

Social-cognitive representational spaces: ontology and consequences

Submitted by:

Joel Eduardo Martinez
Psychology Department

In partial fulfillment of the requirements
For the Degree of Master of Arts
Princeton University
Princeton, New Jersey
Fall 2017

Master's Committee:
Advisor: Alexander Todorov
Elizabeth Levy-Paluck
Uri Hasson

While the nature, existence, and necessity of representations in cognition is debated (Dorffner, 1999; Pereira Jr, 1999), they can be described as relationships between mental or brain states and events or objects in the world (Scheutz, 1999) whose function is to generate adaptive behavior (Parkinson & Wheatley, 2016; Peschl & Alexander, 1999). This predictive function of representations would be particularly useful for navigating social life and relationships. A rich body of classic social psychology already exists on the organization of social memory (e.g., Fiske, 1993; Hastie & Kumar, 1979; Kashima & Kerekes, 1994; Wyer, 1980; Wyer & Srull, 1989), but methodological advances in psychology and neuroscience (Cohen et al., 2017) have allowed more complex inquiries into the structure and representation of social knowledge (Montagrin, Saiote, & Schiller, 2017; Tavares et al., 2015). Thus, this review is a synthesis for what unique insights these innovations have added to the social memory discourse.

Some of the recent methods for assessing the representational stimulus structures used and constructed by the mind from neural or behavioral responses include multi-voxel pattern analysis (Haxby, Connolly, & Guntupalli, 2014; Norman, Polyn, Detre, & Haxby, 2006), representational modeling (Diedrichsen & Kriegeskorte, 2017; Kriegeskorte, Mur, & Bandettini, 2008; Xue, Weng, He, & Li, 2013), repetition suppression (Barron, Garvert, & Behrens, 2016), psychophysics (Guttman & Kalish, 1956; Hearst, 1960), reverse correlation (Dotsch & Todorov, 2012), memory intrusions (Crawford, Sherman, & Hamilton, 2002; Favila, Chanales, & Kuhl, 2016), multidimensional scaling (Edelman, Grill-Spector, Kushnir, & Malach, 1998; Shepard, 1962; Shinkareva, Wang, & Wedell, 2013), and item arrangements or sorting (Alves, Koch, & Unkelbach, 2016; Kriegeskorte & Mur, 2012). While the effective use of many of these methods requires caution (Davis & Poldrack, 2013), their relative success can be attributed to their reliance on interstimulus similarity.

There have been many ways to describe the similarity between stimuli, such as the amount of shared features (Tversky, 1977), analogical relations (Gentner & Markman, 1997; Gentner & Smith, 2012), and psycho-geometric distance (Connolly et al., 2012; Kriegeskorte & Mur, 2012). Stimuli that are considered psychologically or visually similar cluster together in some dimensional space or elicit overlapping neural responses. Despite varied definitions, the important detail is that similarity continues to be an important construct for measuring and organizing mental representations (Kriegeskorte & Kievit, 2013) and structuring social life (Jacoby-Senghor, Sinclair, & Smith, 2015; Leonardelli, Pickett, & Brewer, 2010). As such, specifying the relationship between mental and social structures through the lens of similarity could provide a mechanistic account of the reciprocal link between psychological determinants of sociality and sociality's structuring of the mind's social content.

This review attempts an ontology of the mental representation of fellow human beings with a focus on similarity as an overarching theme. An ontology identifies the essence of representational spaces and requires mapping out foundational components and their interrelations. This overview is separated into four sections: *content*, *process*, *consequences*, and *future directions*. The first aims to tackle how sociality structures the mind by describing important factors that cluster or repel social representations as revealed by neuroimaging and behavioral experiments. The process section proposes associative memory mechanisms by which these representational spaces could be maintained, organized, and updated. The third section tackles the psychological determinants of sociality by describing the social and behavioral consequences of these spaces. To conclude, future directions for this area of research are outlined.

Content of social representational spaces

There are surely an infinite number of dimensions that play a role in social memory, however I will focus on some main themes shared across studies of visual and semantic stimulus similarity: *social category, valence, typicality, distance*. The final section will try to relate these dimensions to models of social cognition that identify fundamental dimensions of social perception.

Social Category

Categorization in social psychology is seen a heuristic in interpersonal cognition (Macrae & Bodenhausen, 2001) and has enjoyed a very imprecise working definition throughout the literature (Klapper, Dotsch, van Rooij, & Wigboldus, 2017). While the vast research on the origins and processes of social categorization is beyond the scope of this review, it suffices to say that that social categorization is ubiquitous (Lieberman, Woodward, & Kinzler, 2017; Macrae & Bodenhausen, 2001; McGarty, 1999; Quinn & Rosenthal, 2012; Tajfel, 1981).

It is thought that a consequence of social categories (e.g., minimal groups, race, gender) is to homogenize within-category memory representations (Hugenberg & Sacco, 2008; Pattyn, Rosseel, Van Overwalle, & Van Hiel, 2015). For example, members that belong to a group are treated as interchangeable as shown by more within-group confusion errors in target identification (Taylor, Fiske, Etcoff, & Ruderman, 1978) and the increased transfer of traits between highly entative (i.e., cohesive) members (Crawford et al., 2002). Thus, memory errors suggest that within-group members occupy a similar psychological space (and are therefore more confusable) relative to individuals across groups, possibly through strategic differences in encoding (Hugenberg & Sacco, 2008; Johnson & Queller, 2003). Of course, an issue with memory performance is that it is an indirect measure of representational similarity, however neural data also supports this homogenizing effect.

One recent study sorted participants into minimal groups and scanned them using fMRI while reading sentences that described ingroup and outgroup members (Cikara, Van Bavel, Ingbreetsen, & Lau, 2017). They trained a multivoxel pattern classifier that was able to accurately classify which stimuli were in or out group members from distributed neural activations. In a second study, they were able to use this same classifier from the minimal group to cross-classify in and out group members in a second category, political party affiliation, providing evidence for a common neural code for ingroups across social categories. Likewise, another study was able to decode personality traits from multivoxel patterns of individuals' stories (Hassabis et al., 2013). Importantly, the voxels used for classification had to respond similarly to instances within each category for the classifier's decision boundary to successfully decode differences (Norman et al., 2006). In a related study using the same minimal group task design, repetition suppression was exhibited to in group members from both minimal groups and political affiliations (Lau & Cikara, 2017). Repetition suppression is a reduction in neural responses that occurs when a stimulus is repeatedly presented. This means that for ingroup members from multiple categories to induce repetition suppression, the brain had to register them as highly similar. A final study used a combination of these methods to show that repetition suppression occurs to different images of the same identity. Moreover, individuals presented with equal amounts of biographical information (i.e., more vs. less known) showed more correlated voxel activity patterns (Verosky, Todorov, & Turk-Browne, 2013). These studies can be considered evidence for the homogenizing effect of social categories on representations.

The relationship between individuals and a social category can also be reflected in stimulus similarity. While not specific to memory, clustering social stimuli by their ratings on multiple social dimensions (i.e., trait space) showed that people sort individuals into subgroups (e.g.,

business women) compared to supraordinate categories (e.g., women) (Pattyn et al., 2015). A recent study using mouse tracking to categorize stimuli showed that mouse deviations from the chosen category was associated with neural representational similarity to multiple categories. In other words, the co-activation of competing social categories by a stimulus is reflected in similarity of the stimulus' representation to category representations. These studies provide evidence that the representation of an individual can be used to designate their respective category, regardless of the mode of representation (e.g., trait space or on-line motor dynamics).

Lastly, the relationship between social categories can be illuminated by the similarity between category representations. Reverse correlation is a technique that “visualizes” the representation of a category or stimulus in participants' minds through successive decisions between images of a base face with random noise filters (Dotsch & Todorov, 2012). The end product is a coarse composite image of the face that the participant was thinking of. This technique was used to show that prejudiced individuals' representation of the poor also contained a racialized component. Those composite images of the poor were rated as depicting black individuals, showing that class and race are highly connected categories in stereotypes (Lei & Bodenhausen, 2017). Likewise, this technique was used to show that subgroups (typical subcategories, e.g., Black athletes) are visually represented more similarly to the supracategory (e.g., Black) relative to subtypes (atypical subcategories, e.g., Black businessmen) (Hinzman & Maddox, 2017).

At the neural level, representational similarity has shown that cultural stereotypes are reflected by neural similarity at the intersection of multiple categories (e.g., angry black man) (Stolier & Freeman, 2016). This intersection can also be shifted by individual characteristics. For example, greater implicit bias can increase the neural representational dissimilarity between black and white faces (Brosch, Bar-David, & Phelps, 2013). Multivoxel patterns have also been used to

examine the relationship between multiple types of moral violations (Koster-Hale, Saxe, Dungan, & Young, 2013; Wasserman, Chakroff, Saxe, & Young, 2017) and can portray the dynamic structure of social semantics, such that context-specific relationships (e.g., which groups are more spiritual? political?) between social groups can be mapped out (Leshinskaya, Contreras, Caramazza, & Mitchell, 2017).

Overall, this section shows that social categories push within-category representations together. These representational distances can help categorize individuals into categories and find commonality between categories.

Valence

Valence is considered a fundamental aspect of person perception (Fiske, Cuddy, & Glick, 2007; Oosterhof & Todorov, 2008; Tamir, Thornton, Contreras, & Mitchell, 2015). In particular, negativity is thought to have primacy in salience (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001) and adaptive value in memory encoding (Bell & Buchner, 2012). However, how valence structures the organization of memory representations has only become a recent inquiry.

According to the density hypothesis which argues for an inherent structure of valenced information (Koch, Alves, Krüger, & Unkelbach, 2016; Unkelbach, Fiedler, Bayer, Stegmüller, & Danner, 2008), positive information is represented more similarly (i.e., dense) than negative information. They argue this information structure exists because the range of positive stimuli is smaller, sandwiched between too little or much. Additionally, the base rate of positive information in the environment is high compared to negative information which translates to greater co-occurrence and therefore associative similarity. In other words, negative information is rare, thereby is represented dissimilarly and therefore stands out. Initial evidence for this hypothesis came from the presentation of positive and negative attitude words were rated for pairwise

similarity and these distances were visualized with multidimensional scaling (Unkelbach et al., 2008). While both negative and positive words clustered within themselves, the negative words were more distant out from each other. This framework has since been applied to understand recognition memory performance (Alves et al., 2015), asymmetries in the valence of shared attributes (Alves, Koch, & Unkelbach, 2017), and the representation of social individuals (Alves et al., 2016). That last study found that liked individuals are considered more similar to each other than disliked individuals. They argue that there are more diverse ways to display negative traits than positive traits which can lead to less overall trait matches between disliked individuals.

A recent study used fMRI-based representational similarity to show that the following categories showed high within-category similarities: social, non-social, positive, neutral, negative. Interestingly, the overall results show about equal amounts of similarity within positive and negative stimuli. However, when examined by specific brain regions, the ventral medial prefrontal cortex and bilateral amygdala showed the pattern predicted by the density hypothesis: greater similarity within negative stimuli relative to positive. These results suggest that more work is needed to directly examine the existence and function of the density effect of valence on neural representations.

This section provides emerging evidence that valence may affect the interstimulus similarity of memory representations such that positive stimuli are inherently more similar to each other than negative stimuli are. This effect may also operate in the social domain.

Typicality

Typicality can be conceptualized at the level of an individual (e.g., acting in unexpected ways) and at the level of society (e.g., this type of person is not common in our group or society) and both are considered salient in interpersonal cognition (Barclay, 2008; Chang & Sanfey, 2009;

Mende-Siedlecki, Baron, & Todorov, 2013; Morewedge & Todorov, 2012). For example, typicality is a determinant of facial trustworthiness (Sofer, Dotsch, Wigboldus, & Todorov, 2015), moral norms (Lindström, Jangard, Selbing, & Olsson, 2017), and tends to be positively valenced (Alves et al., 2017). Conversely, atypicality in person types and attributes is more memorable (O’Sullivan & Durso, 1984; Woll & Craesser, 1982). But how does typicality shape representations?

The most direct evidence comes from research on memory for faces (Corneille, Huart, Becquart, & Brédart, 2004). Participants categorized faces that varied in ethnic typicality and were asked in two time points a week apart to identify the face that was originally presented from a set of foils. The change in response between the two time points showed that participants thought they saw a face that was more typical of their categorization of the original face. In other words, their memory for the face was distorted towards ethnic typicality. Less direct evidence comes from face evaluation (Dotsch, Hassin, & Todorov, 2016; Sofer et al., 2017). These studies show that environmental face statistics, whether one’s own cultural space or being exposed to specific face distributions in a short-duration lab experiment, can shift what’s considered a typical face which also influences social evaluations. While this is not a direct examination of representations, one could argue that these statistical environments are moving internal representations within “face space” towards dynamically changing typicality prototypes. A similar effect is observed in the brain’s adaptation and recalibration to distorted faces (Kloth, Rhodes, & Schweinberger, 2017).

There is very little work on how typicality shapes neural social representations, however relevant insights come from object category learning. One study using multi-voxel pattern analysis found that typical objects are those that show more similarity to objects in same category (Davis & Poldrack, 2014). Another study also showed that the pattern of a category was more similar to

typical objects from that category (Jordan, Greene, Beck, & Fei-Fei, 2016) and, moreover, that typical objects also created sharper category boundaries by increasing within-category cohesiveness and between-category distinctiveness.

This section provides an overview of how typicality serves to shift representations towards a default prototype and creates category boundaries. More research is needed to verify that these effects also occur for social stimuli.

Distance

The definition of distance reviewed in this section can reflect spatial distance (e.g., physically far or close), temporal distance (e.g., distant past or recently), or interpersonal distance (e.g., friendships, strangers, or social standing in a group) as they are related by a common representational code (Parkinson, Liu, & Wheatley, 2014; Parkinson & Wheatley, 2013). However, this section will focus on social distance as it plays a critical role in understanding the self and others (Zink et al., 2008) and people are cognitively and physiologically attuned to this type of information, e.g., social hierarchies (Chiao, 2010; Koski, Xie, & Olson, 2015; Mehta & Prasad, 2015; Utevsky & Platt, 2014). Social distance has been shown to affect two types of mental representations: the self vs. others and others vs. others.

While not a direct examination of representations, early work shows that memory advantages when processing the self are also exhibited when processing intimately close, but not distant or simply familiar, social others (Wagner, Haxby, & Heatherton, 2012). Functional MRI studies using univariate contrasts (i.e., the average neural activation in a region for one condition subtracted by another) also showed that close relationships elicit similar neural activation as the self in medial prefrontal regions (Han, Ma, & Wang, 2016), although this may depend on the specific relationship and culture (Wang et al., 2012). Moreover, competition is a context that

modulates social distance such that confusing one's own performance history with a partners' performance (i.e., self-other mergence) occurs when cooperating and diminishes in competition (Wittmann et al., 2016).

As evidence of distance effects on other-other representations, a recent study mapped the social network of an MBA cohort and presented each student with pictures of close and distant others in their network while they were scanned using fMRI (Parkinson, Kleinbaum, & Wheatley, 2017). Comparing the similarity of the stimuli's distributed voxel patterns showed representational similarity was inversely related to social degree of separation. Students socially closer to the participant were represented more similarly than more distant affiliations. However, a different study that used neural similarity to distinguish between personally familiar individuals found that egocentric factors (e.g., similarity to self) only weakly predicted the neural similarity (Thornton & Mitchell, 2017). Instead, holistically rated interstimulus similarity and confidence in mentalizing the stimuli were better predictors of the representational structure, suggesting the effect of distance is less clear cut.

Beyond familiarity, representations also code for hierarchical or relative status of individuals in a group (Tavares et al., 2015). This study used an innovative "choose your own adventure" narrative task where participants made iterative choices that directly affected the power and affiliative relationship between the participant and multiple characters. The characters represented a variety of possible social stati such as an old high school friend or an owner of a company. Importantly, hippocampal activity was associated with changes of the characters on a vector index that tracked their position on power-affiliation axes. Relatedly, there is evidence that people encode the position of other individuals in social hierarchies that do not contain the participants (Kumaran, Banino, Blundell, Hassabis, & Dayan, 2016; Kumaran, Melo, & Duzel,

2012), where neural similarity between two individuals increases with greater co-occurrence (Dziura & Thompson, 2017).

Collectively these studies suggest that there's a great correspondence between social distance (e.g., familiarity, affiliation, power) and representational distance. Close social distance between oneself and others or between two or more others is encoded as overlapping mental representations. Conversely, the more distant or in competition one is, the more dissimilar the representations.

Social dimensions

The goal of an ontology is to map the interrelations between components. This section will attempt to do so by relating the effects of previously reviewed factors to current models that putatively represent fundamental dimensions of social cognition: face space (Oosterhof & Todorov, 2008), stereotype content model (SCM) (Fiske et al., 2007), ABC model (Koch, Imhoff, Dotsch, Unkelbach, & Alves, 2016), and the RSV model (Tamir et al., 2015). **Table 1** lists the hypothesized dimensions from the models, the data and stimuli used to reach those conclusions, and related factors. While it is likely that all the previously reviewed factors are involved in all of the models, I emphasize the more direct connections. As can be seen from **Table 1**, there are many domains in which these models operate, however the representational level can be considered common ground. This section will speculate what insights a representational similarity perspective could bear on these models.

Table 1. Models of social cognition.

Model	Social Target	Dimensions		Data Type	Relevant factors
Face Space	Faces	I.	Trustworthiness	Trait ratings	V, T
		II.	Dominance		

SCM	Groups	I. Warmth II. Competence	Trait ratings	SC, V, T
ABC	Groups	I. Agency II. Beliefs III. Communion	Trait or similarity ratings	SC, V, T, D
RSV	Mental states	I. Rationality II. Sociality III. Valence	Distributed neural activity	V, D

Note: SC = social category, V = valence, T = typicality, D = distance

Valence is the one theme shared across every model due to their specific dimensions that are proxies for or directly related to valence (e.g., trustworthiness, warmth, communion, valence). Many of these models used dimensional scaling or reduction or clustering techniques that are sensitive to variance in data. If the density hypothesis of valence is correct (Koch, Alves, et al., 2016), the asymmetry in the organization of positive and negative may be a reason these dimensions show up from the analyses. For example, in face space, a computer-generated face model of trustworthiness was created that showed an asymmetrical fit with trustworthiness ratings. Untrustworthy faces showed a linear reduction in their trustworthy ratings. However, there was a plateau in the trustworthy faces where more trustworthiness did not linearly increase trustworthiness ratings. Moreover, a larger range of trustworthy faces were perceived as displaying one emotion (happy) while untrustworthy faces were perceived to display more emotions (angry, neutral). One representational interpretation is that trustworthy faces are represented more similar to each other than the untrustworthy faces. Likewise, a glance at the well-known stereotype content figure where groups are plotted on the warmth-competence dimension, groups low in warmth are more spread out relative to those high in warmth. Likewise, in the ABC model, groups that are high in communion (i.e., warmth) tend to be plotted together in the center of the dimensional space with groups lower in communion more spread out in the outer rings. Interestingly, the RSV

model's valence component shows more spread out positive mental states than negative states, suggesting more work needs to be done on the contingencies of the effect of valence on representations.

The representational effect of social categories is most likely to affect the SCM and ABC models which focus on groups. Specifically, the homogenizing effect of ingroup as a category could be driving the placement of different groups on the respective model dimensions. In the SCM figure, the high warmth/high competence quadrant holds the most closely placed groups (e.g., housewives, Christians, Irish, middle class, Americans). Given that the sample was from the U.S., this clustering effect (and their positive placement) could be due to that being the sample's ingroup. The ABC model acknowledges this more explicitly by showing that the communion dimension (i.e., trusted groups) tend to be placed close to each other in the center of the plot and likely correspond to groups that the participants belong to. Notice that here the homogenizing effect of ingroup categorization cannot be disentangled from the clustering effect of positive valence, since ingroups are very much liked (Cikara & Van Bavel, 2014).

Typicality is most likely to impact the face space, SCM, and ABC models. As already discussed, shifts in face typicality shift which faces are considered trustworthy (Dotsch et al., 2016) and attractive (Potter & Corneille, 2008). The computer-generated faces were based on Caucasian European individuals that look more male. This means the specific face model instantiated in the paper could be determined by the demographic make-up of the sample whom may share or disagree about ideas of typicality. As such, a further step in validation will be to examine variation in computer-generated face spaces created by a variety of demographics. Typicality can also sharpen category boundaries, which means the spatial groupings in the SCM and ABC models may also be highly driven by idiosyncratic familiarity with those groups. The SCM uses clustering

techniques to create quadrants of high and low warmth and competence dimensions. If typicality sharpens category boundaries by pushing representations towards a prototype, the cluster boundaries in the SCM may also be affected. The ABC model acknowledges this in their communion dimension, but an additional effect may be how spread or compressed the whole dimensional space is. The typicality effect predicts that the more familiar all the groups are, more compressed that dimensional space will be and vice versa. Once again, this effect is highly congruent with category and valence effects.

The distance effect is most likely to impact the RSV and ABC models. The RSV model focuses on mental states, which unpublished work from Tamir and colleagues shows that similarity with others can expand their mental state space; states are represented very far from each other. Likewise, the mental state space is compressed for dissimilar others; states are represented very similarly. For the ABC model, social distance can be considered to lead to the same effect as typicality: more similar or typical others will homogenize and cluster groups together, possibly towards the center.

Overall, this section argues that the representational effects of categories, valence, typicality, and social distance are usually congruent with each other when considering their roles in models of social cognition. These models are created by using human evaluative data, which means there is room to examine sources of human variance within these models. In other words, category homogeneity is likely to affect people differently depending on the groups they are evaluating, shifting their placement of groups or faces on a dimensional model. Whether this variance is enough to completely change model conclusions or dimensions or whether this flexibility is simply noise around stable dimensions is an unanswered question. Another important source of variance concerns whether the congruency in the representational effect of the many

factors above is an unchanging feature in social cognition or if there are contexts in which they are at odds and, if so, which effect dominates. Finally, while I conceptualized this section as the effect that factors have on representations, the directionality is unclear. It could be the case that a tightly represented stimulus set is a neural signal that one is dealing with a positive/in group/typical/socially close stimulus, and likewise this signal could be recursively shaped by encountering those stimuli in the world.

Process of social representational spaces

The previous sections have described the content of social representations, however, it is also important to know how that information is stored in the first place. In this section, I make the case that memory integration and separation are mechanisms by which social representations are maintained and changed. First, I provide a very general overview of how memory integration and separation operate neurally and behaviorally followed by evidence that points to their involvement in structuring social memory. The neuromechanistic processes will focus on the temporal lobe (e.g., hippocampus) and frontal cortices (e.g., prefrontal cortex).

Memory integration and separation

Two of the main organizational memory schemes used by the brain are associative and schematic networks (Eichenbaum, 2016). In associative networks, memory for events or items that share overlapping features are linked together, supporting knowledge about direct pairs and transitive inferences about indirect pairs (Preston, Shrager, Dudukovic, & Gabrieli, 2004; Zeithamova, Schlichting, & Preston, 2012). These simpler associative networks can be combined to form schematic organizations (i.e., schemas) that can represent more complex hierarchical concepts (Collin, Milivojevic, & Doeller, 2015; Hasson, Chen, & Honey, 2015; Milivojevic, Vicente-Grabovetsky, & Doeller, 2015). The mechanisms that build these associative

organizations are memory integration and separation whereby novel information is inserted into preexisting schemas by hippocampal mechanisms that shift representations towards or away from each other (Schlichting & Frankland, 2017; Schlichting & Preston, 2015).

When encountering a novel situation, if it relates to previous memories, those memories can be reactivated and become labile and modifiable (Hupbach, Gomez, & Nadel, 2013). There are many factors that determine whether those reactivated memories will be integrated with or separated from the novel information: reactivation strength (Kuhl, Bainbridge, & Chun, 2012; Lewis-Peacock & Norman, 2014), prediction error (Gershman, Monfils, Norman, & Niv, 2017; Gershman, Radulescu, Norman, & Niv, 2014), repeated learning or experience (Favila et al., 2017, 2016; Schlichting, Mumford, & Preston, 2015), valence (Bisby & Burgess, 2017), time (Schapiro, Kustner, & Turk-Browne, 2012; Zeithamova & Preston, 2017), and reward contingencies (Wimmer & Shohamy, 2012). These effects are typically indexed by examining the change in interstimulus representational similarity before and after learning.

This fast-binding hippocampal system is thought to interact with a complimentary system in the neocortex that is attuned to abstracting generalities over time (Antony, Ferreira, Norman, & Wimber, 2017; McClelland, McNaughton, & O'Reilly, 1994). The duration of this learning process is debated, but it is associated with the creation of conceptual schemas (Tse et al., 2007). A related idea states that the hippocampus regularly interacts with the prefrontal cortex in a bidirectional manner to shape memory organization and retrieval (Eichenbaum, 2016, 2017; Preston & Eichenbaum, 2013). The idea is that the prefrontal cortex stores contextual rules that guide the retrieval of relevant and suppression of irrelevant memories in the hippocampus, while the hippocampus provides the prefrontal cortex with cues about the current context. The prefrontal cortex also has memory integration and separation functionality (Schlichting et al., 2015).

Importantly, the prefrontal cortex is not the only system that can build concepts, as the binding mechanisms within the hippocampus are also well suited for capturing commonalities in many domains (Mack, Love, & Preston, 2017).

To summarize, hippocampal mechanisms allow the flexible integration or separation of memory representations to update and form associative networks and concepts. They also interact with prefrontal cortex that builds schemas and guides the retrieval of contextually appropriate memories.

Memory mechanisms for social representations

Given the importance of schemas in social memory theorizing (Baldwin, 1992; Fiske & Linville, 1980; Sherman, 2001), the idea that the brain repurposes more basic cognitive mechanisms for social life (Parkinson & Wheatley, 2015), and the involvement of the hippocampus across many domains (Shohamy & Turk-Browne, 2013) including social cognition (Montagrin et al., 2017; Okuyama, 2017; Ross & Olson, 2010; Rubin, Watson, Duff, & Cohen, 2014), it is not surprising that memory integration and separation could play a role in social information processing. There are two lines of evidence for this idea: shared neural topography between memory systems and areas that represent social stimuli and direct examinations of how associative memory mechanisms process social stimuli.

Prefrontal Cortex

The prefrontal cortex, specifically the medial portion, is thought to be a center for schemas and contextual rules (Preston & Eichenbaum, 2013). Interestingly, it is also a hub for the representation of social information about the self and others (Wagner et al., 2012). Important social information is represented in this area: valence (Chavez & Heatherton, 2015; Koster-Hale et al., 2017), the popularity and connectedness of an individual in a network (Parkinson et al.,

2017), group distinctions (Lau & Cikara, 2017), and multimodal person impressions (Ferrari et al., 2016). Thus, the prefrontal cortex is a hub that both creates schemas and represents social information that could be considered schematic.

Temporal Lobe

The hippocampus, in the medial temporal lobe, is thought to be a binding hub that builds associative networks and concepts. In line with this idea, there is evidence that the hippocampus contains identity information about faces that is insensitive to facial orientation (Anzellotti, Fairhall, & Caramazza, 2014). Recent reviews have also identified the hippocampus as important for social memory due to its flexible relational encoding and ability to make cognitive maps, both important for social navigation (Laurita & Spreng, 2017; Montagrin et al., 2017; Rubin et al., 2014). However, the anterior temporal lobe is more associated with social cognition in the literature (Olson, McCoy, Klobusicky, & Ross, 2013; Ross & Olson, 2010). The anterior temporal lobe has memory integration functionality at the cellular level, it projects to hippocampal and prefrontal areas, and its responses to social information can be derived from associative learning models (Spiers, Love, Le Pelley, Gibb, & Murphy, 2016). In fact, recently a proposition was made that link the anterior temporal lobe, the amygdala, and the orbitofrontal cortex as a associative social learning network (Olson et al., 2013). While the evidence is still nascent, it seems that areas across the temporal lobe have the ability to encode and process social information due to their associative mechanisms.

Associative Social Processing

There have been calls to examine how associative memory systems process socio-relational information (Baldwin, 1992), which is now a trending area of research (Amodio & Ratner, 2011; FeldmanHall, Dunsmoor, Kroes, Lackovic, & Phelps, 2017; Gawronski & Bodenhausen, 2011;

Martinez, Mack, Gelman, & Preston, 2016; Murty, FeldmanHall, Hunter, Phelps, & Davachi, 2016; Rubin et al., 2014). Socio-relational information can occur at the level of the individual, such as associating good or bad behavior with a person (Murty et al., 2016), and at the inter-individual level, such as generalizing the bad reputation of an individual to their friend or a similar looking person (Martinez et al., 2016; Verosky & Todorov, 2013). Memory integration could underlie these effects and more, however neuroscientific research that could identify this specific mechanism is mostly circumstantial at the moment.

Some studies have shown that the medial prefrontal cortex is important for updating and representing important social information using decoding techniques or computational models. For example, medial prefrontal cortex encodes the state of a hierarchy that the participant belongs to and is learning about (Cikara & Gershman, 2016; Kumaran et al., 2016) and represents information about individuals' personalities (Hassabis et al., 2013). Likewise, the medial prefrontal cortex exhibits prediction errors when learning about individuals (Stanley, 2016) or groups (Spiers et al., 2016). While these studies point to integration as a mechanism since they involve making associations or being able to decode between categories, their analyses did not examine shifts at the item representational level and therefore cannot be considered conclusive evidence.

Similarly, there is evidence that the hippocampus is involved in learning of social hierarchies and status (Kumaran et al., 2016, 2012; Tavares et al., 2015), appropriate updating of moral impressions (Croft et al., 2010), detecting expectation violations (Chang & Sanfey, 2009). Further, the anterior temporal cortex is implicated in abstracting individual identities from multiple sources of biographical information (Wang et al., 2017). Interestingly, this study found that the hippocampus is not involved in the retrieval of this person knowledge. We can assume that integration or separation mechanisms were involved in these studies, but the analyses more so

focus on linking hippocampal function to indices of learning rather than directly examining the structure of stimulus representations.

There are three studies that do provide a direct link between affective or social stimuli and memory integration. The first study shows that when learning the structure of a social network, the co-occurrence of face pairs corresponds to their neural representational similarity in medial prefrontal cortex (Dziura & Thompson, 2017). The second provides a mechanistic hypothesis for why negative stimuli increase item memory but degrade information for the surrounding context (Bisby & Burgess, 2017). They attribute the improved item memory to the amygdala's upregulation of negative information and the decreased context memory is due to suppression of hippocampal associative encoding of item and context. This is supported by a study that presented participants with neutral or negative face-occupation pairs (e.g., Face A is a driver or murderer). In later memory tests, the faces were correctly identified in terms of valence, but memory performance for the specific occupation was impaired in the negative condition (Berkers, Klumpers, & Fernández, 2016). Importantly, this effect was associated with increased medial prefrontal-hippocampal coupling where the prefrontal cortex was more active while hippocampal activity was reduced, suggesting their interplay affects associative memory binding. The third study shows that neurons in the monkey anterior temporal cortex code specifically for the association between paired faces and patterns (Eifuku, Nakata, Sugimori, Ono, & Tamura, 2010). These studies provide direct evidence that associative memory can link or separate social representations.

To address a possible criticism, there are studies showing that the hippocampus is not necessary for learning individuals' valenced reputation (Todorov & Olson, 2008; Tranel & Damasio, 1993), suggesting hippocampal mechanisms may not be relevant for social

representation. There are two avenues by which this issue can be addressed: amygdala contributions to memory and fast mapping based on prior knowledge. The amygdala, a structure adjacent and linked to the hippocampus, has been shown to play a role in memory consolidation and the modulation of memory for valenced information (Bisby & Burgess, 2017; LaBar & Cabeza, 2006). Patients with hippocampal damage show enhanced memory for negative items suggesting the amygdala also functions as part of the memory system. It is also linked to the prefrontal cortex, which means it may have functional access to the schema-learning process. Likewise, hippocampal-damaged patients also show evidence of fast mapping (i.e., the rapid acquisition of novel information if it relates to prior knowledge) using neocortical areas (Sharon, Moscovitch, & Gilboa, 2011). These findings can contest the criticism if one considers valence as a well-learned schema that is represented in prefrontal areas (Chavez & Heatherton, 2015). Relatedly, some research says that the prefrontal cortex is not necessary for representing others (Frausin et al., 2016), yet the hippocampus can also build concepts (Mack et al., 2017). While the hippocampus or prefrontal cortex may not be individually necessary for social representation, they along with the amygdala or the anterior temporal lobe could compensate for whichever area is damaged. In other words, this argument for associative social processing is not necessarily a location-based one, instead it focuses on how associative memory mechanisms that are well studied in the hippocampus or prefrontal cortex may be widely distributed and employed to support social memory.

This section makes the case that integration and separation mechanisms in memory systems are fundamental for organizing and updating social representational spaces. It specifically situates dimensions important for person perception with mechanistic explanations from memory neuroscience. One interesting prospect from this framework is that the neocortical system may be

thought of holding the fundamental dimensions of social cognition as contextual rules or schemas, which guide and are shaped by incoming social information from the temporal lobe. This idea is in line with structural and temporal representation binding theory (Krueger, Barbey, & Grafman, 2009), which conceptualizes the prefrontal cortex as containing abstracted representations of repeatedly encountered social situations. These summary representations can interact with limbic areas to guide future social behavior and are akin to schemas. While evidence is converging on this point, it is still indirect since much of the research explicitly studying these memory mechanisms does not focus on social stimuli and the research that does focuses on representational brain mapping without attention to mechanism. This area requires a more explicit examination of integration and separation mechanisms on social representations and their organization.

Consequences of social representational spaces

Previous sections have outlined the type of information that shapes social representations and the associative memory mechanisms involved in creating them. This section will highlight some consequences that arise from the way that social information is represented, thus providing a look at psychological determinants of sociality. The four aspects of social life reviewed are: *transference, mentalizing, stereotyping, and relationships.*

Transference

Akin to evaluative conditioning where the valence of an item is transferred to a neutral item by mere association (Glaser et al., 2014; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010; Walther, 2002), transference is a likely property of memory representations that are clustered together. For example, highly entative groups lead to transference of individual group members' traits to other group members (Crawford et al., 2002). The same confusability occurs

for other social or demographic categories (Fiske, Haslam, & Fiske, 1991), highlighting the representational homogenizing effect of social categories.

Another instantiation of transference is overgeneralization. In one study, participants learned positive, negative, or mixed reputations for virtual characters in an economic game (Martinez et al., 2016). They then learned the association between these characters and novel friend faces. The next day participants played the game with the friends and cooperated more with friends associated with a positive character than friends associated with negative character, suggesting the reputation transferred from character to friend. A similar reputation transfer effect is seen between faces that physically resemble each other (Gawronski & Quinn, 2013; Verosky & Todorov, 2013). While not directly tested, these effects could be due to clustered or reactivated memory representations between associated individuals.

While transference is not guaranteed as there are contextual boundaries to the effect (Martinez & Jonas, n.d.), it is more likely to occur with socially significant affiliations (Andersen & Chen, 2002; Kraus & Chen, 2010; Martinez et al., 2016) highlighting the clustering effect of social distance.

Mentalizing

The structure of representations can also affect we understand and experience emotions. There is evidence that people build emotion mental models from the statistical regularities in others' emotion transitions (Thornton & Tamir, 2017). These mental models allow predictions of future mental states, for example, positive emotions are most likely to transition into each other suggesting they are highly linked in memory. Likewise, there is evidence that we use our own experience to understand others' mental states. This study had participants completed an emotion imagery task in the fMRI scanner where they imagined what they themselves and others would

feel in emotional situations (Oosterwijk, Snoek, Rotteveel, Barrett, & Scholte, 2017). They were able to train the classifier on the self-focused data that correctly classified how specifically they were thinking about others' mental states (e.g., why are they feeling this?, how are they expressing emotions?, what do they feel?), suggesting that representations about self states are used when mentalizing about others. These mental models could also underlie the intersubjectivity that is required for empathy and relating to others on a deeper level (Gallese, 2003) and to communicate effectively by sharing reality (Echterhoff, Higgins, & Levine, 2009).

Stereotyping

The content of stereotypes that implicate multiple categories are reflected in the neural similarity of those categories (Stolier & Freeman, 2016). These intersected representations can be considered as a prejudicial schema since they likely bias how one interprets and processes future social information, leading to a self-maintaining cognitive structure (Fyock & Stangor, 1994). The directionality is unclear: are these similarities byproducts of cognitive processed that then drive stereotype evaluations and behavior, are these similarities shaped by learned stereotypes, or are they recursively reinforced? However, they do predict prejudice between U.S. states (Koch, Kervyn, Kervyn, & Imhoff, 2017), bias how one categorizes a face (Dotsch, Wigboldus, & van Knippenberg, 2011), and correspond with semantic biases found in our text media (Bhatia, 2017).

Relationships

This section relies on the idea of neural synchrony where the time course of neural activity in two individuals is correlated (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004). This phenomenon is akin to the neural similarity between two stimuli, except it is between two individuals. This synchrony is thought to be an important determinant of social resonance or affiliations with others (Wheatley, Kang, Parkinson, & Looser, 2012; Wheatley & Sievers, 2016).

Strikingly, neural synchrony between individuals can predict friendships (Parkinson, 2016) and student engagement in a class (Dikker et al., 2017), suggesting that it captures the intersubjective nature of social interactions. While this research does not operate at the representational level, one obvious prediction to link the two is that individuals with greater synchrony will likely share the same representational space organization.

Conclusions and future directions

This review has attempted to provide an ontology of social representational spaces. It detailed how valence, social category, social distance, and typicality are important factors in social representational spaces that have effects on representational organization. Interestingly, they often have similar effects which have implications for dimensional models of social cognition. It has also argued for memory integration and separation found in memory systems as the mechanisms by which these spaces are formed and changed. Finally, it provided some evidence that the organization of social representations has consequences for social life, such as stereotyping and relationship building. The following are some unresolved issues that may be addressed in future research.

Task or context effects

How flexible or stable are social representational spaces? Several studies have shown that the representational similarity structure within the same stimulus set can change depending on learning, task demands, attentional focus of specific features, or evaluative context in which they are processed (Aly & Turk-Browne, 2016; Bracci, Daniels, & Op de Beeck, 2017; Mack, Love, & Preston, 2016; Nastase et al., 2016; Senoussi, 2013). Recent evidence suggests this also occurs with social stimuli (Kliemann, Jacoby, Anzellotti, & Saxe, 2016; Leshinskaya et al., 2017). For example, rating faces on their emotional expressions allowed a classifier to decode for valence,

but when evaluating the same faces on their age, valence could not be decoded anymore (Kliemann et al., 2016). Likewise, when evaluating social groups by their spirituality or political composition, the representational similarity between the groups reorganizes to reflect those evaluations (Leshinskaya et al., 2017). These studies show that the representational structure captured in a study may be idiosyncratic to task demands and thus more research is needed to understand the implications for research that attempts to find more robust organization of social knowledge.

Individual differences

How shared or idiosyncratic are social representational spaces? On one hand, individual differences in neuroscience has been a neglected source of inquiry as researchers have focused on average group or condition differences (Dubois & Adolphs, 2016). Evidence shows that the unique representational space of each individual is a better predictor of their judgments of stimulus similarity than other individuals' spaces (Charest, Kievit, Schmitz, Deca, & Kriegeskorte, 2014). On the other hand, it is worthwhile to know that despite unique life experiences and brain topology, there are commonalities in how individuals represent stimuli (Raizada & Connolly, 2012; Shinkareva, Malave, Just, & Mitchell, 2012). Knowing the extent of the two could provide avenues for interventions in social conflict as they would provide a mechanism for directly influencing shared reality between individuals.

Limits

What are the limits to social representational spaces? Dunbar's number hypothesized that given the size of the human neocortex our group sizes should be around 150 individuals (Dunbar, 1998), which has been validated using twitter network size (Gonçalves, Perra, & Vespignani, 2011). Moreover, human memory for social partners is bounded by capacity constraints (Stevens,

Volstorf, Schooler, & Rieskamp, 2010). An important direction for this area is how expansive representational spaces can be before they fail to be effective or reach capacity.

These three directions represent necessary inquiries into the nature of social representational spaces. Task or context effects will allow us to examine whether our findings should be considered as representing a truth about social representation or simply a unique task-determined pattern. They would additionally provide information on the contextual flexibility of memory systems (i.e., memory integration). Quantifying idiosyncrasy will provide a powerful avenue to understand how task effects interact with individual characteristics to produce these spaces. The extent to which unique task or individual patterns end up also being shared across individuals may tell us more about which dimensions are more robust, stable, and possibly universal. Finally, understanding limits is an important factor for measurement and theory. If we do not know the constraints, our theories may become detached from the biologically-informed functionality of these spaces.

References

- Alves, H., Koch, A., & Unkelbach, C. (2016). My friends are all alike - the relation between liking and perceived similarity in person perception. *Journal of Experimental Social Psychology, 62*, 103–117. <http://doi.org/10.1016/j.jesp.2015.10.011>
- Alves, H., Koch, A., & Unkelbach, C. (2017). The “common good” phenomenon: Why similarities are positive and differences are negative. *Journal of Experimental Psychology: General, 146*(4), 512–528. <http://doi.org/10.1037/xge0000276>
- Alves, H., Unkelbach, C., Burghardt, J., Koch, A., Krüger, T., & Becker, V. (2015). A density explanation of valence asymmetries in recognition memory. *Memory & Cognition, 43*(6), 896–909. <http://doi.org/10.3758/s13421-015-0515-5>
- Aly, M., & Turk-Browne, N. B. (2016). Attention Stabilizes Representations in the Human Hippocampus. *Cerebral Cortex, 26*(2), 783–796. <http://doi.org/10.1093/cercor/bhv041>
- Amodio, D. M., & Ratner, K. G. (2011). A Memory Systems Model of Implicit Social Cognition. *Current Directions in Psychological Science, 20*(3), 143–148. <http://doi.org/10.1177/0963721411408562>
- Andersen, S. M., & Chen, S. (2002). The Relational Self : An Interpersonal Social – Cognitive Theory, *109*(4), 619–645. <http://doi.org/10.1037//0033-295X.109.4.619>
- Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a Fast Route to Memory Consolidation. *Trends in Cognitive Sciences*. Elsevier Ltd. <http://doi.org/10.1016/j.tics.2017.05.001>
- Anzellotti, S., Fairhall, S. L., & Caramazza, A. (2014). Decoding representations of face identity that are tolerant to rotation. *Cerebral Cortex, 24*(8), 1988–1995. <http://doi.org/10.1093/cercor/bht046>

- Baldwin, M. W. (1992). Relational schemas and the processing of social information. *Psychological Bulletin*, *112*(3), 461–484. <http://doi.org/10.1037/0033-2909.112.3.461>
- Barclay, P. (2008). Enhanced recognition of defectors depends on their rarity. *Cognition*, *107*(3), 817–28. <http://doi.org/10.1016/j.cognition.2007.11.013>
- Barron, H. C., Garvert, M. M., & Behrens, T. E. J. (2016). Repetition suppression: a means to index neural representations using BOLD? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1705), 20150355. <http://doi.org/10.1098/rstb.2015.0355>
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. <http://doi.org/10.1037//1089-2680.5.4.323>
- Bell, R., & Buchner, A. (2012). How Adaptive Is Memory for Cheaters? *Current Directions in Psychological Science*, *21*(6), 403–408. <http://doi.org/10.1177/0963721412458525>
- Berkers, R. M. W. J., Klumpers, F., & Fernández, G. (2016). Medial prefrontal-hippocampal connectivity during emotional memory encoding predicts individual differences in the loss of associative memory specificity. *Neurobiology of Learning and Memory*, *134*, 44–54. <http://doi.org/10.1016/j.nlm.2016.01.016>
- Bhatia, S. (2017). The semantic representation of prejudice and stereotypes. *Cognition*, *164*, 46–60. <http://doi.org/10.1016/j.cognition.2017.03.016>
- Bisby, J., & Burgess, N. (2017). Differential effects of negative emotion on memory for items and associations, and their relationship to intrusive imagery. *Current Opinion in Behavioral Sciences*, *17*, 124–132. <http://doi.org/10.1016/j.cobeha.2017.07.012>
- Bracci, S., Daniels, N., & Op de Beeck, H. (2017). Task Context Overrides Object- and Category-Related Representational Content in the Human Parietal Cortex. *Cerebral Cortex*,

- (January), 310–321. <http://doi.org/10.1093/cercor/bhw419>
- Brosch, T., Bar-David, E., & Phelps, E. A. (2013). Implicit Race Bias Decreases the Similarity of Neural Representations of Black and White Faces. *Psychological Science*, *24*(2), 160–166. <http://doi.org/10.1177/0956797612451465>
- Chang, L. J., & Sanfey, A. G. (2009). Unforgettable ultimatums? Expectation violations promote enhanced social memory following economic bargaining. *Frontiers in Behavioral Neuroscience*, *3*(October), 36. <http://doi.org/10.3389/neuro.08.036.2009>
- Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., & Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences*, *111*(40), 14565–14570. <http://doi.org/10.1073/pnas.1402594111>
- Chavez, R. S., & Heatherton, T. F. (2015). Representational Similarity of Social and Valence Information in the Medial pFC. *Journal of Cognitive Neuroscience*, *27*(1), 72–82. <http://doi.org/10.1162/jocn>
- Chiao, J. Y. (2010). Neural basis of social status hierarchy across species. *Current Opinion in Neurobiology*, *20*(6), 803–9. <http://doi.org/10.1016/j.conb.2010.08.006>
- Cikara, M., & Gershman, S. J. (2016). Medial Prefrontal Cortex Updates Its Status. *Neuron*, *92*(5), 937–939. <http://doi.org/10.1016/j.neuron.2016.11.040>
- Cikara, M., & Van Bavel, J. (2014). The neuroscience of intergroup relations: An integrative review. *Perspectives on Psychological Science*, *9*(3), 254–274. <http://doi.org/10.1177/1745691614527464> pps.sagepub.com
- Cikara, M., Van Bavel, J. J., Ingbretsen, Z. A., & Lau, T. (2017). Decoding “us” and “them”: Neural representations of generalized group concepts. *Journal of Experimental Psychology:*

General, 14(5), 621–631.

Cohen, J. D., Daw, N., Engelhardt, B., Hasson, U., Li, K., Niv, Y., ... Willke, T. L. (2017).

Computational approaches to fMRI analysis. *Nature Neuroscience*, 20(3), 304–313.

<http://doi.org/10.1038/nn.4499>

Collin, S. H. P., Milivojevic, B., & Doeller, C. F. (2015). Memory hierarchies map onto the hippocampal long axis in humans. *Nature Neuroscience*, 18(11), 1562–1564.

<http://doi.org/10.1038/nn.4138>

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., ... Haxby,

J. V. (2012). The representation of biological classes in the human brain. *The Journal of Neuroscience*, 32(8), 2608–18. <http://doi.org/10.1523/JNEUROSCI.5547-11.2012>

Corneille, O., Huart, J., Becquart, E., & Brédart, S. (2004). When memory shifts toward more typical category exemplars: accentuation effects in the recollection of ethnically ambiguous faces. *Journal of Personality and Social Psychology*, 86(2), 236–250.

<http://doi.org/10.1037/0022-3514.86.2.236>

Crawford, M. T., Sherman, S. J., & Hamilton, D. L. (2002). Perceived entitativity, stereotype formation, and the interchangeability of group members. *Journal of Personality and Social Psychology*, 83(5), 1076–1094. <http://doi.org/10.1037/0022-3514.83.5.1076>

Croft, K. E., Duff, M. C., Kovach, C. K., Anderson, S. W., Adolphs, R., & Tranel, D. (2010). Detestable or marvelous? Neuroanatomical correlates of character judgments.

Neuropsychologia, 48(6), 1789–1801.

<http://doi.org/10.1016/j.neuropsychologia.2010.03.001>

Davis, T., & Poldrack, R. A. (2013). Measuring neural representations with fMRI: practices and pitfalls. *Annals of the New York Academy of Sciences*, 1296, 108–34.

<http://doi.org/10.1111/nyas.12156>

Davis, T., & Poldrack, R. A. (2014). Quantifying the internal structure of categories using a neural typicality measure. *Cerebral Cortex*, *24*(7), 1720–1737.

<http://doi.org/10.1093/cercor/bht014>

Diedrichsen, J., & Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS Computational Biology*, *13*(4), 1–33. <http://doi.org/10.1371/journal.pcbi.1005508>

Dikker, S., Wan, L., Davidesco, I., Bavel, J. J. Van, Ding, M., Poeppel, D., ... Rowland, J. (2017). Brain-to-Brain Synchrony Tracks Real-World Report Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Current Biology*, *27*(9), 1375–1380. <http://doi.org/10.1016/j.cub.2017.04.002>

Dorffner, G. (1999). The Connectionist Route to Embodiment and Dynamicism. In *Understanding Representation in the Cognitive Sciences*. (pp. 23–24).

Dotsch, R., Hassin, R. R., & Todorov, A. (2016). Statistical learning shapes face evaluation. *Nature Human Behaviour*, *1*(1), 1. <http://doi.org/10.1038/s41562-016-0001>

Dotsch, R., & Todorov, A. (2012). Reverse Correlating Social Face Perception. *Social Psychological and Personality Science*, *3*(5), 562–571.

<http://doi.org/10.1177/1948550611430272>

Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2011). Biased allocation of faces to social categories. *Journal of Personality and Social Psychology*, *100*, 999–1014.

<http://doi.org/10.1037/a0023026>

Dubois, J., & Adolphs, R. (2016). Building a Science of Individual Differences from fMRI. *Trends in Cognitive Sciences*, *20*(6), 425–443. <http://doi.org/10.1016/j.tics.2016.03.014>

- Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, 6(5), 178–190. [http://doi.org/10.1002/\(SICI\)1520-6505\(1998\)6:5<178::AID-EVAN5>3.3.CO;2-P](http://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.3.CO;2-P)
- Dziura, S. L., & Thompson, J. C. (2017). The neural representational space of social memory. *BiorXiv*, 1–36.
- Echterhoff, G., Higgins, E. T., & Levine, J. M. (2009). Shared reality: Experiencing commonality with others' inner states about the world. *Perspectives on Psychological Science*, 4(5), 496–521. <http://doi.org/10.1111/j.1745-6924.2009.01161.x>
- Edelman, S., Grill-Spector, K., Kushnir, T., & Malach, R. (1998). Toward a direct visualization of the internal shape representation space by fMRI. *Psychobiology*, 26(4), 309–321.
- Eichenbaum, H. (2016). Memory: Organization and Control. *Annual Review of Psychology*, (September), 1–27. <http://doi.org/10.1146/annurev-psych-010416-044131>
- Eichenbaum, H. (2017). Prefrontal – hippocampal interactions in episodic memory. *Nature Publishing Group*, 18(9), 547–558. <http://doi.org/10.1038/nrn.2017.74>
- Eifuku, S., Nakata, R., Sugimori, M., Ono, T., & Tamura, R. (2010). Neural correlates of associative face memory in the anterior inferior temporal cortex of monkeys. *J Neurosci*, 30(45), 15085–15096. <http://doi.org/10.1523/JNEUROSCI.0471-10.2010>
- Favila, S. E., Chanales, A. J. H., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature Communications*, 7, 11066. <http://doi.org/10.1038/ncomms11066>
- Favila, S. E., Kuhl, B. A., Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among Spatial Memories Triggers Repulsion of Hippocampal Representations
Article Overlap among Spatial Memories Triggers Repulsion of Hippocampal

Representations. *Current Biology*, 27(15), 2307–2317.e5.

<http://doi.org/10.1016/j.cub.2017.06.057>

FeldmanHall, O., Dunsmoor, J. E., Kroes, M. C. W., Lackovic, S., & Phelps, E. A. (2017).

Associative Learning of Social Value in Dynamic Groups.

<http://doi.org/10.1177/0956797617706394>

Ferrari, C., Lega, C., Vernice, M., Tamietto, M., Mende-Siedlecki, P., Vecchi, T., ... Cattaneo,

Z. (2016). The Dorsomedial Prefrontal Cortex Plays a Causal Role in Integrating Social Impressions from Faces and Verbal Descriptions. *Cerebral Cortex*, 26(1), 156–165.

<http://doi.org/10.1093/cercor/bhu186>

Fiske, A. P., Haslam, N., & Fiske, S. T. (1991). Confusing One Person With Another: What

Errors Reveal About the Elementary Forms of Social Relations. *Journal of Personality and Social Psychology*, 60(5), 656–674. <http://doi.org/10.1037/0022-3514.60.5.656>

Fiske, S. T. (1993). Social cognition and social perception. *Annual Review of Psychology*, 44,

155–194. <http://doi.org/10.1146/annurev.ps.44.020193.001103>

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition:

warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83.

<http://doi.org/10.1016/j.tics.2006.11.005>

Fiske, S. T., & Linville, P. W. (1980). What does the schema concept buy us? *Personality &*

Social Psychology Bulletin, 6(4), 543–557. <http://doi.org/0803973233>

Frausin, G., Hidalgo, A. D. F., Braga, R., Lima, S., Kinupp, V. F., Chau, L., ... Milliken, W.

(2016). Impaired personal trait knowledge, but spared other-person trait knowledge, in an individual with bilateral damage to the medial prefrontal cortex. *Neuropsychologia*.

<http://doi.org/10.1016/j.jep.2015.07.033>

- Fyock, J., & Stangor, C. (1994). The role of memory biases in stereotype maintenance. *The British Journal of Social Psychology / the British Psychological Society*, 33 (Pt 3), 331–343. <http://doi.org/DOI: 10.1111/j.2044-8309.1994.tb01029>
- Gallese, V. (2003). The roots of empathy: The shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology*, 36(4), 171–180. <http://doi.org/10.1159/000072786>
- Gawronski, B., & Bodenhausen, G. V. (2011). *The Associative – Propositional Evaluation Model: Theory, Evidence, and Open Questions. Advances in Experimental Social Psychology* (1st ed., Vol. 44). Elsevier Inc. <http://doi.org/10.1016/B978-0-12-385522-0.00002-0>
- Gawronski, B., & Quinn, K. A. (2013). Guilty by mere similarity: Assimilative effects of facial resemblance on automatic evaluation. *Journal of Experimental Social Psychology*, 49(1), 120–125. <http://doi.org/10.1016/j.jesp.2012.07.016>
- Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, 52(1), 45–56. <http://doi.org/10.1037//0003-066X.52.1.45>
- Gentner, D., & Smith, L. (2012). *Analogical Reasoning. Encyclopedia of Human Behavior* (2nd ed., Vol. 1). Elsevier Inc. <http://doi.org/10.1016/B978-0-12-375000-6.00022-7>
- Gershman, S. J., Monfils, M. H., Norman, K. A., & Niv, Y. (2017). The computational nature of memory modification. *eLife*, 6, 1–41. <http://doi.org/10.7554/eLife.23763>
- Gershman, S. J., Radulescu, A., Norman, K. a., & Niv, Y. (2014). Statistical Computations Underlying the Dynamics of Memory Updating. *PLoS Computational Biology*, 10(11). <http://doi.org/10.1371/journal.pcbi.1003939>
- Glaser, T., Dickel, N., Liersch, B., Rees, J., Süßenbach, P., & Böhner, G. (2014). Lateral attitude change. *Personality and Social Psychology Review*, 1–57.

<http://doi.org/10.1177/1088868314546489>

Gonçalves, B., Perra, N., & Vespignani, A. (2011). Modeling users' activity on twitter networks:

Validation of Dunbar's number. *PLoS ONE*, 6(8).

<http://doi.org/10.1371/journal.pone.0022656>

Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology*, 51(1), 79–88. <http://doi.org/10.1037/h0046219>

Han, S., Ma, Y., & Wang, G. (2016). Shared neural representations of self and conjugal family members in Chinese brain. *Culture and Brain*, 4(1), 72–86. <http://doi.org/10.1007/s40167-016-0036-5>

Hassabis, D., Spreng, R. N., Rusu, A. a, Robbins, C. a, Mar, R. a, & Schacter, D. L. (2013).

Imagine All the People: How the Brain Creates and Uses Personality Models to Predict Behavior. *Cerebral Cortex*, (August), 1979–1987. <http://doi.org/10.1093/cercor/bht042>

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences*, 19(6), 304–313. <http://doi.org/10.1016/j.tics.2015.04.006>

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject Synchronization of Cortical Activity During Natural Vision. *Science*, 303(MARCH), 1634–1640.

<http://doi.org/10.1126/science.1089506>

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology*, 37(1), 25–38.

<http://doi.org/10.1037//0022-3514.37.1.25>

Haxby, J. V, Connolly, A. C., & Guntupalli, J. S. (2014). Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annual Review of Neuroscience*.

<http://doi.org/10.1146/annurev-neuro-062012-170325>

Hearst, E. (1960). Simultaneous generalization gradients for appetitive and aversive behavior.

Science, *132*(3441), 1769–1770.

Hinzman, L., & Maddox, K. B. (2017). Conceptual and visual representations of racial

categories: Distinguishing subtypes from subgroups. *Journal of Experimental Social Psychology*, *70*, 95–109. <http://doi.org/10.1016/j.jesp.2016.12.012>

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: a meta-analysis. *Psychological Bulletin*, *136*(3), 390–421.

<http://doi.org/10.1037/a0018916>

Hugenberg, K., & Sacco, D. F. (2008). Social Categorization and Stereotyping: How Social

Categorization Biases Person Perception and Face Memory. *Social and Personality*

Psychology Compass, *2*(2), 1052–1072. <http://doi.org/10.1111/j.1751-9004.2008.00090.x>

Hupbach, A., Gomez, R., & Nadel, L. (2013). Episodic memory reconsolidation: An update.

Alberini, Cristina M [Ed] (2013) Memory Reconsolidation (Pp 233-247) Xiv, 320 Pp San Diego, CA, US: Elsevier Academic Press; US, 233–247.

Iordan, M. C., Greene, M. R., Beck, D. M., & Fei-Fei, L. (2016). Typicality sharpens category representations in object-selective cortex. *NeuroImage*, *134*, 170–179.

<http://doi.org/10.1016/j.neuroimage.2016.04.012>

Jacoby-Senghor, D. S., Sinclair, S., & Smith, C. T. (2015). When bias binds: Effect of implicit

outgroup bias on ingroup affiliation. *Journal of Personality and Social Psychology*, *109*(3), 415–433. <http://doi.org/10.1037/a0039513>

Johnson, A. L., & Queller, S. (2003). The mental representations of high and low entativity groups, *21*(2), 101–119.

- Kashima, Y., & Kerekes, A. R. Z. (1994). A distributed memory model of averaging phenomena in person impression formation. *Journal of Experimental Social Psychology, 30*, 407–455.
- Klapper, A., Dotsch, R., van Rooij, I., & Wigboldus, D. (2017). Four meanings of “categorization”: A conceptual analysis of research on person perception. *Social and Personality Psychology Compass, 11*(8), 1–16. <http://doi.org/10.1111/spc3.12336>
- Kliemann, D., Jacoby, N., Anzellotti, S., & Saxe, R. R. (2016). Decoding task and stimulus representations in face-responsive cortex. *Cognitive Neuropsychology, 32*94(December), 1–16. <http://doi.org/10.1080/02643294.2016.1256873>
- Kloth, N., Rhodes, G., & Schweinberger, S. R. (2017). Watching the brain recalibrate: Neural correlates of renormalization during face adaptation. *NeuroImage, 155*(October 2016), 1–9. <http://doi.org/10.1016/j.neuroimage.2017.04.049>
- Koch, A., Alves, H., Krüger, T., & Unkelbach, C. (2016). A General Valence Asymmetry in Similarity: Good Is More Alike Than Bad. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(8), 1171–1192. <http://doi.org/10.1037/xlm0000243>
- Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative–progressive beliefs, and communion. *Journal of Personality and Social Psychology, 110*(5), 675–709. <http://doi.org/10.1037/pspa0000046>
- Koch, A., Kervyn, N., Kervyn, M., & Imhoff, R. (2017). Studying the cognitive map of the U . S . states : Ideology and prosperity stereotypes predict interstate prejudice. *Social Psychological and Personality Science, (May)*.
- Koski, J. E., Xie, H., & Olson, I. R. (2015). Understanding social hierarchies: The neural and psychological foundations of status perception. *Social Neuroscience, 10*(5), 527–550.

<http://doi.org/10.1080/17470919.2015.1013223>

- Koster-Hale, J., Richardson, H., Velez, N., Asaba, M., Young, L., & Saxe, R. (2017). Mentalizing regions represent distributed, continuous, and abstract dimensions of others' beliefs. *NeuroImage*, *161*(March), 9–18. <http://doi.org/10.1016/j.neuroimage.2017.08.026>
- Koster-Hale, J., Saxe, R., Dungan, J., & Young, L. L. (2013). Decoding moral judgments from neural representations of intentions. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(14), 5648–53. <http://doi.org/10.1073/pnas.1207992110>
- Kraus, M. W., & Chen, S. (2010). Facial-feature resemblance elicits the transference effect. *Psychological Science*, *21*(4), 518–522. <http://doi.org/10.1177/0956797610364949>
- Kriegeskorte, N., & Kievit, R. a. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. <http://doi.org/10.1016/j.tics.2013.06.007>
- Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology*, *3*(JUL), 1–13. <http://doi.org/10.3389/fpsyg.2012.00245>
- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(November), 4. <http://doi.org/10.3389/neuro.06.004.2008>
- Krueger, F., Barbey, A. K., & Grafman, J. (2009). The medial prefrontal cortex mediates social event knowledge. *Trends in Cognitive Sciences*, *13*(3), 103–109. <http://doi.org/10.1016/j.tics.2008.12.005>
- Kuhl, B. A., Bainbridge, W. A., & Chun, M. M. (2012). Neural Reactivation Reveals Mechanisms for Updating Memory. *The Journal of Neuroscience*, *32*(10), 3453–3461.

<http://doi.org/10.1523/JNEUROSCI.5846-11.2012>

Kumaran, D., Banino, A., Blundell, C., Hassabis, D., & Dayan, P. (2016). Computations Underlying Social Hierarchy Learning: Distinct Neural Mechanisms for Updating and Representing Self-Relevant Information. *Neuron*, *92*(5), 1135–1147.

<http://doi.org/10.1016/j.neuron.2016.10.052>

Kumaran, D., Melo, H. L., & Duzel, E. (2012). The emergence and representation of knowledge about social and nonsocial hierarchies. *Neuron*, *76*(3), 653–66.

<http://doi.org/10.1016/j.neuron.2012.09.035>

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, *7*(1), 54–64. <http://doi.org/10.1038/nrn1825>

Lau, T., & Cikara, M. (2017). fMRI Repetition Suppression During Generalized Social Categorization. *Scientific Reports*, *7*(1), 4262. <http://doi.org/10.1038/s41598-017-04115-8>

Laurita, A. C., & Spreng, R. N. (2017). *The Hippocampus from Cells to Systems*. *The Hippocampus from Cells to Systems*. <http://doi.org/10.1007/978-3-319-50406-3>

Lei, R. F., & Bodenhausen, G. V. (2017). Racial Assumptions Color the Mental Representation of Social Class. *Frontiers in Psychology*, *8*(April), 1–7.

<http://doi.org/10.3389/fpsyg.2017.00519>

Leonardelli, G. J., Pickett, C. L., & Brewer, M. B. (2010). *Optimal Distinctiveness Theory. A Framework for Social Identity, Social Cognition, and Intergroup Relations*. *Advances in Experimental Social Psychology* (1st ed., Vol. 43). Elsevier Inc.

[http://doi.org/10.1016/S0065-2601\(10\)43002-6](http://doi.org/10.1016/S0065-2601(10)43002-6)

Leshinskaya, A., Contreras, J. M., Caramazza, A., & Mitchell, J. P. (2017). Neural Representations of Belief Concepts: A Representational Similarity Approach to Social

- Semantics. *Cerebral Cortex*, 1–14. <http://doi.org/10.1093/cercor/bhw401>
- Lewis-Peacock, J. A., & Norman, K. A. (2014). Competition between items in working memory leads to forgetting. *Nature Communications*, 5, 1–10. <http://doi.org/10.1038/ncomms6768>
- Liberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). The Origins of Social Categorization. *Trends in Cognitive Sciences*, xx, 1–13. <http://doi.org/10.1016/j.tics.2017.04.004>
- Lindström, B., Jangard, S., Selbing, I., & Olsson, A. (2017). The Role of a “Common Is Moral” Heuristic in the Stability and Change of Moral Norms. *Journal of Experimental Psychology: General*, 1–40. <http://doi.org/10.1037/xge0000365>
- Mack, M. L., Love, B. C., & Preston, A. R. (2016). Dynamic updating of hippocampal object representations reflects new conceptual knowledge. *bioRxiv*, 113(46), 1–21. <http://doi.org/10.1073/pnas.1614048113>
- Mack, M. L., Love, B. C., & Preston, A. R. (2017). Building concepts one episode at a time: The hippocampus and concept formation. *Neuroscience Letters*. <http://doi.org/10.1016/j.neulet.2017.07.061>
- Macrae, C. N., & Bodenhausen, G. V. (2001). Social cognition: categorical person perception. *British Journal of Psychology*, 92(Pt 1), 239–255. <http://doi.org/10.1348/000712601162059>
- Martinez, J. E., & Jonas, K. J. (n.d.). Transference in social judgments of sexual behavior and use of HIV pre-exposure prophylaxis.
- Martinez, J. E., Mack, M. L., Gelman, B. D., & Preston, A. R. (2016). Knowledge of social affiliations biases economic decisions. *PLOS ONE*, 11(7), e0159918. <http://doi.org/10.1371/journal.pone.0159918>
- McClelland, J. L., McNaughton, B. L., & O’Reilly, R. (1994). Why there are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and

- Failures of Connectionist Models of Learning and Memory, *102*(3), 419–457.
- McGarty, C. (1999). *Categorization in Social Psychology*. London: SAGE Publications Ltd.
- Mehta, P. H., & Prasad, S. (2015). The dual-hormone hypothesis : a brief review and future research agenda. *Current Opinion in Behavioral Sciences*, *3*, 163–168.
- Mende-Siedlecki, P., Baron, S. G., & Todorov, A. (2013). Diagnostic value underlies asymmetric updating of impressions in the morality and ability domains. *The Journal of Neuroscience*, *33*(50), 19406–15. <http://doi.org/10.1523/JNEUROSCI.2334-13.2013>
- Milivojevic, B., Vicente-Grabovetsky, A., & Doeller, C. F. (2015). Insight reconfigures hippocampal-prefrontal memories. *Current Biology*, *25*(7), 821–830. <http://doi.org/10.1016/j.cub.2015.01.033>
- Montagrin, A., Saiote, C., & Schiller, D. (2017). The Social Hippocampus. *Hippocampus*, 1–19.
- Morewedge, C. K., & Todorov, a. (2012). The Least Likely Act: Overweighting Atypical Past Behavior in Behavioral Predictions. *Social Psychological and Personality Science*, *3*(6), 760–766. <http://doi.org/10.1177/1948550611434784>
- Murty, V., FeldmanHall, O., Hunter, L., Phelps, E., & Davachi, L. (2016). Episodic memories predict adaptive value-based decision-making. *Journal of Experimental Psychology.*, *145*(5), 1–11. <http://doi.org/10.1037/xge0000158>
- Nastase, S. A., Connolly, A. C., Oosterhof, N. N., Halchenko, Y. O., Guntupalli, J. S., Visconti di Oleggio Castello, M., ... Haxby, J. V. (2016). Attention selectively reshapes the geometry of distributed semantic representation. *bioRxiv*, (603), 45252. <http://doi.org/10.1101/045252>
- Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–30.

<http://doi.org/10.1016/j.tics.2006.07.005>

O'Sullivan, C. S., & Durso, F. T. (1984). Effect of schema-incongruent information on memory for stereotypical attributes. *Journal of Personality and Social Psychology*, *47*(1), 55–70.

<http://doi.org/10.1037/0022-3514.47.1.55>

Okuyama, T. (2017). Social memory engram in the hippocampus. *Neuroscience Research*.

<http://doi.org/10.1016/j.neures.2017.05.007>

Olson, I. R., McCoy, D., Klobusicky, E., & Ross, L. A. (2013). Social cognition and the anterior temporal lobes: A review and theoretical framework. *Social Cognitive and Affective Neuroscience*, *8*(2), 123–133. <http://doi.org/10.1093/scan/nss119>

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–92.

<http://doi.org/10.1073/pnas.0805664105>

Oosterwijk, S., Snoek, L., Rotteveel, M., Barrett, L. F., & Scholte, H. S. (2017). Shared states : using MVPA to test neural overlap between self-focused emotion imagery and other-focused emotion understanding. *Social Cognitive and Affective Neuroscience*, *12*(7), 1025–

1035. <http://doi.org/10.1093/scan/nsx037>

Parkinson, C. (2016). Neural Homophily: Similar Neural Responses Predict Friendship, (310), 1–13.

Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2017). Spontaneous neural encoding of social network position. *Nature Human Behaviour*, *1*(April), 72. <http://doi.org/10.1038/s41562-017-0072>

Parkinson, C., Liu, S., & Wheatley, T. (2014). A common cortical metric for spatial, temporal, and social distance. *The Journal of Neuroscience*, *34*(5), 1979–87.

<http://doi.org/10.1523/JNEUROSCI.2159-13.2014>

Parkinson, C., & Wheatley, T. (2013). Old cortex, new contexts: re-purposing spatial perception for social cognition. *Frontiers in Human Neuroscience*, 7(October), 1–7.

<http://doi.org/10.3389/fnhum.2013.00645>

Parkinson, C., & Wheatley, T. (2015). The repurposed social brain. *Trends in Cognitive Sciences*, 19(3), 133–141. <http://doi.org/10.1016/j.tics.2015.01.003>

Parkinson, C., & Wheatley, T. (2016). Reason for optimism: How a shifting focus on neural population codes is moving cognitive neuroscience beyond phrenology. *The Behavioral and Brain Sciences*, 39(2013), e126. <http://doi.org/10.1017/S0140525X15001600>

Pattyn, S., Rosseel, Y., Van Overwalle, F., & Van Hiel, A. (2015). Social classification occurs at the subgroup level: Evidence based on a multidimensional scaling study. *Social Psychology*, 46(1), 13–23. <http://doi.org/10.1027/1864-9335/a000217>

Pereira Jr, A. (1999). Representation in Cognitive Neuroscience. *Understanding Representation in the Cognitive Sciences: Does Representation Need Reality?*, 49–56.

http://doi.org/10.1007/978-0-585-29605-0_6

Peschl, M. F., & Alexander, R. (1999). Does Representation Need Reality? Rethinking Epistemological Issues in the Light of recent Developments and Concepts in Cognitive Science. In *Understanding Representation in the Cognitive Sciences*. (pp. 9–17).

http://doi.org/10.1007/978-0-585-29605-0_1

Potter, T., & Corneille, O. (2008). Locating attractiveness in the face space: Faces are more attractive when closer to their group prototype. *Psychonomic Bulletin and Review*, 15(3), 615–622. <http://doi.org/10.3758/PBR.15.3.615>

Preston, A. R., & Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in

- memory. *Current Biology*, 23(17), R764–R773. <http://doi.org/10.1016/j.cub.2013.05.041>
- Preston, A. R., Shrager, Y., Dudukovic, N. M., & Gabrieli, J. D. E. (2004). Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus*, 14(2), 148–52. <http://doi.org/10.1002/hipo.20009>
- Quinn, K. a., & Rosenthal, H. E. S. (2012). Categorizing others and the self: How social memory structures guide social perception and behavior. *Learning and Motivation*, 43(4), 247–258. <http://doi.org/10.1016/j.lmot.2012.05.008>
- Raizada, R. D. S., & Connolly, A. C. (2012). What Makes Different People’s Representations Alike: Neural Similarity Space Solves the Problem of Across-subject fMRI Decoding. *Journal of Cognitive Neuroscience*, 24(4), 868–877. http://doi.org/10.1162/jocn_a_00189
- Ross, L. A., & Olson, I. R. (2010). Social cognition and the anterior temporal lobes. *NeuroImage*, 49(4), 3452–3462. <http://doi.org/10.1016/j.neuroimage.2009.11.012>
- Rubin, R. D., Watson, P. D., Duff, M. C., & Cohen, N. J. (2014). The role of the hippocampus in flexible cognition and social behavior. *Frontiers in Human Neuroscience*, 8(September), 1–15. <http://doi.org/10.3389/fnhum.2014.00742>
- Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, 22(17), 1622–7. <http://doi.org/10.1016/j.cub.2012.06.056>
- Scheutz, M. (1999). The ontological status of representations. In *Understanding Representation in the Cognitive Sciences*. (pp. 33–34).
- Schlichting, M. L., & Frankland, P. W. (2017). Memory allocation and integration in rodents and humans. *Current Opinion in Behavioral Sciences*, 17, 90–98. <http://doi.org/10.1016/j.cobeha.2017.07.013>

- Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, *6*, 8151. <http://doi.org/10.1038/ncomms9151>
- Schlichting, M. L., & Preston, A. R. (2015). Memory integration: neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*, *1*, 1–8. <http://doi.org/10.1016/j.cobeha.2014.07.005>
- Senoussi, M. (2013). Multivoxel Object Representations in Adult Human Visual Cortex Are Flexible: An Associative Learning Study. *Journal of Cognitive Neuroscience*, *26*(3), 194–198. <http://doi.org/10.1162/jocn>
- Sharon, T., Moscovitch, M., & Gilboa, A. (2011). Rapid neocortical acquisition of long-term arbitrary associations independent of the hippocampus. *Proceedings of the National Academy of Sciences*, *108*(3), 1146–1151. <http://doi.org/10.1073/pnas.1005238108>
- Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, *27*(2), 125–140. <http://doi.org/10.1007/BF02289630>
- Sherman, J. W. (2001). The dynamic relationship between stereotype efficiency and mental representation.
- Shinkareva, S. V., Malave, V. L., Just, M. A., & Mitchell, T. M. (2012). Exploring commonalities across participants in the neural representation of objects. *Human Brain Mapping*, *33*(6), 1375–1383. <http://doi.org/10.1002/hbm.21296>
- Shinkareva, S. V., Wang, J., & Wedell, D. H. (2013). Examining similarity structure: Multidimensional scaling and related approaches in neuroimaging. *Computational and Mathematical Methods in Medicine*, *2013*. <http://doi.org/10.1155/2013/796183>
- Shohamy, D., & Turk-Browne, N. B. (2013). Mechanisms for widespread hippocampal

- involvement in cognition. *Journal of Experimental Psychology. General*, 142(4), 1159–70.
<http://doi.org/10.1037/a0034461>
- Sofer, C., Dotsch, R., Oikawa, M., Oikawa, H., Wigboldus, D. H. J., & Todorov, A. (2017). For Your Local Eyes Only: Culture-Specific Face Typicality Influences Perceptions of Trustworthiness. *Perception*, 30100661769178. <http://doi.org/10.1177/0301006617691786>
- Sofer, C., Dotsch, R., Wigboldus, D. H. J., & Todorov, A. (2015). What is typical is good: the influence of face typicality on perceived trustworthiness. *Psychological Science*, 26(1), 39–47. <http://doi.org/10.1177/0956797614554955>
- Spiers, H. J., Love, B. C., Le Pelley, M. E., Gibb, C. E., & Murphy, R. a. (2016). Anterior Temporal Lobe Tracks the Formation of Prejudice. *Journal of Cognitive Neuroscience*.
<http://doi.org/10.1162/jocn>
- Stanley, D. A. (2016). Getting to know you: General and specific neural computations for learning about people. *Social Cognitive and Affective Neuroscience*, 11(4), 525–536.
<http://doi.org/10.1093/scan/nsv145>
- Stevens, J. R., Volstorf, J., Schooler, L. J., & Rieskamp, J. (2010). Forgetting constrains the emergence of cooperative decision strategies. *Frontiers in Psychology*, 1(January), 235.
<http://doi.org/10.3389/fpsyg.2010.00235>
- Stolier, R. M., & Freeman, J. B. (2016). Neural pattern similarity reveals the inherent intersection of social categories. *Nature Neuroscience*, 19(November 2015), 1–8.
<http://doi.org/10.1038/nn.4296>
- Tajfel, H. (1981). *Human groups and social categories. Human groups and social categories*.
<http://doi.org/10.1017/S0021932000023336>
- Tamir, D. I., Thornton, M. A., Contreras, J. M., & Mitchell, J. P. (2015). Neural evidence that

- three dimensions organize mental state representation: Rationality, social impact, and valence. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(1), 194–199. <http://doi.org/10.1073/pnas.1511905112>
- Tavares, R. M., Mendelsohn, A., Grossman, Y., Williams, C. H., Shapiro, M., Trope, Y., & Schiller, D. (2015). A Map for Social Navigation in the Human Brain. *Neuron*, *87*(1), 231–243. <http://doi.org/10.1016/j.neuron.2015.06.011>
- 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. <http://doi.org/10.1037//0022-3514.36.7.778>
- Thornton, M. A., & Mitchell, J. P. (2017). Consistent Neural Activity Patterns Represent Personally Familiar People. *Journal of Cognitive Neuroscience*, *29*(9), 1583–1594.
- Thornton, M. A., & Tamir, D. I. (2017). Mental models accurately predict emotion transitions. *Proceedings of the National Academy of Sciences*, *2017*. <http://doi.org/10.1073/pnas.1616056114>
- Todorov, A., & Olson, I. R. (2008). Robust learning of affective trait associations with faces when the hippocampus is damaged, but not when the amygdala and temporal pole are damaged. *Social Cognitive and Affective Neuroscience*, *3*(3), 195–203. <http://doi.org/10.1093/scan/nsn013>
- Tranel, D., & Damasio, A. R. (1993). The Covert Learning of Affective Valence Does Not Require Structures in Hippocampal System or Amygdala. *Journal of Cognitive Neuroscience*, *5*(1), 79–88. <http://doi.org/10.1162/jocn.1993.5.1.79>
- Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. a, Wood, E. R., ... Morris, R. G. M. (2007). Schemas and memory consolidation. *Science (New York, N.Y.)*, *316*(5821), 76–

82. <http://doi.org/10.1126/science.1135935>
- Tversky, A. (1977). Features of Similarity. *Psychological Review*, 84(4), 290–302.
<http://doi.org/10.1016/B978-1-4832-1446-7.50025-X>
- Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: the density hypothesis. *Journal of Personality and Social Psychology*, 95(1), 36–49. <http://doi.org/10.1037/0022-3514.95.1.36>
- Utevsky, A. V, & Platt, M. L. (2014). Status and the brain. *PLoS Biology*, 12(9), e1001941.
<http://doi.org/10.1371/journal.pbio.1001941>
- Verosky, S. C., & Todorov, A. (2013). When physical similarity matters: Mechanisms underlying affective learning generalization to the evaluation of novel faces. *Journal of Experimental Social Psychology*, 49(4), 661–669. <http://doi.org/10.1016/j.jesp.2013.02.004>
- Verosky, S. C., Todorov, A., & Turk-Browne, N. B. (2013). Representations of individuals in ventral temporal cortex defined by faces and biographies. *Neuropsychologia*, 51(11), 2100–2108. <http://doi.org/10.1016/j.neuropsychologia.2013.07.006>
- Wagner, D. D., Haxby, J. V., & Heatherton, T. F. (2012). The representation of self and person knowledge in the medial prefrontal cortex. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(4), 451–470. <http://doi.org/10.1002/wcs.1183>
- Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology*, 82(6), 919–934.
<http://doi.org/10.1037//0022-3514.82.6.919>
- Wang, G., Mao, L., Ma, Y., Yang, X., Cao, J., Liu, X., ... Han, S. (2012). Neural representations of close others in collectivistic brains. *Social Cognitive and Affective Neuroscience*, 7(2), 222–229. <http://doi.org/10.1093/scan/nsr002>

- Wang, Y., Collins, J. A., Koski, J., Nugiel, T., Metoki, A., & Olson, I. R. (2017). Dynamic neural architecture for social knowledge retrieval. *Proc*, 1–10.
<http://doi.org/10.1073/pnas.1621234114>
- Wasserman, E. A., Chakroff, A., Saxe, R., & Young, L. (2017). Illuminating the conceptual structure of the space of moral violations with searchlight representational similarity analysis. *NeuroImage*. <http://doi.org/10.1016/j.neuroimage.2017.07.043>
- Wheatley, T., Kang, O., Parkinson, C., & Looser, C. E. (2012). From Mind Perception to Mental Connection: Synchrony as a Mechanism for Social Understanding. *Social and Personality Psychology Compass*, 6(8), 589–606. <http://doi.org/10.1111/j.1751-9004.2012.00450.x>
- Wheatley, T., & Sievers, B. (2016). Toward a science of social resonance. In *Positive Neuroscience* (pp. 37–54). <http://doi.org/10.1093/acprof:oso/9780199977925.001.0001>
- Wimmer, G. E., & Shohamy, D. (2012). Preference by association: How memory mechanisms in the hippocampus bias decisions. *Science*, 338(6104), 270–273.
<http://doi.org/10.1126/science.1223252>
- Wittmann, M. K., Kolling, N., Faber, N. S., Scholl, J., Nelissen, N., & Rushworth, M. F. S. (2016). Self-Other Mergence in the Frontal Cortex during Cooperation and Competition. *Neuron*, 91(2), 482–493. <http://doi.org/10.1016/j.neuron.2016.06.022>
- Woll, S. B., & Craesser, A. C. (1982). Memory Discrimination for Information Typical or Atypical of Person Schemata. *Social Cognition*. <http://doi.org/10.1521/soco.1982.1.4.287>
- Wyer, R. S. (1980). The acquisition and use of social knowledge: basic postulates and representative research. *Personality & Social Psychology Bulletin*, 6(4), 558–573.
<http://doi.org/0803973233>
- Wyer, R. S., & Srull, T. K. (1989). *Memory and cognition in its social context*. Psychology

Press.

Xue, S. W., Weng, X. C., He, S., & Li, D. W. (2013). Similarity representation of pattern-information fMRI. *Chinese Science Bulletin*, 58(11), 1236–1242.

<http://doi.org/10.1007/s11434-013-5743-0>

Zeithamova, D., & Preston, A. R. (2017). Temporal Proximity Promotes Integration of Overlapping Events. *Journal of Cognitive Neuroscience*, 26(3), 1–13.

http://doi.org/10.1162/jocn_a_01116

Zeithamova, D., Schlichting, M. L., & Preston, A. R. (2012). The hippocampus and inferential reasoning: building memories to navigate future decisions. *Frontiers in Human Neuroscience*, 6, 70.

<http://doi.org/10.3389/fnhum.2012.00070>

Zink, C. F., Tong, Y., Chen, Q., Bassett, D. S., Stein, J. L., & Meyer-Lindenberg, A. (2008).

Know Your Place: Neural Processing of Social Hierarchy in Humans. *Neuron*, 58(2), 273–283. <http://doi.org/10.1016/j.neuron.2008.01.025>