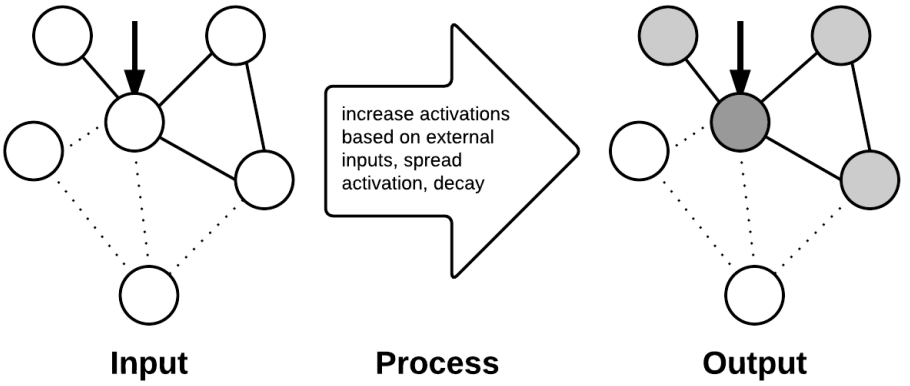

FIGURE 3. An illustration of the distinction between input, process, and output in the *perception* part of a connectionist model. The *input* refers to the starting state of the network, which consists of a set of nodes with initial activation levels (usually zero), external inputs that indicate the degree to which each node is excited by an observed stimulus (in the figure above, only the nodes with an arrow are excited by the currently observed stimulus), and a set of excitatory (solid lines) and inhibitory (dashed lines) associations. The *process* is the set of rules by which the starting state of the network (input) is transformed into its final state (output). Specifically, the process entails to (1) increase the activations of nodes to the degree that they are excited by an observed stimulus (i.e., influences of observation), (2) spread activation between nodes via excitatory and inhibitory links (i.e., influences of knowledge/prior experiences), and (3) gradual activation decay (McClelland & Rumelhart, 1989). The *output* refers to the final activations after all activations have stabilized.

ing mechanism; Thagard & Verbeurgt, 1998). The connectionist *process* (algorithmic level) refers to the set of rules that are used to update the activations of each node (Thagard & Verbeurgt, 1998). These rules entail to (1) increase activations of nodes to the degree that they are excited by an observed stimulus (i.e., influences of observation), (2) spread activation between nodes via excitatory and inhibitory links (i.e., influences of learned knowledge/prior experiences), and (3) gradual activation decay (McClelland & Rumelhart, 1989). This continues iteratively until all activation levels stabilize in an equilibrium. Finally, the *output* (second part of the computational level) refers to the final activations after all activations have stabilized or until the process is interrupted (for formal details see Appendix B).

To develop our framework in sufficient detail to conduct computer simulations, it is necessary to situate the categorization-individuation distinction in the interplay between learning and perception. Recall that we proposed a distinction between nodes that can receive positive external input from any member of a social group (category nodes) and nodes that can receive positive external input exclusively from specific individuals (individual nodes). This distinction has two consequences. First, it affects which associations are formed with category nodes and individual nodes during *learning*. Given that category and individual nodes receive different external inputs, their correlations with external inputs of other nodes will be different. Consequently, their associations with other nodes will be different as well (in spite of being based on the same learning mechanism). Second, the distinction affects the activation pattern that is generated during *perception*. Given that category and individual nodes receive different external inputs and also have dif-

ferent associations with other nodes, they influence the person perception output in distinct ways (in spite of being based on the same perception mechanism). As a result, dissociable patterns may emerge in the person perception output such as the ones reported in past experiments (Gawronski et al., 2003; Klauer et al., 2014; Klauer & Wegener, 1998; see also Young & Hugenberg, 2011).

COMPUTER SIMULATIONS

Some of the most straightforward evidence for a cognitive dissociation comes from person memory research. This research has shown that people have tendency to confuse people within groups with each other during retrieval—especially when the group membership is made salient. This has been shown most extensively using the "Who said what" paradigm, which we will use as our main example here (Gawronski et al., 2003; Klauer, Hölzenbein, Calanchini, & Sherman, 2014; Taylor et al., 1978; for an overview see Klauer & Wegener, 1998). In the learning phase of this paradigm, participants read statements made by several speakers who fall into two different social categories (e.g., male and female). In the test phase, the statements are presented again and participants need to select the correct speaker of the statement.

It is a highly robust finding that participants tend to confuse members within groups (e.g., male speakers with other male speakers) more often than they confuse members between groups (e.g., male speakers with female speakers)—especially when the group membership is made salient (Blanz, 1999; Gawronski et al., 2003; Klapper, Dotsch, van Rooij, & Wigboldus, 2016; Klauer et al., 2014; Klauer & Wegener, 1998; Taylor et al., 1978; Van Twuyver & van Knippenberg, 1995). More importantly, multinomial processing tree (MPT) analyses showed evidence for a dissociation between two underlying cognitive components: one that distinguishes between individual speakers (individuation) and one that distinguishes the speakers at a group level (social categorization; Klauer & Wegener, 1998).

In the following, we aim to demonstrate that our framework can reproduce within group confusions and most importantly: MPT-based evidence of dissociable cognitive components.[1] For this purpose, we conducted computer simulations of the connectionist mechanisms described in the previous section. The key aspect of our account of the person memory findings lies in the idea that the perceiver can learn an associative link between the statement and a specific individual (e.g., Peter) and/or an associative link between the statement and the social category of the speaker (e.g., male). If the perceiver learned exclusively a statement-category link (learning output) then the statement will activate the category, which then activates all speakers that are associated with that category (perception output).

1. It is worth noting that these are not independent results. If the MPT analysis detects evidence of social categorization, this necessitates that there were more within group confusions than between group confusions. However, if there are more within group confusions than between group confusions, this does not necessitate that a cognitive dissociation is found, because the MPT analysis makes additional assumptions that can be false. To successively build up our conclusions, we analyze group confusions first and subsequently focus on the MPT analysis.

This prevents between-category confusions but does not prevent within-category confusions. In contrast, if the statement is directly associated with the correct individual (e.g., Peter) then the statement will activate exclusively the correct individual, causing correct recognition of the speaker of the statement. Importantly, whether the statement will be associated with the individual or merely the social category of this individual depends on the degree to which the individual and category nodes received external *input* during learning. Most importantly, because the effects of external inputs to individual and category nodes are independent of each other, an MPT analysis will show evidence of two dissociable cognitive components if applied to the simulated behavioral data. These ideas were tested in Simulations 1–4, which were implemented in R (R Core Team, 2017). Original scripts are available on Open Science Framework (osf.io/ade2h).

Our simulation of the "Who said what" paradigm required simulating two types of learning. First, we needed to simulate lifetime learning, which leads to a network that is already present before the participant begins with the "Who said what" task. Second, we needed to simulate the learning that takes place during the learning phase of the "Who said what" task. In the following, we describe the simulation of lifetime learning. Learning that takes place during the "Who said what" task was embedded with the simulation of specific test trials and is therefore described in the procedure of Simulations 2–4 below. Unless specified otherwise, we followed the rule to set the external inputs of a node to the value 1 for properties that are currently perceived as present (e.g., the person is perceived as *male*), to the value -0.1 for properties that are not perceived as present (e.g., the person is not perceived as *male*), and to the value 0 for currently unobservable properties (e.g., the node *male* receives no external input because no person is perceived).[2]

SIMULATION 1: LIFETIME LEARNING

To simulate lifetime learning, we initialized a network of eight nodes that denote the identities of the speakers in a "Who said what" paradigm (I_1–I_8) and two category nodes (C_1 and C_2). During their lifetime, people usually learn that some properties (e.g., the name "Peter") are associated with certain social groups (e.g., male). To simulate this, we created eight learning inputs corresponding to eight observed people. In each of these learning inputs, one identity node (e.g., I_3) and one category node (e.g., C_1) had an external input of one (e.g., the person *Peter* who is *male*) while all other nodes had an external input of -0.1 (with added normally distributed noise; $\mu = 0$, $\sigma = 0.1$). At the beginning of the learning simulation, all association weights were set to zero. Next, we updated the weights by applying Hebbian learning (see Appendix B for details) 1000 times to each of the eight learning inputs. An illustration of the average network structure that resulted from this simulation is depicted in Figure 4. A possible interpretation of this simulation is

2. We assumed a larger absolute value for properties that are perceived as present based on research suggesting that absence is harder to detect than presence (Agostinelli, Sherman, Fazio, & Hearst, 1986). As such, it seems likely that absence is perceived with less certainty.

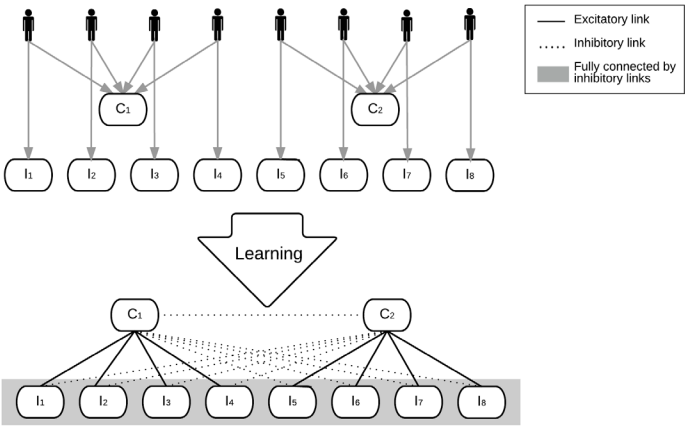

FIGURE 4. The learning input and resulting network in the simulation of lifetime learning. During our simulation of lifetime learning, eight people were observed repeatedly which each excited one identity and one category node respectively (see grey arrows). This led to a network with excitatory category-individual links while category-category and individual-individual links were inhibitory (on average over simulated participants).

that perceivers learn during their life that (1) the names *Peter*, *Carl*, *Jon*, and *Marc* (I_1–I_4) are associated with *male* (C_1), while (2) the names *Jane*, *Maria*, *Lara*, and *Anne* (I_5–I_8) are associated with *female* (C_2). This constitutes the type of knowledge with which a participant enters an experiment. In the following three simulations, we simulated learning and speaker selection during the "Who said what" paradigm.

GENERAL PROCEDURE OF SIMULATIONS 2–4

Our simulations followed an iterative procedure where each iteration consisted of a simulation of learning in a particular learning trial and the selection of the speaker of the statement in the corresponding test trial. In recent applications of the "Who said what" paradigm (Gawronski et al., 2003; Klauer et al., 2014; Klauer & Wegener, 1998) participants were asked two questions in each test trial of the "Who said what" paradigm: (1) was the displayed statement shown during the learning phase, and if yes: (2) who said the statement? The results that we aimed to replicate were based primarily on the responses to test question 2 (Klauer & Wegener, 1998). Therefore, we simulated exclusively the cognitive mechanisms that may underlie answering test question 2 (who said the statement?). Responses to test question 1 were set directly without simulating any cognitive process to ensure that we can apply the same analytical approach as in past research but without complicating the simulations (Klauer & Wegener, 1998).

Specifically, each iteration of our simulation started by setting with equal probability whether the statement considered in the iteration would be treated as a target statement (which was shown during the learning phase) or a distractor (which was exclusively shown during the test phase). Next, we directly set the response to test question 1 (was the statement shown during the learning phase?) with a constant probability to give a correct response (.8). If the response was "no," the it-

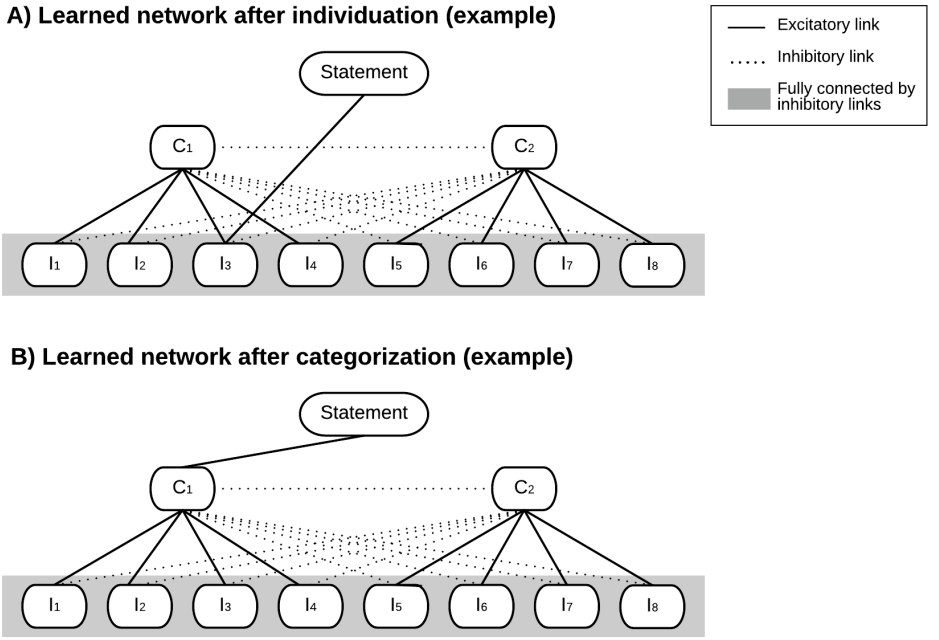

FIGURE 5. An illustration of the cognitive consequences of social categorization and individuation during learning trials of the "Who said what" paradigm. If the perceiver individuates the speaker, an association is learned between the statement and the identity of the speaker (e.g., the name). In contrast, if the perceiver categorizes the speaker, an association is learned between the statement and the category of the speaker (e.g., the sex). These associations have dissociable effects (correct speaker selection or within-category confusions respectively) during the test phase when the speaker needs to be retrieved based on the statement.

eration was terminated (consistent with the design of past studies). If the response was "yes," we next simulated a learning trial and used the resulting connectionist network to simulate the retrieval of the speaker in the corresponding test trial.

To simulate a learning trial, we took the network that resulted from Simulation 1 (i.e., lifetime learning; Figure 4) and added a statement node. Learning during a "Who said what" experiment was then simulated by applying our Hebbian learning mechanism a single time with the external input of the statement node set to one while sampling the external inputs of the corresponding identity node (e.g., I_3) and social category node (e.g., C_1) from normal distributions (for details see the descriptions of Simulations 2–4). All external inputs of other (non-observed) nodes were set to -0.1. Taken together, this simulated that participants read the statement on every learning trial (in line with instructions) but paid varying amounts of incidental attention to the (group and individual) properties of the speaker. Consequently, in some trials the simulated participants associated the statement with the specific identity of the speaker (e.g., I_3 which could denote the name *Peter*) and in other trials the simulated participants associated the statement to the social category of the speaker (e.g., C_1 which could denote the category *male*; see Figure 5).

Subsequently, we simulated the retrieval process in the corresponding test trial. The general idea was that the statement acts as a retrieval cue to retrieve the speaker of the statement. Speaker selection therefore depends on the associative links between the statement with the properties of the speaker. We simulated perception of the statement node by applying the connectionist perception mechanism with the external input of the statement node set to one and all other external inputs set to zero (i.e., reflecting that no person properties were observable at this moment). The individual node with the highest output activation was taken as the response in the test trial (i.e., the selected speaker).

Response frequencies of 50 simulated participants with 100 trials each were generated iteratively by repeating the procedure above 5000 times. To simulate several participants, we used the weights that resulted from Simulation 1 (lifetime learning) for 100 iterations (which can therefore be seen as trials performed by the same participant with one lifetime learning history) before applying Simulation 1 again to generate weights that were used for the next 100 iterations (i.e., to simulate a new participant with a different lifetime learning history). The response frequencies that resulted from this iterative simulation procedure were analyzed with regard to the difference between within-category and between-category confusions and also using the standard multinomial processing tree analysis for the "Who said what."

We programmed three versions in which we simulated the situations that perceived speakers are encoded both as group members and individuals (Simulation 2), perceived speakers are encoded exclusively as individuals (Simulation 3), and perceived speakers are encoded exclusively as group members (Simulation 4). The first purpose of these simulations was to reproduce the finding that people make more within-category than between-category errors dependent on how salient group properties are (Simulation 2). The second purpose was to reproduce evidence of two underlying cognitive components based on an MPT analysis (Simulations 3 and 4 compared to Simulation 2). Specifically, the standard MPT analysis estimates two parameters for social categorization (the probability of encoding category 1 and category 2 respectively) and two parameters for individuation (the probability of encoding the identity of an individual from category 1 and category 2 respectively). For the present discussion, the critical question was whether the two social categorization parameters can be influenced relatively independent of the two individuation parameters (and vice versa) based on the external inputs to category and individual nodes. It is worth mentioning that the MPT analysis also makes a number of other assumptions (Klauer & Wegener, 1998), which are not of central interest here but which are also tested below through the fit of the MPT model with our simulation data.

SIMULATION 2: CATEGORIZATION AND INDIVIDUATION

To simulate a situation in which both social categorization and individuation occur during learning trials, we sampled the external inputs of the identity node of

the correct speaker (μ = -0.1, σ = 1) and the category node of the correct speaker (μ = 2, σ = 1) from normal distributions with -0.1 as lower limit.[3] This simulated varying amounts of attention to the identity and category of speakers during learning trials. Before we assessed confusions between speakers, we first corrected between-category confusions for their overall higher chance of occurrence (relative to within-category confusions) by multiplying their frequency with 3/4 (Klauer & Wegener, 1998). In line with past findings, the results of a paired samples *t*-test showed that within-category confusions (M = 17.68; SD = 3.68) occurred significantly more often than (corrected) between-category confusions (M = 8.34; SD = 2.48), $t(49) = 14.63$, $p < .001$.

Next and importantly, the standard MPT analysis was applied to the simulated data (Klauer & Wegener, 1998). The fit of the standard MPT model with the data was satisfactory, $G^2 = 0.54$, $df = 1$, $p = .463$. The critical MPT parameter estimates and confidence intervals are depicted in Table 1. Importantly, both social categorization and individuation parameter estimates and the lower boundaries of their 95% confidence intervals were well above zero. Moreover, constraining the social categorization parameters to be equal to zero significantly reduced the model fit, $G^2 = 199.47$, $df = 2$, $p < .001$.[4] The same was true if the individuation parameters were constrained to be equal to zero, $G^2 = 133.78$, $df = 2$, $p < .001$. Thus, according to these results both social categorization and individuation occurred in our simulation. However, do the social categorization and individuation parameters capture two dissociable cognitive components? This was addressed in the next two simulations in which we tested whether social categorization and individuation can be eliminated independently.

SIMULATION 3: ONLY INDIVIDUATION

Simulation 3 was equivalent to Simulation 2 except that the external inputs of social category nodes were always set to -0.1. This simulated a situation in which participants did not perceive the speaker as a group member and therefore never encoded a statement-category link. As expected, the results of a paired samples *t*-test showed no significant evidence that within-category errors (M = 14.00; SD = 3.69) occurred more often than (corrected) between-category errors (M = 13.53; SD = 3.19), $t(49) = 0.64$, $p = .528$.

Next and importantly, the standard MPT analysis was applied to the data. The fit of the MPT model with the data was satisfactory, $G^2 = 0.41$, $df = 1$, $p = .523$. The critical MPT parameter estimates and confidence intervals are depicted in Table 2.

3. We set these means such that rates of correct speaker selection and within-category speaker confusions are relatively similar to those observed in past research (Klauer & Wegener, 1998). However, these means are relatively arbitrary in relation to the discussion above.

4. Setting the parameter values to the boundaries is somewhat problematic because it compromises the chi squared distribution of G^2 (Self & Liang, 2017). Ideally, this situation requires a more elaborate analytic approach. However, due to the high statistical power derived from the extensive simulation data, the conclusions are the same for either analytic approach. For the sake of simplicity, we report the simpler analyses.

TABLE 1. Critical Parameter Estimates and 95% Confidence Intervals (CIs) for Simulation 2

Parameter	Estimate	Lower CI	Upper CI
c_1	0.212	0.164	0.260
c_2	0.147	0.100	0.194
d_1	0.540	0.414	0.665
d_2	0.502	0.382	0.621

Note. The parameters c_1 and c_2 are the probabilities of remembering members of category C_1 and C_2 respectively (individuation) and the parameters d_1 and d_2 are the probabilities of remembering the social category of members of C_1 and C_2 respectively (social categorization).

Most importantly, social categorization parameter estimates were virtually zero while individuation parameter estimates and the lower boundaries of the confidence intervals were well above zero. Moreover, constraining the social categorization parameters to be equal to zero did not significantly reduce the model fit, $G^2 = 1.13$, $df = 2$, $p = .570$, whereas constraining individuation parameters to be equal to zero significantly reduced the model fit, $G^2 = 60.13$, $df = 2$, $p < .001$. Thus, according to these results only individuation occurred in our simulation.

SIMULATION 4: ONLY CATEGORIZATION

Simulation 4 was equivalent to Simulation 2 except that external inputs of individual nodes were always set to -0.1. This simulated a situation in which participants never construed a perceived speaker as an individual and therefore never learned a statement-individual link. As expected, the results of a paired samples *t*-test showed that there were significantly more within-category confusions ($M = 20.10$; $SD = 4.29$) than (corrected) between-category confusions ($M = 9.77$; $SD = 2.18$), $t(49) = 13.94$, $p < .001$. Hence, simulating zero reliance on individuation did not eliminate the phenomenon that people systematically confuse speakers within categories, as one would expect.

Next and importantly, the standard MPT analysis was applied to the simulated data (Klauer & Wegener, 1998). The fit of the MPT model with the data was satisfactory, $G^2 = 1.59$, $df = 1$, $p = .208$. The critical MPT parameter estimates and confidence intervals are depicted in Table 3. Most importantly, social categorization parameter estimates and the lower boundaries of the confidence intervals were well above zero while individuation parameter estimates were virtually zero. Moreover, constraining the social categorization parameters to be equal to zero significantly reduced the model fit, $G^2 = 212.71$, $df = 2$, $p < .001$, whereas constraining the individuation parameters to be equal to zero did not significantly reduce the model fit, $G^2 = 0.00$, $df = 2$, $p = 1$. Thus, according to these results only social categorization occurred in our simulation.

Taken together, our simulations successfully reproduced the established empirical findings. First, they reproduced the finding that within-group confusions happen more often than between-group confusions and that this effect increases with category salience. Second and importantly, the simulations reproduced evidence

TABLE 2. Critical Parameter Estimates and 95% Confidence Intervals (CIs) for Simulation 3

Parameter	Estimate	Lower CI	Upper CI
c_1	0.107	0.067	0.147
c_2	0.099	0.059	0.139
d_1	0.000	-0.151	0.151
d_2	0.066	-0.080	0.212

Note. The parameters c_1 and c_2 are the probabilities of remembering members of category C_1 and C_2 respectively (individuation) and the parameters d_1 and d_2 are the probabilities of remembering the social category of members of C_1 and C_2 respectively (social categorization).

of two dissociable underlying cognitive components (social categorization and individuation). This suggests that our interpretation of social categorization and individuation as different inputs of the connectionist process is consistent with this evidence. Hence while dissociations based on an MPT analysis are conventionally referred to as "process dissociations," they need not be: they may also reflect a dissociation between different types of inputs. Therefore, our single process synthesis of social categorization and connectionist assumptions is compatible with the dissociation.

DISCUSSION

In the present article, we aimed to contribute to the coherence and conceptual integration of the person perception literature by addressing long-standing conceptual issues. On the one hand, there is the influential assumption that people employ two processing strategies: categorization and individuation. This idea has helped substantially in explaining how people encode and memorize information about perceived people. Moreover, the idea that perceivers employ these two cognitive strategies is supported particularly by evidence of independent cognitive components in person memory (among others). On the other hand, it has been pointed out that the categorization-individuation distinction is conceptually problematic and (single process) connectionist models have been proposed as an alternative. However, two questions have remained open. First, it has remained unclear how (single process) connectionist models can be reconciled with the categorization-individuation distinction in (dual process) social categorization models. As a result, it remained unclear whether both models can be true at the same time. Second, it remained unclear how connectionist models could account for evidence of two independent cognitive components in person memory without assuming dual processes (see Figure 1).

Regarding the first question, we suggested that categorization and individuation can be seen as two different types of inputs to a single connectionist process. Consequently, person perception can be a "dual-model" in the input (computational level) while being a "single-model" with regard to the employed processes (algorithmic level), making social categorization and connectionist models compatible. Regarding the second question, we argued that evidence of a cognitive

TABLE 3. Critical Parameter Estimates and 95% Confidence Intervals (CIs) for Simulation 4

Parameter	Estimate	Lower CI	Upper CI
c_1	0.000	-0.039	0.039
c_2	0.000	-0.041	0.041
d_1	0.521	0.417	0.625
d_2	0.402	0.281	0.524

Note. The parameters c_1 and c_2 are the probabilities of remembering members of category C_1 and C_2 respectively (individuation) and the parameters d_1 and d_2 are the probabilities of remembering the social category of members of C_1 and C_2 respectively (social categorization).

dissociation can emerge not because of two different processes but because the connectionist process is applied to two different types of inputs. To support this argument, we showed in computer simulations that our interpretation of social categorization and individuation is consistent with relevant findings from person memory research—most importantly, a dissociation between two cognitive components (social categorization and individuation). In these computer simulations, we varied exclusively the input of the model while keeping the process part constant and showed that this was sufficient to replicate the cognitive dissociation. Overall, this leads us to conclude that social categorization and connectionist models (1) seem compatible with each other and (2) consistent with evidence of a cognitive dissociation. In the following, we will discuss how our framework can further help to bridge the social categorization and connectionist literatures and how it advances existing theories.

INTEGRATION OF SOCIAL CATEGORIZATION AND CONNECTIONIST MODELS

Our framework may help the social categorization and connectionist literatures to inform each other more effectively in the future. First, social categorization research can inform connectionist models about input assumptions. Our simulations can be seen as an illustration of this idea. By adjusting the input of our connectionist models based on the theorizing in the social categorization literature, we were able to coherently explain key findings in the person memory literature. Other findings may be explained in a similar fashion. For example, Van Rooy, Van Overwalle, Vanhoornissen, Labiouse, and French (2003) presented a connectionist model that explained the phenomenon that grouping people (e.g., based on ethnicity) accentuates their perceived between-group differences. Essentially their model assumed that there are some nodes (which we would call category nodes) that are excited during the perception of any member of a social group. These nodes therefore became associated with the traits (e.g., likability) of the members of this group (e.g., an ethnic group), which subsequently accentuated judgements on the trait dimension (e.g., making the group members appear more likable). Our framework makes explicit that this connectionist explanation (and potentially oth-

ers) can also be seen as a social categorization explanation. Future research may apply this integrative explanatory approach to other documented phenomena.

Conversely, connectionist models may contribute to the social categorization literature by specifying the processes by which social categorization notions translate into measurable phenomena (i.e., outputs). This idea converges with recent arguments that social categorization appears to be driven by a dynamic (e.g., connectionist) process (Freeman & Ambady, 2011). An implication of this idea is that research on learning mechanisms in human cognition may shape the predictions of social categorization models. That is, dependent on the proposed connectionist learning mechanism (e.g., Hebbian learning), construing perceivers as either group members or individuals (i.e., learning inputs) may result in different associative networks, which lead to different predicted person perception outputs. According to this view, social categorization and connectionist models are natural extensions of each other that shed light on different aspects of the same underlying cognitive system.

THE INTERNAL COHERENCE OF SOCIAL CATEGORIZATION MODELS

The labels "social categorization" and "individuation" have not been defined unequivocally in past research and as a result these labels have been applied somewhat inconsistently (Cox & Devine, 2015; Kunda & Thagard, 1996; Quinn & Macrae, 2005). In fact, some researchers have argued that the distinction is artificial and may be better avoided (Cox & Devine, 2015; Kunda & Thagard, 1996). Our framework provides steps to address this issue by providing a more formal interpretation of the categorization-individuation distinction. Specifically, in our connectionist model social categorization constitutes mapping people onto representations (e.g., male) that are generally mapped onto any member of some social group (e.g., any man) while individuation constitutes mapping people onto representations (e.g., Peter) that are generally mapped exclusively onto that individual. This distinction was implemented in computer simulation, which addresses concerns that the distinction is conceptually problematic: that is, the computer simulation would not work if the conceptual distinction was not internally consistent.

Our connectionist interpretation may also help to derive more unequivocal explanations and predictions from social categorization models. For instance, it was theorized that memory confusions are caused by "social categorization" but it remained ambiguous what constitutes "social categorization." For example, personality traits were often explicitly distinguished from "social categories" in past research (Fiske & Neuberg, 1990; Tajfel, 1969; Tajfel & Wilkes, 1963), which may cause a reader of the literature to believe that representing a person in terms of a personality trait (e.g., trustworthy) would not lead to the commonly found memory confusions (let alone a cognitive dissociation). In contrast, our framework and computer simulations suggest that all that is necessary for these memory confusions to occur is that the perceiver maps several observed people onto the same mental representation—and there seems to be no a priori reason why mapping

several people onto the same personality trait (e.g., trustworthy) should be an exception. In line with this argument, we recently found evidence that memory confusions occur between (un)trustworthy looking faces and that these confusions occur more often when trustworthiness is made salient by instructions (Klapper, Dotsch, van Rooij, & Wigboldus, 2016). Moreover, a multinomial processing tree analysis showed evidence that two cognitive components were driving this phenomenon: one that distinguishes between people based on their trustworthiness (social categorization) and one that distinguishes between individuals (individuation). These findings were not clearly predicted from the past literature but follow naturally from our framework and computer simulations.

This can also help to make more unequivocal predictions for future research on person memory. Although we discussed our computer simulations primarily as a model of the "Who said what" paradigm, the general logic of the simulations also applies to other memory contexts such as identifications of potential culprits though eye witnesses. In general, the simulation describes how a perceiver may recognize a previously perceived person (e.g., a speaker or culprit) based on a provided cue (e.g., a statement or crime). Consequently, our work has implications for the interpretation of witness testimony. Past research has demonstrated that people seem to rely more on categorization when perceiving other-race faces (relative to same-race faces) causing more memory confusions (Bernstein, Young, & Hugenberg, 2007; Hugenberg et al., 2007, 2010). The current work and recent evidence (Klapper et al., 2016) suggest that this phenomenon may be more general. To give one possible example, suppose that a witness observed an untrustworthy looking person committing a crime and that this witness has a relatively strong tendency to categorize people into trustworthy and untrustworthy. In that case, the witness may falsely accuse an untrustworthy appearing suspect of being the culprit because of an increased tendency to confuse untrustworthy with other untrustworthy looking people. A consequence of this theoretical prediction is that testimony of witnesses who have a general tendency to categorize may generally (e.g., even if same-race faces are observed) produce more false positives than testimony of witnesses who have a tendency to individuate. Future research may further investigate such possibilities.

LIMITATIONS AND FUTURE RESEARCH

Our computer simulation (necessarily) adopts specific processing assumptions such as the Hebbian learning mechanism. As mentioned before, those details are not important to the general argument that social categorization and connectionist models are compatible in principle (with each other and with evidence of a cognitive dissociation) but it limits the degree to which that insight can be generalized to specific models in the literature (e.g., connectionist models with a different learning mechanism). Similarly, some social categorization theories make more specific assumptions (e.g., that individuation entails featural or more graded processing; Blair, Judd, & Chapleau, 2004; Brewer, 1988; Fiske & Neuberg, 1990).

Our computer simulations abstract away from such specific assumptions and may better be seen as a general framework based on a relatively global distinction between categorization and individuation. Future research may complement our work by exploring the compatibility of different theories using different processing assumptions (e.g., learning mechanisms) and by adding more details to the categorization-individuation distinction (e.g., featural vs. holistic processing). In addition, future research can further test the framework by extending it to more specific findings such as the higher malleability of individuation to cognitive load (Klauer & Wegener, 1998) or effects of structural fit and situational relevance on categorization (Blanz, 1999; Van Twuyver & van Knippenberg, 1995).

CONCLUSION

The idea that perceivers can construe other people as individuals (individuation) or group members (social categorization) has been highly influential in the social perception literature. Another influential idea is that social perception is driven by a dynamic connectionist process in which activation spreads via learned associations between internal representations. We presented a framework that synthesizes these broad theoretical ideas and demonstrated how this framework can account for key evidence in the person memory literature. This framework aims to contribute to the coherence of the literature, provides conceptual bridges between the social categorization and connectionist literature, and helps to alleviate concerns that the categorization-individuation distinction is a conceptual artifact. Finally, we hope that our work can serve as a stepping stone toward more unified models of person memory and perception.

APPENDIX A. MARR'S LEVELS

While Marr's levels have been influential for decades, the exact interpretation of these three levels (computational, algorithmic, and implementational) has varied somewhat in the literature (McClelland, 2009). Our interpretation of Marr's levels is consistent with recent mathematical interpretations (Thagard & Verbeurgt, 1998; van Rooij & Wareham, 2012; van Rooij, Wright, & Wareham, 2010). To clarify our usage of Marr's levels, we provide a brief explanation below with the main focus on the computational and algorithmic level.

To begin with, consider a simplified explanation: while the computational level describes "objects" or "states," the algorithmic levels describe active "events" that happen to these "objects." In our coffee machine example, the objects are the coffee capsules, the water, and the cup (computational level) while the active events are pressing the water through the coffee capsule into the cup (algorithmic level). In a connectionist model, the main objects (or states) are nodes, activation levels of these nodes, association weights, and external inputs to the nodes (see Figure 3). These objects constitute (much of) the computational level. The algorithmic level is the active events that happen to these objects. In the connectionist model, the main events are (1) updating the activation levels based on external inputs, (2) updating the activation levels based on spread of activation via associations, and (3) continuous activation decay. These points belong to the algorithmic level because they describe what actively happens to the objects.

More formally, the computational level describes a function $f{:}I \rightarrow 0$ where I is the set of inputs, O is the set of outputs. A function is a mapping that defines which input $i \in I$ belongs to which output $o \in O$. As such, the computational level entails three aspects: (1) what the input domain I is, (2) what the input domain O is and (3) how elements of the input domain are mapped onto elements of the output domain ($\rightarrow$). The specification of the input domain I and output domain O is what we informally referred to as describing "objects" in the previous paragraph. This was oversimplified in the sense that the computational level also entails a mapping ($\rightarrow$) from elements of the input domain to elements of the output domain. For example, a full computational level description of the coffee machine would not only describe the coffee capsule, water, and cup but also clarify which type of coffee is returned for which type of coffee capsule (e.g., that caramel coffee will be the output if a caramel capsule is provided as input and that a mocha coffee will be the output if a mocha capsule is provided as input or more generally that the returned coffee will have acquired the taste of the content of the coffee capsule). The latter constitutes the mapping. Similarly, an informal computational level description of the perception mechanism that we employ in the present article is that it takes initial activation levels, external inputs, and association weights (inputs), and returns final activation levels (outputs) such that an equilibrium is achieved between all forces that act on the activation of a node (mapping). However, the mapping aspect of the computational level is not directly relevant to the argument in the present article and so we have devoted less attention to it. Interested readers are referred to the work of Thagard and Verbeurgt (1998) or McClelland and Rumelhart (1989) for more detailed computational level descriptions of connectionist models. For the purpose of the present article, it suffices to think of the computational level of a description of objects and their properties or more mathematically the input domain I and the output domain O.

What is the algorithmic level? Simply put, the algorithmic level is the algorithm that derives the outputs from the inputs. In programming, the computational level constitutes the

documentation of a function ("the function takes x and returns y") while the algorithmic level constitutes the algorithm (/code) behind the function. Especially, input and algorithm are relatively easy to distinguish in computer code. For example, our computer simulations employ two main functions: the learning and the perception function. Both of these functions take inputs (part of the computational level), which are defined between brackets in the first line of the definition of the function (although some of them were hard coded because they were not of interest here). Below this line one can find various lines of code (algorithmic level), which constitute the algorithm that is applied to these inputs.

Based on this outline of the computational and algorithmic level in our computer simulations, we can describe the main argument of our article in more general terms. A key point of our argument is that the "process" dissociation (i.e., a dissociation at the algorithmic level) shown through MPT modeling need not to be interpreted as a process dissociation but may alternatively be seen as an input dissociation (i.e., a dissociation at the computational level). In the language of a programmer, this claim sounds as follows: it is not necessary to vary the algorithms in the computer simulations in order to reproduce the MPT dissociation. Instead, it suffices to keep the algorithms constant and vary two different parts of the inputs (i.e., what is entered between the brackets of the function). In our case, this entailed varying the external inputs to nodes while keeping the learning and perception algorithms constant. We hope that this helps to clarify the distinction between computational and algorithmic level and how that distinction maps onto our argumentation and computer simulations.

Finally, what is the implementational level? Essentially, the implementational level describes the inputs, outputs, and the algorithm in terms of physical entities. For example, saying that the brain adds two numbers x and y constitutes an algorithmic level description while saying that the brain adds two number x and y by feeding the firing rate x of neuron 1 and the firing rate y of neuron 2 into the same neuron constitutes an implementational level description. As such, whether a model belongs to the algorithmic or implementational level depends on whether the model describes the algorithm in terms of physical entities or merely abstractly without a clear mapping onto physical entities in the world.

Some readers may feel that connectionist models may better be placed at the implementational level given that they seem conceptually close to neural networks in the brain. We take a neutral position on this point. Although connectionist models certainly resemble neural networks in the brain, the exact mapping onto neural processes is usually left open. In addition, there are other models, which are considerably closer to neural networks in the brain than the connectionist model we employ (e.g., Schröder & Thagard, 2013). As such, whether connectionist models should be situated at the implementational level is somewhat debatable. More importantly though, whether the connectionist model belongs to the implementational level is not essential for the present article. Our unification works if the "dual process" part of social categorization models (i.e., the categorization-individuation distinction) is situated at a different level than the "single process" part of connectionist models. Our proposal is to situate the "dual part" at the computational level and situating the "single part" at either the algorithmic or the implementational level would enable social categorization and connectionist models to be compatible. As such, situating connectionist models at the implementational level would not change the conclusions of the present article.

APPENDIX B. FORMAL DETAILS OF COMPUTER SIMULATIONS

Learning Mechanism. Learning starts with a set of nodes with weighted links between the nodes. The strength of a link between two nodes *i* and *j* is represented by a numerical weight w_{ij}. At the onset of learning all weights are set to zero. Next, weights are updated iteratively based on a set of stimuli that constitutes the learning input. Each stimulus in this learning input is formally represented by a vector of external inputs ext_i for each node *i* in the network. We used the standard Hebbian learning algorithm for auto-associators by Rumelhart and McClelland (1989) to update the association weights based on the external inputs. Specifically, weights were updated by applying the following learning rule:

$$\Delta w_{ij} = \eta * ext_i * ext_j$$

where η is the learning rate, ext_i is the external input of node *i* and ext_j is the external input of node *j*. In all simulations, the learning rate was $\eta = 0.01$, which is a standard value (Freeman & Ambady, 2011; McClelland & Rumelhart, 1989). Weights of self-connections were permanently set to zero ($w_{ii} = 0$).

Perception Mechanism. Our formal implementation of the person perception mechanism adopts standard connectionist assumptions (Freeman & Ambady, 2011; McClelland, 1991; McClelland & Rumelhart, 1989; Rumelhart, Hinton, & McClelland, 1986). Each node in the network had a numerical activation level, which was initially set to zero. The activation of each node *i* was then updated iteratively based on its net input. The net input of node *i* is

$$net_i = \sum_j w_{ij} * 0_j + ext_i + \varepsilon_{0.01}$$

where w_{ij} is the association weight between node *i* and *j*, o_j is the output of node *j*, ext_i is the external input of node *i*, and $e_{0.01}$ is normally distributed noise with a mean of 0 and standard deviation of 0.01. The latter reflects the noisy conditions under which the brain processes information (see also Freeman & Ambady, 2011). The output of node *j* is the amount of positive activation of node *j*:

$$o_j = max(a_j, 0)$$

where *aj* is the activation of node *j*. In other words, if the activation of a node becomes negative, it does not spread activation to other nodes. This is a common assumption in connectionist models of this type (Freeman & Ambady, 2011; McClelland, 1991; Rumelhart et al., 1986). Once the net input for all nodes had been computed, the activations of all nodes were updated in parallel as follows:

If $net_i > 0$:
$\Delta \alpha_i = I(M - \alpha_i)a_i - D * \alpha_i$
If $net_i \leq 0$:
$\Delta \alpha_i = I(\alpha_i - m)\, a_i - D * \alpha_i$

where *M* and *m* are the maximum and minimum activations respectively, *I* is a constant that scales the effect of the net input on the activation of the node, and *D* is a constant that scales the tendency of activations to decay to zero. In all simulations, we used the standard values $M = 1$, $m = -0.2$, $D = 0.1$, and $I = 0.4$ (Freeman & Ambady, 2011; McClelland, 1991; Rumel-

hart et al., 1986). Activations of nodes were updated iteratively according to the formulas above until one of two standard stopping conditions was met: (1) the maximum change in activations is smaller than 0.01 or (2) the number of iterations exceeds 200. At this point, the updating stopped and the activations were interpreted as the output of the person perception process (e.g., a memory retrieval result).

REFERENCES

Agostinelli, G., Sherman, S., Fazio, R., & Hearst, E. (1986). Detecting and identifying change: Additions versus deletions. *Journal of Experimental Psychology: Human Perception and Performance, 12*(4), 445-454.

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

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect. *Psychological Science, 18*(8), 706–712.

Blair, I. V, Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science, 15*(10), 674-679. https://doi.org/10.1111/j.0956-7976.2004.00739.x

Blanz, M. (1999). Accessibility and fit as determinants of the salience of social categorizations. *European Journal of Social Psychology, 29*, 43-74.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition, Vol. 1. A dual model of impression formation* (pp. 1-36). Hillsdale, NJ: Erlbaum.

Cox, W. T. L., & Devine, P. G. (2015). Stereotypes possess heterogeneous directionality: A theoretical and empirical exploration of stereotype structure and content. *PlosOne*, *10*(3), 1–27.

Dalege, J., Borsboom, D., Van Harreveld, F., Van Den Berg, H., Conner, M., & Van Der Maas, H. L. J. (2016). Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychological Review, 123*(1), 2-22. https://doi.org/10.1037/a0039802

De Houwer, J., & Moors, A. (2015). Levels of analysis in social psychology. In B. Gawronski & G. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 24-40). New York: Guilford.

Ehret, P. J., Monroe, B. M., & Read, S. J. (2014). Modeling the dynamics of evaluation: A multilevel neural network implementation of the iterative reprocessing model. *Personality and Social Psychology Review, 19*(2), 148-176. https://doi.org/10.1177/1088868314544221

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1-74). New York: Academic.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review, 118*(2), 247-279. https://doi.org/10.1037/a0022327

Gawronski, B., Ehrenberg, K., Banse, R., Zukova, J., & Klauer, K. C. (2003). It's in the mind of the beholder: The impact of stereotypic associations on category-based and individuating impression formation. *Journal of Experimental Social Psychology, 39*(1), 16-30. https://doi.org/10.1016/S0022-1031(02)00517-6

Hugenberg, K., Miller, J., & Claypool, H. M. (2007). Categorization and individuation in the cross-race recognition deficit: Toward a solution to an insidious problem. *Journal of Experimental Social Psychology, 43*(2), 334-340. https://doi.org/10.1016/j.jesp.2006.02.010

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review, 117*(4),

1168-1187. https://doi.org/10.1037/a0020463

Klapper, A., Dotsch, R., van Rooij, I., & Wigboldus, D. H. J. (2016). Do we spontaneously form stable trustworthiness impressions from facial appearance? *Journal of Personality and Social Psychology, 111*(5), 655-664.

Klauer, K. C., Hölzenbein, F., Calanchini, J., & Sherman, J. W. (2014). How malleable is categorization by race? Evidence for competitive category use in social categorization. *Journal of Personality and Social Psychology, 107*(1), 21-40. https://doi.org/10.1037/a0036609

Klauer, K., & Wegener, I. (1998). Unraveling social categorization in the "who said what?" paradigm. *Journal of Personality and Social Psychology, 75*(5), 1155-1178.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review, 103*(2), 284-308.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51*, 93-120. https://doi.org/10.1146/annurev.psych.51.1.93

Macrae, C. N., & Bodenhausen, G. V. (2001). Social cognition: Categorical person perception. *British Journal of Psychology, 92*(1), 239-255. https://doi.org/10.1348/000712601162059

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.

McClelland, J. L. (1991). Stochastic interactive activation and the effects of context on perception. *Cognitive Psychology*, *23*(1), 1–44.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science, 1*, 11-38. https://doi.org/10.1111/j.1756-8765.2008.01003.x

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*(3), 419-457.

McClelland, J. L., & Rumelhart, D. E. (1989). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*. Cambridge, MA: MIT Press.

Quinn, K., & Macrae, C. N. (2005). Categorizing others: The dynamics of person construal. *Journal of Personality and Social Psychology, 88*(3), 467-479. https://doi.org/10.1037/0022-3514.88.3.467

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, *1*, 45-76.

Schröder, T., & Thagard, P. (2013). The affective meanings of automatic social behaviors: Three mechanisms that explain priming. *Psychological Review, 120*(1), 255-280. https://doi.org/10.1037/a0030972

Seidenberg, M. S., & Plaut, C. (2014). Quasiregularity and its discontents: The legacy of the past tense debate. *Cognitive Science, 38*, 1190-1228. https://doi.org/10.1111/cogs.12147

Self, S. G., & Liang, K. (2017). Asymptotic properlies of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association, 82*(398), 605-610.

Smith, E. R. (2009). Distributed connectionist models in social psychology. *Social and Personality Psychology Compass, 3*(1), 64-76. https://doi.org/10.1111/j.1751-9004.2008.00160.x

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.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*(2), 108-131. https://doi.org/10.1207/S15327957PSPR0402

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Biosocial Science, 1*, 173-191. https://doi.org/10.1017/S0021932000023336

Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgement. *British Journal of Psychology , 54*, 101–114.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology, 36*(7), 778-793. https://doi.org/10.1037//0022-3514.36.7.778

Thagard, P., & Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science, 22*(1), 1-24. https://doi.org/10.1207/s15516709cog2201_1

Van Overwalle, F., & Labiouse, C. (2004). A recurrent connectionist model of person impression formation. *Personality and Social Psychology Review, 8*(1), 28-61. https://doi.org/10.1207/S15327957PSPR0801_2

van Rooij, I., & Wareham, T. (2012). Intractability and approximation of optimization theories of cognition. *Journal of Mathematical Psychology, 56*(4), 232-247.

van Rooij, I., Wright, C. D., & Wareham, T. (2010). Intractability and the use of heuristics in psychological explanations. *Synthese, 187*(2), 471-487. https://doi.org/10.1007/s11229-010-9847-7

Van Rooy, D., Van Overwalle, F., Vanhoomissen, T., Labiouse, C., & French, R. (2003). A recurrent connectionist model of group biases. *Psychological Review, 110*(3), 536-563. https://doi.org/10.1037/0033-295X.110.3.536

Van Twuyver, M., & van Knippenberg, A. (1995). Social categorization as a function of priming. *European Journal of Social Psychology, 25*(560), 695-701.

Young, S. G., & Hugenberg, K. (2011). Individuation motivation and face experience can operate jointly to produce the own-race bias. *Social Psychological and Personality Science, 3*(1), 80-87. https://doi.org/10.1177/1948550611409759

Zebrowitz, L. A., Fellous, J.-M., Mignault, A., & Adreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. *Personality and Social Psychology Review, 7*(3), 194-215. https://doi.org/10.1207/s15327957pspr0101_1