Neuroscience Area – PhD course in

Cognitive Neuroscience

The Relevance of Affective Information

to the Semantic Representation

of Social Groups

Candidate:

Tiziano Suran

Advisor:

Prof. Raffaella Rumiati

Co-advisors:

Prof. Giorgia Silani

Prof. Wael El-Deredy

Academic Year 2018-19

INDEX

Abstract

The ability to readily access social group representations and automatically categorize others as members of social groups plays a central role in our social lives. This ability helps us guide our behavior by providing a rich set of information about unknown individuals based on the existing knowledge stored in our semantic system. Neuropsychological studies reporting the presence of patients with dementia disproportionally impaired at processing social group knowledge, with spared knowledge about other categories, together with patients presenting the opposite pattern, led to propose that social groups might well be represented separately from categories such as animals, plants or objects.

The organization of the semantic system in categories is consistent with several theories arguing that the relevance of the sensory and functional information, or features, forming a representation varies depending on the semantic category. The presence of a dissociation for social groups is however only accommodated by more recent theories that go beyond the simpler sensory/functional distinctions made by traditional models. Some of these theories suggest that for the representation of social groups a critical role is played by affective features. This hypothesis was corroborated by a study with brain tumor patients whose lexical semantic processing of social groups was affected by lesions of the left amygdala, insula and inferior frontal gyrus, all of which are also part of the brain network involved in processing affective information. However, beyond this finding, the empirical evidence in support of the affective features hypothesis put forward about social groups is to date relatively scarce. Thus, the aim of my dissertation is to provide the evidence of the greater relevance of affective features in social group representations by describing three original studies.

First, I have reviewed the main theories of semantic memory and affect, and then I reported three original studies. In Study 1, by applying transcranial magnetic stimulation on the left inferior frontal gyrus, previously linked to processing negative affect, I showed a link between processing social group knowledge and this area. Results showed an increase in the speed at which negative

1

social categories are categorized, without affecting responses to other stimuli. In Study 2, I reported that social category names are more susceptible to affective priming effects compared to nonsocial categories. The behavioral priming effect was also reflected in a late neural component measured via electroencephalography, where social group names displayed different amplitudes based on the affective congruence with the preceding prime. In Study 3 I documented that evaluative responses to social group names and pictures tend to be delayed compared to those towards nonsocial categories, despite no differences in categorical semantic access. Additionally, a multivariate pattern analysis (MVPA) of the neural correlates associated with this evaluation highlighted a better decoding of affective content for social groups in both an early, and a late time window across input modalities.

The findings reported in my thesis provide an affirmative answer to the hypothesized greater relevance of affective features in social group representations. I argue that such relevance is expressed in terms of an affect processing region contributing to their semantic categorization, a greater benefit from affective priming and an enhanced affective decoding. These results hopefully add valuable neuroscientific evidence about how social groups are represented, contributing towards the identification of their neural substrate and affective electrophysiological correlates in terms of response magnitude and temporal dynamics.

CHAPTER 1

Introduction

"Man is by nature a social animal; an individual who is unsocial naturally and not accidentally is either beneath our notice or more than human. Society is something that precedes the individual."

- Aristotle, Politics

In order to survive and thrive we live next to other people in a social environment. The observation made by Aristotle of society preceding the individual can indicate the great influence that a perceived social group has on what we think of and feel about other people. In just a fraction of a second, faces tend to automatically be categorized as members of a given social group, based on their gender, ethnicity, social status or occupation, a process that precedes the processing of the face's personal identity (Ito & Urland, 2003). In this instance, the 'society' preceding the individual is her/his social group, carrying along all the characteristics that are proper of that group, automatically attributed to the individual. For such categorization to occur, one needs to have a mental representation of the different categories present in our society, each with specific characterizing features. Moreover, this representation needs to be stored in the brain for later recollection, as indeed argued by several models of semantic memory described in the following paragraphs. The existence of a dedicated set of brain regions that support social group representations has recently been reinforced by the observation of patients with impaired social group representation following brain damage (Rumiati, Carnaghi, Improta, Diez, & Silveri, 2014). Building on extant theories suggesting that semantic representations are embodied by virtue of their reliance on sensory and functional processing modalities, the argument has been put forward of a special link between the representation of conspecifics and the processing of affective information. Accordingly, based on evidence showing the presence of damage

to brain regions supporting the processing of emotions, the resulting reduced ability to do so has been suggested to be the leading cause to an impoverished representation of social groups (Piretti et al., 2015).

Beyond correlational data, the hypothesis of a special link between social groups and affective processes has not found supporting or disproving evidence to date, especially as compared to the effort that has poured in research on other categories of knowledge. By presenting an overview of the main theories of semantic cognition and describing a series of original studies, the aim of my thesis is to provide a better understanding of (a) how our conceptual knowledge about categories of conspecifics is represented within the affective/semantic system, (b) how the affective information about social groups is processed by the brain, and (c) whether the representation of social groups possesses unique characteristics that makes it different from that of non-social categories (in the affective domain).

1. **Semantic memory and social cognition**

1.1 Semantic memory and semantic representations

As a component of long-term memory, semantic memory is described as the storage of our encyclopedic, conceptual knowledge. We may know what a pen is when we see one, but we do not necessarily know when and where we acquired such knowledge: while the former information is stored in the semantic memory, the latter information is stored by the episodic memory (Patterson, Nestor, & Rogers, 2007). When we think about a concept, our brain recreates a representation of that concept, which also leads to the renewal of the related multisensory perceptual experience in the absence of sensory stimulation. "Representation" will be thus intended here as the act by which consciousness, as a result of sparse neural activity (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006), reproduces an external object, or an internal object such as mood or a fantastic product, and also the content of that act of recreation, i.e., a simulation (M. Wilson, 2002). In the

4

human brain, the representation of the concept 'pen' will thus comprise the features that characterize it, which could be its shape, purpose and use. As pens come in different shapes and forms, according to the Prototype theory, the immediate mental representation we form after reading the word 'pen' would consist of the features that our idea of the standard, prototypical pen possesses (Fillmore, 1975). All these features will also be reciprocally linked to the entry in the mental lexicon for the word 'pen', the symbol used to refer to that concept (Binder, Desai, Graves, & Conant, 2009; Leshinskaya & Caramazza, 2016). The single symbols (i.e., words, or lexical items) linking to our conceptual knowledge are held by some theories to be organized in a lexical-semantic network made of nodes and relationships, allowing words to be also represented separately from the sensory and functional features of the concepts they refer to (e.g., Leshinskaya & Caramazza, 2016). The strength of the links between entries in the network is dictated by two factors. The first factor is their spatio-temporal contingency, or statistical co-occurrence (Simmons, Hamann, Harenski, Hu, & Barsalou, 2008; Vigliocco, Meteyard, Andrews, & Kousta, 2009), as in the word 'pen' often appearing next to the word 'paper', and the second factor consists of the similarities and overlap of their features (Farah & McClelland, 1991; Rogers et al., 2004), such as the color and use of the words 'paper' and 'board'. When it comes to the question as to how the brain represents concepts in the first place, a very popular view, common to several theories, argues that semantic knowledge is *grounded* into perception and action. Our conceptual understanding would be thus dependent on the way our body interacts with the environment and the brain areas processing that information, for which semantic knowledge would be itself 'embodied'. As will be described more in detail in the following sections, there is convergent evidence that shows how semantic representations follow a hierarchical organization which is partially embodied. As such, this organization contains both distributed modality- and attribute-specific systems and amodal single-locus systems that abstract knowledge from combinations of features (Leshinskaya & Caramazza, 2016; Patterson et al., 2007).

1.2   Evidence for category-specificity in semantic knowledge

The first cases of category-specific semantic impairments were described by Warrington and colleagues (Warrington, 1975; Warrington & Mccarthy, 1983; Warrington & Shallice, 1984). Some patients were disproportionally impaired at recognizing pictures representing animals and plants when presented with their names during an identification task; in contrast, they performed much better when they were presented with names of man-made objects. Later reports of patients presenting the reverse pattern, that is, a disproportionate recognition of objects compared to animals and plants (for reviews, see Capitani, Laiacona, Mahon, & Caramazza, 2003, and Thompson-Schill, 2003), led to suggest a living/non-living distinction in semantic representations. The correlation between the type of semantic deficit and lesioned brain areas (Warrington, 1975; Warrington & Shallice, 1984), with damage to visual regions and motor regions leading, respectively, to deficits for living and non-living things, was later also observed in healthy individuals using imaging techniques. Martin and colleagues, using positron emission tomography (PET), reported how naming animals activated brain regions associated with visual processing, whereas for tools led to a greater activation in areas associated with motor actions, such as the left premotor area (Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995; Martin, Wiggs, Ungerleider, & Haxby, 1996).

1.3   Category-specificity in social cognition

The idea of our knowledge about other people as also being represented by a distinct neural substrate was first suggested by neuropsychological studies reporting patients with an inability to name and recall more detailed information about otherwise familiar people as a consequence of brain damage (e.g., Miceli et al., 2000; Thompson et al., 2004). The category-specificity of conspecifics was made evident not only by patients with an impairment sparing knowledge about the semantic categories of objects or animals, but also by patients showing the reverse pattern i.e., an impairment for objects and animals with intact knowledge about people (Haslam & Sabah, 2013; Kay & Hanley, 2002;

Lyons, Kay, Hanley, & Haslam, 2006; Thompson et al., 2004). Rumiati and colleagues (Rumiati et al., 2014) later also described patients with specific impairments in their knowledge about social groups, revealing how impairments for conspecifics can also occur at a supraordinate, categorical level. In order to better understand what makes the representation of social categories unique, the following sections will provide a description of the different theories initially formulated to explain the category-specific deficits first reported only among non-social categories (for reviews, see Capitani, Laiacona, Mahon, & Caramazza, 2003, and Thompson-Schill, 2003), their evolution, and their explanation of the emergence of semantic deficits for conspecifics.

1.4 Theories of semantic memory

One of the first theoretical accounts on the emergence of the category-specific deficits is the *sensory/functional theory* (SFT) first proposed by Warrington and Shallice (1984). The theory holds that categorical deficits emerge because of damage to systems processing modality-specific information, either sensory or functional, in which living and non-living things, respectively, are represented. With this interpretation, Warrington and Shallice first introduce the seeds for the later theories of *embodied* cognition, whereby the contents of semantic cognition are based on perception and action (e.g., Barsalou, 2010), an organizational principle also found to emerge in computational models of the conceptual semantic system (Farah & McClelland, 1991).

A different theory, the *domain-specific hypothesis* (DSH; Caramazza & Shelton, 1998), suggests instead that distinct semantic categories are processed by specific brain networks developed as a result of evolutionary pressure. Although, as I will discuss later, the theory has since been updated (Mahon & Caramazza, 2009), in its original formulation, semantic categories emerged because their unique survival relevance related to nutrition, escaping from predators or reproduction. According to this interpretation, any relationship that a category has with a specific sensory or motor feature or brain region is only secondary to its pre-determined implementation. As the SFT, also the DSH

endorses to some extent the embodied view of semantic memory. Nevertheless, an emphasis is put on the greater independence of conceptual knowledge systems from sensory/motor ones.

1.4.1 *Beyond the animate/inanimate distinction*

Although sufficient in explaining the living/non-living dissociations initially reported, both the SFT and the DSH needed to be refined in order to accommodate later findings of patients with deficits in sub-categories within the living/non-living distinction. Within the living category, beyond the deficits for conspecifics (e.g., Miceli et al., 2000), single cases have been reported with selective deficits also for fruits/vegetables (Hart et al., 1983; Samson & Pillon, 2003) and for animals (Blundo et al., 2006; Caramazza & Shelton, 1998). Such findings led to both a reformulation of the existing theories, but also the formulation of new ones, each nowadays enjoying different degrees of support. I will here briefly describe some of the most accredited theories by dividing them according to the way in which embodiment is interpreted (Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012; Meteyard & Vigliocco, 2008). Based on how the interaction between semantic representations and perceptual/motor modalities is explained, theories either belong to the 'secondary activation' (or secondary embodiment) or to the 'mediation' (or weak embodiment) categories.

1.4.2 *Secondary activation theories*

Secondary activation theories argue that semantic representations are not dependent on sensory information and thus are amodal. The link to sensory-motor content is present in the form of a secondary activation deriving from the existent connections between an independent semantic system, and the sensory and motor systems. Damage to sensory/functional modules would thus result in impoverished representations with minimal semantic impairments (Meteyard et al., 2012).

As a first theory of the kind, the *hub-and-spokes* model (Lambon Ralph, Jefferies, Patterson, & Rogers, 2016; Noonan, Jefferies, Visser, & Lambon Ralph, 2013; Patterson, Nestor, & Rogers, 2007; Rogers et al., 2004) went beyond the initial binary sensory/functional distinction by holding that the process by which conceptual representations are formed consists of an interplay between a central amodal module (the *'hub'*) and several, modality-specific areas (*'spokes'*). Accordingly, the modality-specific information of the *spokes* would be abstracted away by the *hub*, rendering it purely symbolic. Whereas damage to the *hub*, located in the anterior temporal lobes (ATL), would result in semantic impairments across categories, as observed in semantic dementia patients (e.g., Jefferies, Patterson, Jones, & Lambon Ralph, 2009), damage to modality-specific areas would impair the processing of knowledge features related to sound, praxis, function, vision, valence, and mental lexicon. Thus, damage to the praxis *spoke* would in turn impair the knowledge of a specific semantic category mostly relying on that knowledge. For instance, lesions to the temporo-parietal area, involved in the processing of praxis-related information, have been found to lead to an impairment of the semantic representation of man-made tools, for which their most relevant semantic feature relates mainly to their use (e.g., Buxbaum & Saffran, 2002).

Another secondary activation theory, the *distributed domain-specific hypothesis* (distributed DSH; Mahon & Caramazza, 2009) holds that semantic content is distributed across several regions and interacts with sensory and motor information via non-arbitrary connections. Like the *hub-and-spokes* model, the distributed DSH also expands the number of properties relevant for conceptual knowledge to include sensory processes, motor action and emotional responses. As in the original formulation of the theory, an organization into domain-specific networks would be the result of evolutionary pressure, and all semantic knowledge would be represented in an amodal, abstract manner. The representation of each concept would produce an activation of sensory/motor information, but this would only be secondary to it. Thus, areas processing semantic information would only connect to sensory/functional brain regions, without overlapping.

1.4.3 *Mediation theories*

Unlike secondary activation theories, the key assumption of mediation theories is that sensory/motor information constitutes at least part of the semantic representations. Rather than being the result of a secondary activation, the activity in modality/specific cortical areas, and thus their content, plays a key role in conceptual representations. Furthermore, the integration of single features within a modality would take place nearby those areas, as for the combination of features across modalities, according to a principle of spatial proximity. The explanation of interaction effects between word processing and perception/action is to be found in a mediation process by areas adjacent to primary sensory and motor areas, and presenting reciprocal connections (Meteyard et al., 2012). In this way the activation of semantic content and the processing in primary areas are not completely independent, and the strength of the reciprocal influence depends on the characteristics of the connections, but also on the task demands, which could make some sensory/motor features more salient than others (Simmons & Barsalou, 2003).

One of such theories is the Convergence Zone (CZ) theory first proposed by Damasio (Damasio, 1989) and later extended by Simmons and Barsalou (Simmons & Barsalou, 2003). Accordingly, higher-order CZs present in association cortices respond to patterns of activation in adjacent sensory/motor cortices, which themselves respond to modality-specific features. Convergence zones would thus be part of a hierarchical neuronal organization, going from neurons responding to simple features (e.g., grey color) to combinations of features (e.g., visual input deriving from the sight of a grey dog, including its shape and motion), defined as *conjunctive neurons*. Unlike the original theorization of Damasio, which argued that CZs act as a bidirectional bridge that reinstates patterns of activity across the neuronal hierarchy, Simmons and Barsalou (Simmons & Barsalou, 2003) proposed CZs to have a representational capacity on their own. This means that damage to simpler CZs would not lead to impaired representations, as feature combinations are stored

in higher-order CZs. The authors also add to the theory a Similarity in Topography (SIT) principle, which predicts that the distance between two neurons in a convergence zone would be inversely proportional to the similarity of the features they associate. For instance, in a visual CZ, the conjunctive neurons responding to human faces would be located closer together to those that respond to faces of other mammals, as opposed to chairs.

The *grounding representation in perception, action and emotion* (GRAPES) model proposed by Martin (Martin, 2016) is another instance of mediation theories. The model extends the original *sensory/functional theory* and differentiates itself from the CZ theory by suggesting that semantic memory is organized according to the properties of concepts, rather than according to sensory modalities. In turn, the property acquisition modality would not define the location of its neural substrate, which would therefore be biologically pre-determined. This proposal was driven by studies on the similarities between typically developing individuals and individuals with congenital modality-specific deficits. For example, in congenitally blind individuals (Amedi, Raz, Azulay, Malach, & Zohary, 2010), the recognition of an object based on tactile information, deriving from its never-before-seen shape, would rely on the same areas recruited in the typically developing for processing shape information, and partially overlapping with visual shape processing areas in the ventral occipital stream (Amedi, Jacobson, Hendler, Malach, & Zohary, 2002). These areas store information about object properties and partially overlap with the sensory-motor regions allowing to experience them in a specific modality. Nevertheless, as it would be the case for hearing the sound of an engine or reading the word 'car', the information about the visual properties of a car can be readily accessed from different input modalities. Although the representation of a concept can involve several areas, each dealing with a different type of property, the relevance of a property in the representation varies from one category to another, and this relative distribution is also continuously updated through learning.

1.5   Social group representations and the affective features hypothesis

While the link between the category of tools and the processing of praxis has been extensively studied (e.g., Martin, 2016), it remains to be fully understood which kind of features are more relevant for processing social categories. In reporting cases of patients impaired in processing social groups, Rumiati and colleagues (Rumiati et al., 2014) highlighted the importance of distinguishing between person-specific and group categorical knowledge in the semantic domain. This distinction is particularly important when drawing conclusions on the processing of conspecifics based on comparisons with other categories of knowledge. Accordingly, any difference emerging from the comparison of person-specific knowledge with knowledge about animals and objects would be confounded by the required level of analysis. As an example, while Ronald Reagan is a unique individual with very specific characteristics and only one individual will be identified as such, all chairs of any shape and color will always be considered as chairs. While we can discuss what all chairs have in common, that is, that they are supposed to be sat on, we cannot talk in the same terms about Ronal Reagan for obvious reasons. On the contrary, despite differences in their individual identities, all policemen will be considered as having the common social function. The idea that these two levels of analysis also imply different neural substrates is also made evident by studies showing a greater efficiency of social categorical knowledge extraction from faces relative to person-specific knowledge (Cloutier, Mason, & Macrae, 2005; Mason & Macrae, 2004), and also by evidence of subordinate (instance-specific) knowledge being generally more impaired than that for super-ordinate (categorical) concepts in neurodegenerative diseases such as semantic dementia (Patterson et al., 2006; Rogers, Patterson, Jefferies, & Lambon Ralph, 2015).

The first reports of neurological patients with impairments in recognizing social groups (Rumiati et al., 2014) and the co-occurrence of lesions in brain areas commonly linked to affective processes, including the amygdala and the insula (Piretti et al., 2015), together with the presence of affective deficits (Carnaghi, Silveri, & Rumiati, 2015), led to suggest that social group representations

may rely to a greater extent on their affective features (Piretti et al., 2015). Unlike objects, people are agents, that is, they originate action that can directly affect the individual. Humans therefore focus on knowing other people's intentions and dispositions, whose expected consequences would serve as a useful source of information. Neuroimaging studies identified affect processing brain areas which selectively respond to social, as opposed to nonsocial stimuli, as well as areas, such as the amygdala, that show additive effects of a picture's emotionality and sociality (Norris, Chen, Zhu, Small, & Cacioppo, 2004).

The above theories also contain hints linking affective features to the representation of conspecifics. In the *hub-and-spokes* model, the authors suggested that impairments in the processing of conspecifics may derive from damage to cortical areas involved in the processing of affective features. This processing of valence, as they defined it, would be carried out by the orbitofrontal cortex, which is connected to subregions of the ATL *hub* via the uncinate fasciculus (Lambon Ralph et al., 2016). However, the authors also pointed to the lack of empirical evidence in understanding how affective knowledge interacts with knowledge about conspecifics.

In one of its latest formulations, the *distributed domain-specific hypothesis* also suggests the presence of an integration of affective information especially when representing conspecifics, which would be less evolutionarily relevant for tools. As one of the neural correlates of this integration process, the theory points to the fusiform face area, which would selectively respond to faces by virtue of its connectivity with several regions implied in extracting socially relevant information from stimuli, among which the amygdala (Mahon & Caramazza, 2011).

The link between affect and social cognition is also made explicit by other theories. For example, in their *conceptual topography theory*, Simmons and Barsalou also include an emotional processing modality (Simmons & Barsalou, 2003). Accordingly, emotional CZs would contain conjunctive neurons integrating information coming from different modalities which, based on previous studies, might be located in the orbitofrontal cortex and in the right somatosensory cortex.

While the former is hypothesized to integrate emotional and conceptual information (and is also highlighted by the *hub-and-spokes* model), the latter would bind emotional features from different sources, and has been indicated to be particularly relevant for decoding information from facial expressions and bodily postures.

Lastly, in his GRAPES theory, Martin proposes emotion to be a conceptual property relevant for social knowledge, and that such socio-emotional knowledge is processed by a circuitry of regions for processing biological motion (superior temporal sulcus) and emotion (amygdala). Nevertheless, the author also recognizes the need for further studies to clarify the roles of frontal and anterior temporal cortices in our social knowledge (Martin, 2016).

1.6   *Affective features hypothesis and abstract concepts*

Given the lack of causal evidence that our knowledge about social groups is characterized by a greater reliance on affective features, either by secondary activation or by mediation, my thesis will be dedicated to the investigation of how the processing of affective information interacts with semantic knowledge, and whether this interaction presents some unique properties when representing social groups.

Regardless of the semantic category, the contribution of affective information to conceptual knowledge has so far mostly been linked to the representation of abstract words. These words, such as "honor", as opposed to "knight", relate to concepts regarded as difficult, if not impossible, to perceive and act upon (Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011; Vigliocco, Meteyard, Andrews, & Kousta, 2009). As for the hypothesis of a greater reliance on affective features put forward about social categories, it has been suggested that abstract concepts may be grounded in affective states, or that, among the several features contributing to their representation, affect plays a predominant role (Vigliocco et al., 2014). While the reason for such grounding has been attributed to

the lack of external perceptual features of abstract concepts (Barsalou & Wiemer-Hastings, 2005), this explanation does not apply to social categories, or at least not to the same extent. Indeed, it would be hard to fully attribute an increased reliance on affective features to an increased abstractness of social groups. Although these concepts could be considered as relatively more abstract than object concepts, concepts such as 'policeman' do possess a combination of sensory correlates, including visual (e.g., wearing a uniform) and motor features (i.e., bipedal motion; Papeo, Wurm, Oosterhof, & Caramazza, 2017), that concepts like "grace" do not possess. For these reasons, the greater weight of affective features will mostly be attributed to the occurrence of exclusively social emotional processes like mentalization and empathy.

2. **Affective semantic knowledge**

2.1 *Affective features in semantic cognition*

From an evolutionary perspective, the ability to perform fast evaluations is of utter importance to respond quickly to stimuli in the environment. This automatic process allows us to determine whether whatever we encounter is potentially harmless, beneficial or dangerous (Cacioppo & Berntson, 1999). Thus, at the most basic level, evaluative processes produce a positive or negative affective response to determine whether something is good or bad for us, which not only guides our immediate behavior, but stores the affective association in memory for it to be retrieved in a second moment. The immediate response and its recollection from memory are not processed and experienced in the same manner. For example, eating a candy will be associated with a pleasant experience. The concept of a candy elicited by reading the word will also be linked to something pleasant. In both cases, positive valence information is being processed, nonetheless these take place at different levels of processing. In his hierarchical emotion theory, Panksepp (Panksepp, 1998, 2012) argues that there are three levels at which affective processing can occur. The first level comprises positive or negative response elicited by a direct experience with an object (e.g., eating a candy); the second level is related to

Pavlovian learning (e.g., anticipating the positive experience of eating a candy at the sight of the candy); the third-level processing would coincide with the long-term memory association, which can be recollected even in the absence of a stimulus (e.g., reading the word 'candy' and knowing that it is associated with a pleasant experience). Some authors make an even simpler distinction between the terms 'affective valence' and 'semantic valence', the former being the valence of a direct emotional response, and the latter referring to the semantic knowledge about valence (Itkes, Kimchi, Haj-Ali, Shapiro, & Kron, 2017). Valence is also addressed by several theories on emotion that consider affect as a core component from which discrete emotional experiences, like happiness, fear, and anger are derived (Kuhlmann, Hofmann, & Jacobs, 2017). In turn, even if emotions are more complex than affective reactions, on which they are built, the two terms are often used interchangeably. As in the present thesis this will also be the case, the terms will always refer to the basic positive/negative distinction just described.

When it comes to which brain regions support positive and negative valence processing, contrasting theories argued about the two valences as being processed by the same neural substrate (e.g., Carroll, Yik, Russell, & Barrett, 1999; Russell & Carroll, 1999), or about there being distinct neural substrates processing positive and negative affective information (e.g., Cacioppo, Gardner, & Berntson, 1999). More recent meta-analytic studies present support for both theories and suggest the presence of an interplay between valence-general and valence-specific regions (Cacioppo, Gardner, & Berntson, 1999; for a review see Lindquist et al., 2016). The presence of an even partially separable neural network for affective processing makes the study of affective processing increasingly complex. The possibility of positive and negative valence being processed separately also imply the possibility that the same stimulus can be associated with both positive and negative affect, whose immediate relevance is context and goal-dependent (Man, Nohlen, Melo, & Cunningham, 2017). An example of this can be the concept of 'exercising', linked to both fatigue and the reward of keeping fit. Even with separate representations of positive and affect for the same concept, it has been suggested that the

more dominant association will influence early, automatic evaluative processes regardless of task demands, whereas later processing may involve adjusting the initial affective association with a conflicting one, and thus elicit greater reflective processing driven by contextual information and task demands (Cunningham, Raye, & Johnson, 2004; Cunningham & Zelazo, 2007). As affective responses are highly subjective, if different brain areas are involved in the processing of negative and positive affect, this means that the processing of the same stimulus by two individuals may not always involve the same neural substrate, a factor that needs to be taken into account.

Either in Panksepp's third-level processing, in the concept of 'affective valence', affective theories converge with the previously outlined semantic theories suggesting valence (or, more generally, affect and emotion) to constitute a distinct feature of semantic representations (e.g., Lambon Ralph et al., 2016; Mahon & Caramazza, 2011; Martin, 2016; Simmons & Barsalou, 2003). Although related to semantic knowledge, affective knowledge has been shown to possess an at least a partially distinct neural substrate, leading to different behavioral outcomes. For instance, whereas semantic information (i.e., nonaffective) would drive instrumental goal-directed behaviors in frontal brain regions, affective information has been argued to directly drive behaviors of approach and avoidance with the contribution of the amygdala (Amodio & Ratner, 2011). Other studies also show how the areas involved in affective and non-affective processing in general are highly anatomically interconnected and communicating reciprocally (Ghashghaei & Barbas, 2002), which allows for a study of their interaction and their interdependence. If we think of features as the building blocks of semantic representations, it may seem straightforward to assume that their perception precedes the full conscious conceptual representation. Identifying an object involves first processing its single physical properties, like its color and shape. However, affective features can only be extracted after the perceptual encoding, and they will differ from individual to individual. One can then ask whether it is possible to discern between valence as being stored in memory, rather than being derived from a post-hoc evaluation (Lebrecht, Bar, Barrett, & Tarr, 2012), even though a post-hoc evaluation may

also be based on a memory storage, although accessed in a more controlled manner (Cunningham & Zelazo, 2007).

2.2   *Affective features in priming studies*

A first empirical indication of how affect can be conceived as a conceptual property is given by studies showing the influence of pre-activating affective valence processing on semantic access. As previously stated, the strength of the associations within the semantic network is also determined by the features shared by two concepts. Semantic priming paradigms have been extensively used to measure these associations also by means of spreading of activation, which could be derived from the processing of common specific features (e.g., the word 'strawberry' preceding the word 'cherry, both referring to something red, sweet and edible; Collins & Loftus, 1975; Neely, 2012). Affective priming, as a special case of semantic priming, is considered to be a demonstration that affective features are also representational constituents (Storbeck & Robinson, 2004). As such, these features were absent in initial semantic representation theories (e.g., Warrington & Shallice, 1984). Affective priming refers to the process by which the representation of a concept is facilitated by the prior processing of affectively congruent information, like a word, which is considered to activate a valence-processing node in the network (De Houwer, Hermans, Rothermund, & Wentura, 2002; Fazio, Jackson, Dunton, & Williams, 1995; Fazio, Sanbonmatsu, Powell, & Kardes, 1986). For example, people are on average faster at identifying a positive word such as 'love' if this was immediately preceded a positive word like 'plush', as opposed to a more negative word as 'knife'. Moreover, this facilitation, only obtained with shorter prime-target delays (for a review, see Klauer & Musch, 2003), is considered to be independent of any non-affective association between the prime words and the targets. Rather, it is thought to depend on automatic processing of the congruent affective features that the prime concept activates (De Houwer et al., 2002; Spruyt, Houwer, & Hermans, 2009). These affective prime-target relationships can in some cases have an even greater

influence compared to nonaffective relationships. This is the case of when affective features are made salient via task demands, like explicitly asking to focus on the affective content of a concept (e.g., Bargh, Chaiken, Govender, & Pratto, 1992; Bargh, Chaiken, Raymond, & Hymes, 1996; Klinger, Burton, & Pitts, 2000; Storbeck & Robinson, 2004).

2.3   *The neural correlates of affective features processing*

At the neurophysiological level, event-related potential (ERP) studies indicate that the mechanisms behind affective priming are similar to those of semantic priming, as suggested by the elicitation of a common N400 effect (Yao & Wang, 2014; but see also Herring, Taylor, White, & Crites, 2011). If we consider affective information as a type of semantic feature, this should not be surprising. Nevertheless, neuroimaging studies show the presence of both common and distinct networks implementing the two processes. In a study by Liu and colleagues (H. Liu, Hu, Peng, Yang, & Li, 2010), the left middle frontal gyrus (MTG)/superior temporal gyrus (STG) have been found to be active in both types of priming, activity in the left inferior frontal gyrus (IFG) and in the STG was specific to semantic priming, and the left fusiform gyrus (FG) and insula were involved in affective priming. The common activation in the temporal cortex is in line with semantic theories indicating the area as supporting amodal conceptual processing (Lambon Ralph et al., 2016), while the left IFG has been repeatedly shown to be involved in semantic memory retrieval and selection (Moss et al., 2005). The FG also finds an association with semantic processes in other studies, both for positive and negative affective information (Luo et al., 2004), whereas the insula is more commonly associated with subjective emotional responses (Amodio, 2014), but presents variations when it comes to the valence being processed, with the left part responding more strongly to negative than to positive stimuli (Lindquist et al., 2016).

Several of the areas involved in affective and nonaffective semantic priming are also active when processing emotional information in social stimuli. For example, activity in the amygdala has

also been associated with both positive and negative social clues, like those concerning social group membership of individual faces, and has thus been proposed to support the processing of social threat and social reward (Amodio, 2014). The anterior insula is found to support the subjective experience of negative affect also when processing social stimuli, while the area involved in mentalizing and perspective taking, the medial prefrontal cortex (mPFC) more broadly, may be engaged more strongly towards ingroup (positive valence) than outgroup (negative valence) members (Gilbert, Swencionis, & Amodio, 2012). This latter region, receiving projections from the amygdala and insula, is also thought to support the integration of general affective processes with processes such as mentalizing and empathy. Such use of affective information for processes which are unique to social encounters, like mentalizing and empathy, does provide an initial indication of how affective information related to social stimuli may undergo distinct processes which do not take place when representing nonsocial concepts.

2.4   *Affective features as attitudes*

In absence of person-specific information, when it comes to evaluative judgments of other individuals, the main driving factor stems from the affective knowledge we have about the social group we associate that individual with. In the social psychological literature, this knowledge is defined as an attitude, and it is conceptualized as a subjective evaluation represented in memory that determines our positive or negative disposition towards others (Fazio et al., 1995, 1986; Greenwald & Banaji, 1995). The combined effects of categorization and evaluation makes it possible for attitudes to influence novel social encounters, as how people feel about a social group is generalized to all individuals identified as its members. In turn, it could be argued that the most salient valence information about individuals comes from their social groups. This affective trace in semantic memory does not necessarily require a prior interaction with the members of a social group. The information simply acquired from our social environment (Hilton & von Hippel, 1996), from our

parents (Castelli, Zogmaister, & Tomelleri, 2009) and through mass media (Tukachinsky, Mastro, & Yarchi, 2015) is enough for attitudes to be created. Based on the data conveyed by different sources, we start to link a social group with an affective evaluation. But whereas in social psychology the main 'attitude objects' of interest are represented by people, they can also consist of any non-social entity, such as places, objects, or even abstract concepts, to which several of the same processes apply. The unanswered question is the extent to which the attitude towards different categories of concepts, as part of our knowledge about them, contributes to their representation, and whether such affective knowledge plays a special role when social groups are compared with nonsocial categories.

3. **Introduction to the studies**

In the following chapters I will present three experiments designed to study how we process social and non-social categories and the role of affective semantic information. To achieve this aim, I will employ transcranial brain stimulation and neurophysiological recordings during semantic categorization and evaluation tasks. In particular, I will compare social groups and objects of different types, both on the positive and the negative side of the valence spectrum. As the extent of embodiment in semantics is still under debate, I will attempt to accommodate my findings within the multiple semantic models that argue for an affective processing modality. As this modality represents a meeting point between semantic and affective theories, its mechanisms are expected to follow those of what the latter theories described as semantic affect. The affective semantic component will thus be treated as having two dimensions, a positive and a negative one, each represented in distinct valence-specific areas at the level of the cortex (Lindquist et al., 2016).

In Study 1 I will focus on the contribution of a frontal cortical area - the left inferior frontal gyrus - found to be involved in processing negative affective responses (Kensinger & Schacter, 2006) and damaged in patients showing semantic deficits for social groups (Piretti et al., 2015). The aim of the study was to assess whether the stimulation of this area would affect to a greater extent the

categorization of names of social groups which are generally evaluated as more negative in comparison to more positively evaluated social groups, and to non-social categories overall. In Study 2, while recording the corresponding neurophysiological signatures via electroencephalography (EEG), I will require participants to focus on the affective content of the same types of stimuli in order to provide an evaluative response during an affective priming paradigm. Here too, the comparison between social and non-social categories will involve the affective dimension, evaluating whether the brain response to positive and negative information changes based on the semantic category from which it is derived and on the affective congruency with prime words. In Study 3 I will apply multivariate pattern analyses (MVPA) on EEG data to measure the temporal dynamics of the brain in processing the affective content of social and non-social stimuli. In this study I will manipulate the modality of presentation of the stimuli, allowing to measure the activity of a cross-modal neural substrate for processing affective knowledge of social groups.

CHAPTER 2

Study 1: The Contribution of the Left Inferior Frontal Gyrus in Affective Processing of Social Groups

Abstract

We investigated the contribution of the *pars opercularis* of the left inferior frontal gyrus (LIFGop) in representing knowledge about social groups. We asked healthy individuals to categorize words preceded by semantically congruent or incongruent primes while stimulating the LIFGop. Previous studies showing an involvement of the LIFGop both in processing social stimuli and negative valence words led us to predict that its stimulation would affect responses to negative social category words. Compared to the Vertex as control site, the stimulation of the LIFGop increased the speed of categorization of negative social groups, and disrupted the semantic priming effect for negative words overall. Within the framework of recent theories of semantic memory, we argue that the present results provide initial evidence of the representation of social groups being characterized by affective properties, whose processing is supported by the LIFGop.

1. Introduction

The study of neurological patients' cognitive and affective abilities allows highlighting the underlying neural systems. Thus, brain damage has been shown to selectively and severely impair the ability of

individuals to name and recall detailed information about otherwise familiar people, leaving their semantic knowledge about objects or animals intact (e.g., Miceli et al., 2000; Thompson et al., 2004); or it can produce the reverse pattern, that is, impaired recognition of objects and animals with intact knowledge about people (Haslam & Sabah, 2013; Kay & Hanley, 2002; Lyons, Kay, Hanley, & Haslam, 2006; Thompson et al., 2004). These dissociations argue in favor of distinct brain systems for processing conspecifics and other non-social stimuli. However, the theories formulated to explain the category-specific deficits initially reported only among non-social categories have been a little silent on how concepts about conspecifics might be represented (for reviews, see Capitani, Laiacona, Mahon, & Caramazza, 2003; Caramazza, Anzellotti, Strnad, & Lingnau, 2014; Thompson-Schill, 2003).

Extant neuropsychological research has shown that patients' failure to name/recognize famous people such as Ronald Reagan (Miceli et al., 2000; Thompson et al., 2004) is selectively limited to the individual identities, while their ability to recognize non-social entities remains intact. Nevertheless, the level of the information required for the successful recognition of individual conspecifics might be intrinsically different from the one required for recognizing non-social entities. Recently, Rumiati, Piretti and colleagues reported that some patients were selectively impaired at recognizing conspecifics when presented as social groups (e.g., *doctors*; Piretti et al., 2015; Rumiati, Carnaghi, Improta, Diez, & Silveri, 2014). This deficit is argued to originate at the semantic level as it has been observed when either sorting names of social groups (Rumiati et al., 2014) or naming images depicting them (Piretti et al., 2015). In the latter study, using voxel-based lesion-symptom mapping (VLSM), Piretti et al. (2015) documented tumor patients with lesions in regions associated also with affective processes such as insula, amygdala, basal ganglia, and inferior frontal gyrus, who were impaired at naming photographs of social groups. This association between damage to affective processing areas and recognition deficits involving social groups led us to hypothesize that successful recognition of such social categories may depend on an intact ability to process affective information.

As the recognition of living and non-living things primarily weights on sensory and functional properties respectively, we hypothesized that recognition of social groups might require a privileged access to affective properties. Indeed, compared with patients without the deficit and healthy controls, patients with dementia and semantic deficits for social groups expressed less extreme affective evaluations for both positive and negative social groups, an effect that is known as *emotional blunting* (Carnaghi et al., 2015).

From a theoretical perspective, Barsalou and colleagues proposed that, in addition to the other sensory modalities, a model of conceptual knowledge should also include an emotion modality that is relevant for processing the affective properties of conceptual representations (Simmons and Barsalou, 2003; Barsalou, 2008). In their view, the internal affective states represent a component of knowledge that is as important as the external perceptual experience processed by other modalities (Barsalou, 2008). Thus, the patients' deficit at recognizing social groups (Rumiati et al., 2014; Piretti et al., 2015) can be interpreted as being due to degraded affective properties that play a prominent role in their representation. Martin (2016) too suggested that the affective properties are stored in emotional systems, with affective information being a salient component of social concepts. An emotion modality is also present in the *hub-and-spokes* model proposed by Lambon Ralph and collaborators (Lambon Ralph, Jefferies, Patterson, & Rogers, 2016; Noonan, Jefferies, Visser, & Lambon Ralph, 2013; Patterson, Nestor, & Rogers, 2007; Rogers et al., 2004) whereby the conceptual representations result from integrating several, modality-specific areas or *spokes* – that process information related to sound, praxis, function, vision, valence, and verbal description – which in turn feed a central modality-independent module (the *'hub'*). Interestingly, the model predicts that impairments involving the representation of conspecifics may derive from damage to cortical areas that process valence (Lambon Ralph et al., 2016).

As to the neural correlates, the *pars opercularis* of the left inferior frontal gyrus (LIFGop) singled out by Piretti et al (2015) is one of the candidates in which the affective processing of social

groups may take place. This area has also been found across several neuroimaging studies. First, the LIFGop is specifically involved in the semantic representation of familiar people when individual identities are categorized based on their profession at the level of social group (Chedid et al., 2016). Additionally, the LIFGop is one of the joint regions being active when healthy subjects read sentences describing behaviors of both individuals and social groups (Van der Cruyssen, Heleven, Ma, Vandekerckhove, & Van Overwalle, 2014).

As to the involvement in the processing of affective information related to conspecifics, the LIFGop has been associated with both emotional empathic responses (Shamay-Tsoory, Aharon-Peretz, & Perry, 2009), and in processing emotional facial expressions (Decety & Chaminade, 2003). Moreover, as the LIFGop is also part of Broca's area, its damage has also been shown to impair linguistic processes linked to affect. Thus damage to this region impairs the retrieval of the names of the emotions conveyed by facial expressions (Adolphs, Damasio, Tranel, Cooper, & Damasio, 2000), while neuroimaging techniques repeatedly highlighted a link between LIFGop and the affective processing of words with a negative valence. For instance, the cortical activity in this area increases more when participants judge the animacy of words with negative valence than positive valence (Kensinger & Schacter, 2006; but see also Leclerc & Kensinger, 2011). Additionally, a greater LIFGop activation to negative than neutral or positive words has been recorded in working memory tasks with emotional word distractors (García-Pacios, Garcés, del Río, & Maestú, 2017). Last, the area is increasingly activated when subjects read sentences about negative social groups or an individual with a negative trait performing a negative behavior, compared to when positive behaviors are performed by positive social groups or individuals (Van der Cruyssen et al., 2014). In sum, the reviewed evidence suggests a key role of the LIFGop in both lexical and affective processing involving conspecifics. Based on the hypothesis that the representation of social groups relies on processing affective information, and that the LIFGop supports to a greater extent the processing of

negative words, this area might in turn be expected to support the representation of social groups with a negative valence.

In the present study we applied transcranial magnetic stimulation (TMS) over the LIFGop when healthy participants performed a semantic priming task. In fact, there is a consensus that the contribution of an area to a category-specific representation can be established by applying TMS during a semantic priming task (e.g., Cattaneo, Devlin, Salvini, Vecchi, & Silvanto, 2010; Fuggetta, Rizzo, Pobric, Lavidor, & Walsh, 2009). The stimulation, by influencing the activity of an area based on the initial state of activation, should induce the same facilitatory effects in categorizing an unprimed target as if it was preceded by a congruent prime (Cattaneo, Rota, Vecchi, & Silvanto, 2008). Consistently, we expected a smaller priming effect for social groups than for objects when stimulating the LIFGop, with no significant differences in RTs between congruent and incongruent trials containing social group targets. More specifically, this reduced difference should result from the categorization of social groups in incongruent trials targets becoming faster when stimulating the LIFGop relative to a control site, leaving the responses in congruent trials unaffected. Given the previously established link between the LIFGop and the processing of negative valence stimuli, we additionally expected to record stronger or unique effects for social group words associated with a more negative valence.

2. **Method**

2.1 *Participants*

Twenty participants (10 females, age range: 20 – 31 years) took part to the study for a monetary compensation. The inclusion criteria consisted in the absence of common contraindications to TMS (Rossi, Hallett, Rossini, & Pascual-Leone, 2009, 2011), speaking Italian as first language and self-reported right-handedness. This study protocol was approved and carried out in accordance with the recommendations of the local Ethics Committee (2793/10), and in accordance with the Declaration of Helsinki. All subjects gave written informed consent prior to participating.

2.2 Materials

Stimuli were selected from a dataset of 175 Italian plural related nouns of Objects (e.g., *chiavi* [keys]) and Social groups (e.g., *studenti* [students]). Ratings were collected from a sample of 14 participants, and were based on familiarity, imageability, and valence. The pool of selected stimuli consisted of 30 nouns of Social groups and 30 nouns of Objects (see Appendix 1). The two categories were matched on word length, $t(58) = 1.00$. $p = .34$, familiarity $t(58) = 1.53$, $p = .40$, imageability, $t(58) = -1.92$, $p = .06$, and valence, $t(58) = 1.00$, $p = .32$. Additionally, each category was divided into two sub-samples of equal size based on their valence. Within each category, Positive (e.g., modelle [female models], gioielli [jewels]) and Negative words (e.g., alcolisti [alcoholics], pistole [guns]) were matched based on length ($t_{Social\_Groups}(28) = 1.19$, $p = .24$; $t_{Objects}(28) = .34$, $p = .74$), familiarity ($t_{Social\_Groups}(28) = 1.33$, $p = .19$; $t_{Objects}(28) = .46$, $p = .64$), and imageability ($t_{Social\_Groups}(28) = -.11$, $p = .91$; $t_{Objects}(28) = .43$, $p = .67$), but differed significantly in their valence ($t_{Social\_Groups}(28) = 9.48$, $p < .001$; $t_{Objects}(28) = 11.22$, $p < .001$; see Table 1). The words 'PERSONE' (people) and 'OGGETTI' (objects) were used as primes, with the former being treated as semantically congruent to Social groups, and the latter to Objects. Primes and targets were always presented in uppercase and lowercase, respectively, to make it easier for participants to distinguish between the two and reduce the likelihood of erroneously responding to the primes.

Category	Valence	Imageability	Familiarity	Length
Social positive	6.23 (.46)	7.03 (.42)	6.50 (1.12)	7.93 (1.79)
Object positive	6.06 (.48)	7.36 (.83)	5.90 (1.33)	7.20 (1.47)
Social negative	3.98 (.80)	7.04 (.38)	5.99 (.96)	7.13 (1.89)
Object negative	3.45 (.76)	7.25 (.42)	5.69 (1.08)	7.00 (1.77)

Table 1. Mean valence, imageability, and familiarity on a scale from 0 to 7 (0 = extremely low; 9 = extremely high) and average word length for each word category. Standard deviations reported within parentheses.

To assess whether there were significant differences between each target Category/Valence group in terms of their semantic relationship with the primes, we employed the Italian version of the *snaut* open online software (http://meshugga.ugent.be/snaut-italian-2/; Mandera, Keuleers, & Brysbaert, 2017) to extract their respective semantic distances, calculated based on word co-occurrences in large text corpora. An ANOVA on the resulting values using Category and Valence as between factors and Congruence as within factor revealed only a significant main effect of Congruence [$F(1, 56) = 45.73$, $p < .001$, $\eta_p^2 = .45$], characterized by a lower semantic distance for Congruent ($M = .79$, $SD = .10$), relative to Incongruent primes ($M = .92$, $SD = .08$), with all other effects being nonsignificant (all ps > .35).

2.3   *Transcranial Magnetic Stimulation*

Three-dimensional MRI data of each subject were co-registered to the volunteer's cranium to provide a navigational template for the positioning of the TMS coil. The positioning was conducted with the Brainsight software (version 2.1.5; Rogue Research, Montreal Canada) connected to a Polaris Vicra Optical Tracking System (Polaris, Northern Digital, Ontario, Canada). The system tracked the coil's position with respect to the head using a stereotactic camera, which senses both the coil and the reflectors located on a strap tied to the subject's head. Stimulation sites were identified on each brain reconstruction based on macro-anatomical landmarks (nasion, inion, lateral canthi, and tragi), and the locations of the induced field and stimulation spots were displayed on the MRI data. LIFGop location was based on the findings of Piretti and colleagues (2015) through MNI coordinates ($x = -48$, $y = 1$, $z = 15$), whereas Vertex location was set manually as the midpoint between nasion and inion and between the tragi.

Paired-pulse TMS was delivered via two 200² units connected to a 70 mm figure-of-eight coil through a BiStim[2] module (Magstim Company, Whitland, UK). During the experiment, the coil was held by a mechanical arm. For the LIFGop, the coil was oriented with an angle of approximately 45°

from the nasion-inion line with the handle pointing inwards. For the Vertex, the coil was oriented tangentially to the scalp and perpendicular to the sagittal plane. The two pulses were separated by a 100 ms interval (10 Hz), held to induce an inhibition of cortical activity (as in, Chen, Wassermann, Canos, & Hallett, 1997; Oshio et al., 2010; but see also Opie, Ridding, & Semmler, 2015; Shirota, Sommer, & Paulus, 2016), and delivered approximately every 6 seconds.

2.4   Procedure

Prior to positioning the coil, each participant's resting motor threshold (RMT) was measured. The determination of the RMT was conducted via observed movement motor threshold estimation (OM-MT; Pridmore, Fernandes Filho, Nahas, Liberatos, & George, 1998), defined as the percentage of total machine output (PTMO) of single-pulse stimulation over the left primary motor cortex (M1; ~5 cm lateral from the Vertex) that evoked 5 out of 10 visible contralateral finger/hand motor responses. Stimulation intensity was then set to 90% of each subject's RMT ($M = 30.7$, $SD = 3.50$).

After positioning the coil over the target area, participants completed a semantic priming paradigm. A PC running E-Prime (version 2.1, Psychological Tools, Inc) controlled the presentation of the stimuli, the recording of responses, and the onset of the TMS pulses. Stimuli were projected on a white background via a 19" LCD monitor with resolution of 1280*1024 pixels and a screen frame rate of 60 Hz. During each trial, a fixation cross was presented for 500 ms, followed by the presentation of a prime word (50 ms). The prime was replaced by a blank screen (100 ms) that preceded the target which remained on the screen for 1500 ms, or until participants gave a response via button press (see Fig. 1).
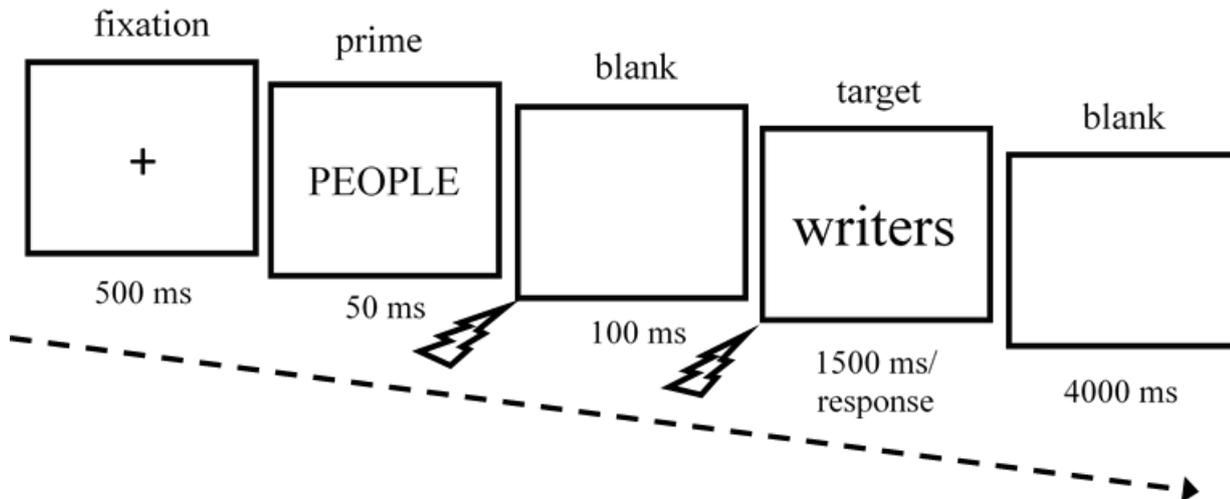
Figure 1. Temporal progression of a trial containing a Positive Social group target preceded by the semantically congruent prime. Bolts indicate TMS pulse delivery, with the first pulse coinciding with the offset of the prime and the second with the onset of the target word, separated by a 100 ms interval.

Within the task, the first TMS pulse was delivered at the offset of the prime, whereas the second one coincided with the onset of the target word. To avoid potential effects associated to the stimulation in the preceding trial, a 4000 ms inter-trial interval was used, making a single trial last up to a maximum of 6150 ms. Participants were instructed to just attend the first word (prime) and to categorize the second word (target) based on whether it represented a living or a non-living entity. The response was given via button press, with the category-button mapping counterbalanced across participants. Since the TMS was mainly targeted to a left-lateralized area, participants always responded using the index and middle fingers of their left hand. Before starting with the main task, participants were given 5 practice trials with a different sample of stimuli to familiarize with the task and the stimulation. During the main task, all target stimuli were presented twice in a random order, once with the semantically congruent, and once with the semantically incongruent prime. The resulting 120 trials were then split into two blocks of 60 trials each. After the first block, participants could take a short self-paced break, after which they continued by pressing a response button. The same procedure, lasting approximately 7 minutes, was completed by each participant twice, once per

stimulation site, with the order of the site being counterbalanced across participants. The experiment thus consisted of a 2 (Site: LIFG vs. Vertex) x 2 (Category: Social groups vs. Objects) x 2 (Valence: Positive vs. Negative) x 2 (prime-target Congruence: Congruent vs. Incongruent) within-subjects design, with response time (RT) as the dependent variable. Given the subjectivity of the valence ratings previously acquired, to check whether the present subjects considered the two semantic categories having an equal valence within each level of the pre-established Valence factor, we asked them to rate the valence of each word at the end of the session.

2.5 Data analysis

We used R (R Core Team, 2016) and lme4 (Bates, Mächler, Bolker, & Walker, 2015) to perform a linear mixed effects analysis using RTs as predicted variable and Site, Category, Valence, Congruence and their interactions as fixed effects. The model also controlled for the effects of item length, familiarity and imageability to improve the accuracy of the fixed effect estimates. Random intercepts for each item nested within each subject were used as random effects. Incorrect trials and trials with RTs greater than 2 SD from their average (629 ms, $SD = 161$) were filtered prior to analysis. To fix deviations from normality, as from visual inspection of residual plots, RT were transformed via reciprocal log-transformation. Following transformation, no visible deviations from normality we detected. P-values are reported using restricted maximum likelihood (REML) estimation and the Satterthwaite approximation to calculate the denominator degrees of freedom. Follow-up analyses to explore the significant interactions were conducted via an analysis of simple effects, with Bonferroni-corrected contrasts between congruence and stimulation site conditions. The effects size of the contrasts were calculated using the Cohen's d for repeated measures (Cohen's d_z; Lakens, 2013; Rosenthal, 1986).

As a manipulation check, to assess whether the subjects' valence ratings coincided with the pre-determined Valence factor, we ran an ANOVA on the average valence ratings of the target words given at the end of the session, using Category and Valence as factors.

3. Results

3.1 Main results

The analysis revealed a significant main effect of Congruence [$F(1, 3325) = 50.51$, $p < .001$], with participants being overall slower on Incongruent trials compared to Congruent trials, and no significant main effects of Site [$F(1, 3342) = 1.56$, $p = .21$], Category [$F(1, 1129) = .18$, $p = .67$] and Valence [$F(1, 1128) = .71$, $p = .40$]. Two significant three-way interactions emerged between Valence, Category and Site [$F(1, 3348) = 7.71$, $p = .006$], and between Valence, Congruence and Site [$F(1, 3328) = 3.85$, $p = .05$], but not a significant four-way Congruence x Category x Valence x Site interaction [$F(1, 3329) = .01$, $p = .93$]. To disentangle the Valence x Category x Site interaction, we tested how RTs to each Category were influenced by stimulation Site based on their valence. To do so, we ran two random effects models, one for each level of Valence, using Category and Site as fixed factors, controlling for imageability, length and familiarity. Random intercepts for Congruence nested within items and subjects were used as random effects.

For Negative stimuli, the results showed a significant Site x Category interaction [$F(1, 1147) = 7.19$, $p = .007$]. Bonferroni-corrected contrasts addressing stimulation effects showed a significant effect for Social groups [$t(1158) = -2.58$, $p = .03$, Cohen's $d = -.58$], with faster responses when stimulating the LIFGop compared with the Vertex, and no effect for Objects [$t(1135) = 1.20$, $p = .68$, Cohen's $d = .27$] (see Figure 2). In addition, there was a significant difference in the stimulation effect between the two categories [$t(1147) = 2.68$, $p = .02$, Cohen's $d = .60$]. As for the Positive stimuli, neither the interaction [$F(1, 2150) = 1.58$, $p = .21$], nor the main effect of Site [$F(1, 2150) = .61$, $p = .43$], or Category [$F(1, 2150) = 1.58$, $p = .21$], were observed.
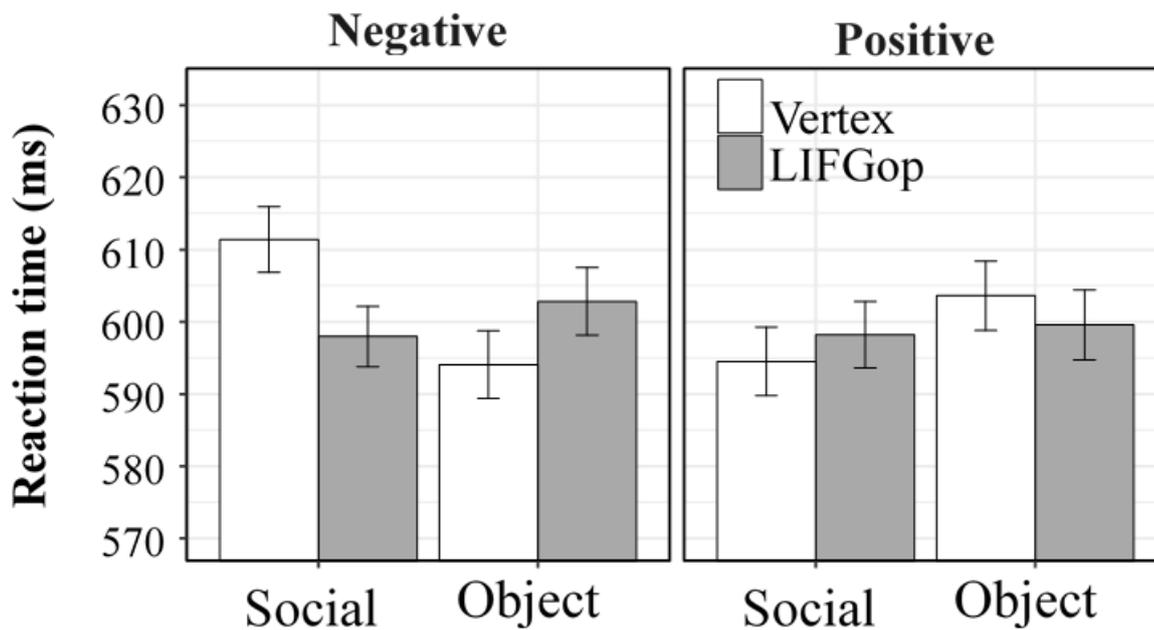
Figure 2. Mean RTs (ms) for Negative (left panel) and Positive targets (right panel) separated by stimulation site and Category. Relative to the Vertex, participants were significantly faster to respond to social groups of more negative valence when the LIFGop was being stimulated. Error bars represent ± 1 SE.

To disentangle the Valence x Congruence x Site interaction, we tested how RTs in each Congruence condition were influenced by the stimulation Site based on whether they belonged to the Negative or Positive valence (see Figure 3). As for the previous interaction, we ran two random effects models, one for each level of Valence, using Congruence and Site as fixed factors, controlling for word imageability, familiarity and valence ratings. Random intercepts for items nested within categories and within subjects were used as random effects.

The results for Negative stimuli showed a main effect of Congruence [$F(1, 1679) = 12.60$, $p < .001$], no significant effect of Site [$F(1, 1686) = .97$, $p = .32$], and a marginally significant Site x Congruence interaction [$F(1, 1676) = 3.35$, $p = .07$]. Bonferroni-corrected contrasts on priming effects (Congruent vs Incongruent trials) showed the presence of a significant priming during the

stimulation of the Vertex [$t(1681) = 3.79$, $p < .001$, Cohen's $d = .84$], but not during the stimulation of the LIFGop [$t(1673) = 1.22$, $p = .67$, Cohen's $d = -.27$]. For the Positive valence stimuli, the significant main effect of Congruence [$F(1, 1647) = 40.35$, $p < .001$], did not interact with the Site factor [$F(1, 1654) = .96$, $p = .33$].
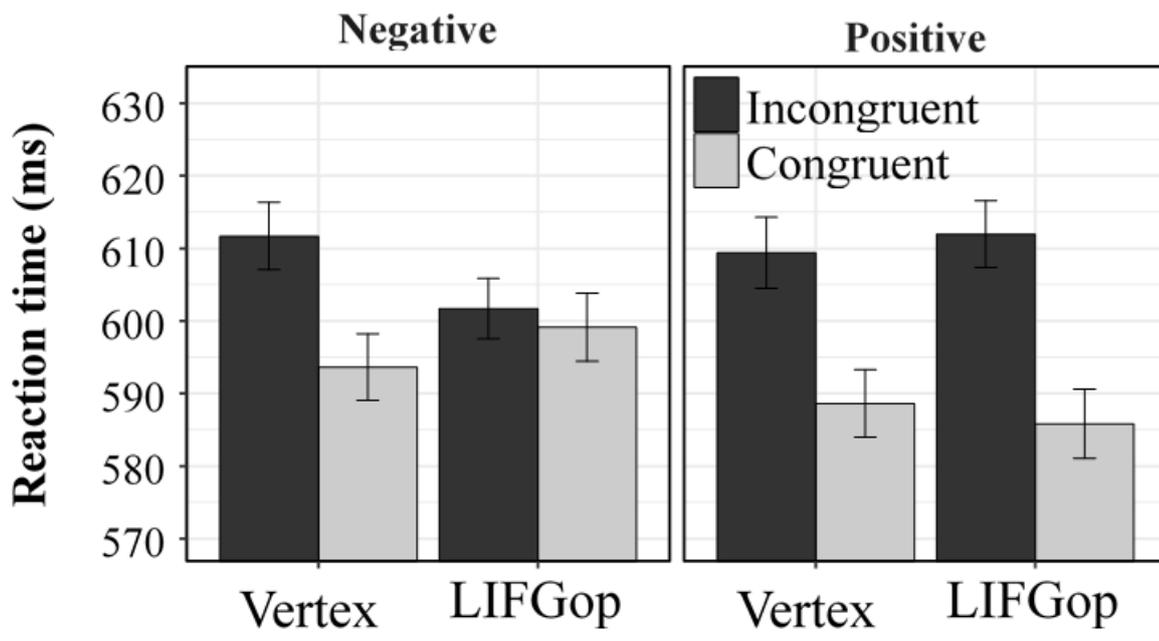
Figure 3. Mean RTs (ms) for Negative (left panel) and Positive targets (right panel) separated by stimulation site and semantic congruence with the preceding prime. Error bars represent ± 1 SE.

3.2   *Manipulation check*

The analysis of the valence ratings showed a significant main effect of the pre-established Valence factor [$F(1, 19) = 65.30$, $p < .001$, $\eta^2 = .77$], with positive valence words having higher ratings ($M = 6.29$, $SD = 1.07$) compared to negative words ($M = 4.26$, $SD = 1.20$). The main effect of Category and the interaction between Category and Valence were not significant (both $ps > .10$).

3.3 Exploratory analysis

We further explored the data to assess the extent to which the faster responses for Negative Social groups during LIFGop stimulation were driven by changes in RTs in Incongruent or Congruent trials relative to the Vertex. To this end, we ran Bonferroni-corrected contrasts for stimulation effects (Vertex vs. LIFG) on Incongruent and Congruent trials containing Negative Social groups or Negative Objects.

For Negative Social groups, a significant effect of the stimulation emerged in Incongruent trials [$t(1686) = -2.74$, $p = .01$, Cohen's $d = -.61$], but not in the Congruent ones [$t(1687) = -.93$, $p = .71$, Cohen's $d = -.21$], indicating that the overall faster responses to Negative Social groups, and the same lack of a priming in the LIFGop stimulation condition, was mostly driven by faster responses to unprimed targets (see Figure 4). In the case of Negative Objects, no significant effects of stimulation were observed for either Incongruent [$t(1677) = -.12$, $p = 1.00$, Cohen's $d = -.03$], or Congruent trials [$t(1670) = 1.76$, $p = .15$, Cohen's $d = .39$].
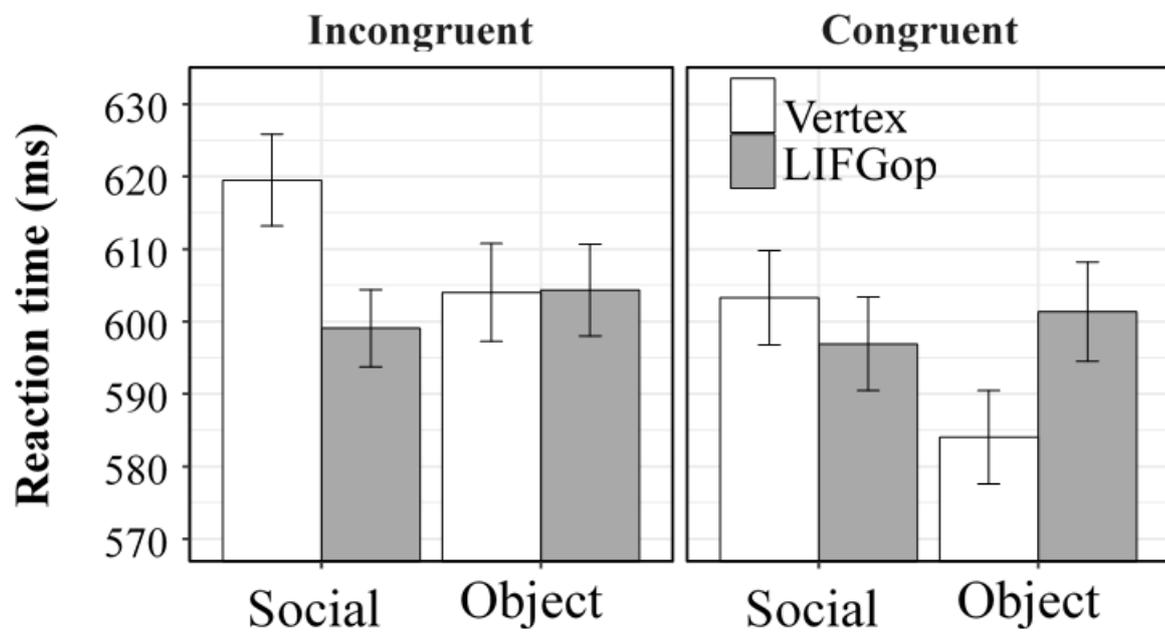

Figure 4. Mean RTs (ms) for Incongruent (left panel) and Congruent (right panel) of negative valence divided by semantic category and site of stimulation. Error bars represent ± 1 SE.

4. **Discussion**

With the present study we aimed at assessing the role played by the LIFGop in representing social groups. This prediction was based on empirical facts and extant theories. First, brain-damaged patients were described with a selective deficit in processing social groups (Rumiati et al., 2014), and the LIFGop was identified as the underlying common cortical region associated with patients' poorer naming performance on social groups (Piretti et al., 2015). Second, some theoretical propositions (Barsalou, 2008; Lambon Ralph et al., 2016; Martin, 2016) led us to hypothesize that the semantic representation of social groups might be better captured by affective properties, with the affective valence playing a critical role. Moreover, the observed association between LIFGop activity and the processing of negative valence words (García-Pacios et al., 2017; Kensinger & Schacter, 2006; Leclerc & Kensinger, 2011) suggested that this area might be specifically involved in representing social groups of negative affective valence.

When TMS was applied over the LIFGop, relative to the Vertex, we found a significant decrease in RTs to negative social categories, with no effect for positive social categories or objects of any valence. In line with our hypothesis, we interpret this finding as an indication of a preferential involvement of the LIFGop, boosted by TMS, in the representation of social groups when processing negative affective attributes. An exploratory analysis also revealed that, when the stimulation was applied to the LIFGop (relative to the Vertex), the faster responses were mostly driven by lower RTs to unprimed trials. These findings, together with the unaltered responses to primed trials, are consistent with previous TMS studies on object representations, also reporting faster responses in unprimed trials (Cattaneo et al., 2010, 2008). According to the state-dependent hypothesis, the TMS effect on the target neuronal population in a state of lower activation (during the unprimed condition) should lead to an increased excitability and faster processing (Cattaneo et al., 2010, 2008; Silvanto, Muggleton, & Walsh, 2008). It is thus plausible that the overall facilitation we found in categorizing

negative social groups derives from a combination of the TMS-induced facilitation in incongruent trials and a stable priming effect in congruent trials.

Additionally, our results have shown a disruption of the priming effect, which, contrary to our expectations, was not unique to social groups, but it was present for negative words in general, including objects. In the latter case, although not significant, the disruption of the priming was numerically mostly led by a decrease in RTs for primed object targets, contrary to what was found for social groups. If the LIFGop supports the activation of negative affective representations and these are more strongly linked to the processing of social groups, then an increase in its excitability through stimulation might have increased participants' readiness to categorize a social group when reading a negative word regardless of its category. We hypothesize that such readiness might have produced a lag in the categorization of primed object targets. Although purely speculative, this idea is in line with electrophysiological studies suggesting the presence of effects driven by the affective content of words preceding and influencing later semantic processing (Kissler, Assadollahi, & Herbert, 2006).

It is still not clear whether the involvement of the LIFGop might extend beyond accessing the existing social semantic knowledge. For instance, a recent TMS study failed to find a significant involvement in the formation of social impressions (Ferrari et al., 2016). However, unlike the present study, the possibility of a valence-specific effect was not considered by Ferrari and colleagues, suggesting that a possible contribution of this region in creating novel social impressions, especially for negative stimuli, is not to be excluded. To this end, it is worth noting that an fMRI study on impression formation from social behavioral descriptions reported an increased activation of the LIFGop when contrasting negative vs. positive social attributes (Van der Cruyssen et al., 2014).

When it comes to accessing the representation of social groups, our findings support the hypothesis on the relevance of affective information, consistently with the proposal that social concepts are characterized by a predominance of affective properties (Piretti et al., 2015). This conclusion is consistent with the role of LIFGop in processing negative valence words (García-Pacios

et al., 2017; Kensinger & Schacter, 2006; Leclerc & Kensinger, 2011; Van der Cruyssen et al., 2014), and with multiple theorizations linking the processing of social concepts with affect (e.g., Lambon Ralph et al., 2016; Martin, 2016). Consistently with the Conceptual topography theory (Simmons & Barsalou, 2003), the present findings also support the distinction in processing of positive and negative affective features within the emotional modality, additionally suggesting the latter being implemented in the LIFGop. Finding the same pattern of results for positive words following the stimulation of an area involved in the processing of positive affective information would bring additional evidence in support of this distinction and of the hypothesis that social concepts rely on affective information.

Our results are not easily accommodated within previous theories of emotion lateralization suggesting that the processing of negative affect is right-lateralized in the prefrontal regions (Davidson, 1992, 1995; Davidson & Irwin, 1999). Consistent with our findings, however, in a more recent neuroimaging study the affective processing of picture and word stimuli led to a bilateral activation of prefrontal lateral regions (Kensinger & Schacter, 2006). In the domain of language, Kensinger & Schacter (2006) found a greater activation for negative compared with positive words only in the LIFGop. Yet, our results do not exclude the possibility that also other areas might support the processing of social groups with a negative valence. We rather favor the idea that the LIFGop is part of a wider network of cortical and subcortical regions supporting affective and semantic processing, that in our study we might have affected indirectly with the stimulation.

In conclusion, our data support the hypothesis of a predominant role of the LIFGop in affective processing of social groups. Confirming the previously evidenced involvement of the region in processing negative features, we suggest that its activation facilitates the processing of lexical stimuli that rely more heavily on such features. In this view, compared to the categorization of objects, the categorization of social group names is facilitated because their representation relies more on affective information than objects do. As such, our findings bring additional evidence to the

theoretical propositions that link semantic representations to affective processes (Barsalou, 2008; Lambon Ralph et al., 2016; Martin, 2016; Simmons & Barsalou, 2003) and suggest where in the brain this interplay is likely to take place. While a comparison with only the objects category poses a limit to the strength of our conclusions, we encourage future studies on the representation of social groups to replicate the present findings by taking into account comparisons with semantic categories such as animals. As previous research has argued objects and animals to be distinguished for the variety of their uses and of their visual properties, respectively, we argue that the factor playing a greater role in enabling us to discern among social groups is likely to be the affective valence we attribute to them.

CHAPTER 3

Study 2: Electrophysiological Correlates of Social Group Representations in Affective Priming

Abstract

There is growing evidence in cognitive neuroscience that processing of information about social groups involves the associated affective features, compared with processing information about nonsocial semantic categories. With the present study we aimed at assessing the extent of such involvement by measuring event-related potentials in healthy individuals while they performed an affective priming paradigm requiring evaluative responses. Behavioral results showed a greater affective priming for social group than for nonsocial category targets, while the analysis of the neural correlates revealed a modulation in the late positive component, which was higher in the positive valence social groups as compared to positive valence nonsocial categories. The present findings complement previous neuropsychological and brain stimulation studies by showing how the engagement in affective processing enhances the representation of social groups compared to nonsocial categories, as indicated by the emergence of a distinct behavioral and neurophysiological response.

1. Introduction

When we categorize our conspecifics into social groups based according to characteristics such as gender, ethnicity, social status or occupation, we rely on both semantic and affective information. For example, the label "criminals", referred to people being guilty of crimes, comes together with its generally negative evaluation; likewise, the label "students", in addition to informing us about the

young age of individuals that attend some type of educational system, would generally be attached to a positive evaluation. The semantic component informs us about the appearance and behavior of the group, whereas the affective one determines our attitude, a positive or negative disposition towards it (T. D. Wilson, Lindsey, & Schooler, 2000). The semantic and affective systems have often been regarded as being apart (Pessoa, 2008), however arguments for their interdependence have also been put forward. For instance, it has been suggested that affective information would be a core component of our conceptual knowledge as much as any other type of information, and that its relevance would differ depending on the semantic category (Barsalou, 2008; Lambon Ralph et al., 2016). In particular, some authors have proposed that the affective information largely contributes towards the representation of conspecifics (Lambon Ralph et al., 2016; Mahon & Caramazza, 2011; Martin, 2016). Indeed, neuropsychological studies documented patients with a selective deficit at processing social groups, thus suggesting that concepts about them might have a representation of their own (Carnaghi, Silveri, & Rumiati, 2015; Rumiati, Carnaghi, Improta, Diez, & Silveri, 2014). Moreover, widespread tumor lesions in areas commonly linked to affective processes affected the patients' accuracy at naming pictures of social groups (Piretti et al., 2015). What needs to be explained is the role played by valence in processing the affective information. Using transcranial magnetic stimulation on healthy subjects, the increased excitability of the inferior frontal gyrus, associated to the processing of negative valence stimuli, has been found to facilitate the categorization of negative social groups (Suran, Rumiati, & Piretti, 2019).

To further assess the extent to which affective valence contributes to the semantic representation of social groups, we employed an affective priming paradigm (Fazio, 2001; Fazio, Sanbonmatsu, Powell, & Kardes, 1986). During the test, participants are shown either a positive or negative target word upon which they build an evaluative response. To influence their speed in evaluating the target, a prime word of the same or opposite valence is presented right before it, pre-

activating an affective processing. The priming effect corresponds to faster responses to targets preceded by affectively congruent, relative to affectively incongruent primes.

In order to establish the stage at which the primes influence the processing of the targets, we complemented the behavioral paradigm with the measurement of event-related potentials (ERPs). The two most studied components in affective priming are the N400 (~400 ms post-target onset) and the late positive component (LPC; ~600 ms post-target onset), whose effects emerge from their differences between congruent and incongruent trials. In this context, the former effect has been associated with the process of integrating the semantic content of primes and targets (Zhang, Lawson, Guo, & Jiang, 2006; Zhang, Li, Gold, & Jiang, 2010). The LPC, reported to be sensitive to word valence (for a review, see Citron, 2012), has been regarded to also capture affective prime-target incongruities and thus to be a more reliable index of affective priming when controlling for non-affective prime-target relationships (Herring et al., 2011; Hinojosa, Carretié, Méndez-Bértolo, Míguez, & Pozo, 2009).

In line with previous studies showing a facilitated response to manipulable object targets following sensory-motor priming (Labeye, Oker, Badard, & Versace, 2008; Myung, Blumstein, & Sedivy, 2006), at the behavioral level we expected that activating the processing of affective information would result in a facilitated response (i.e. shorter reaction time) to social groups. Moreover, such facilitation, if not specific to social groups, was expected be significantly greater when compared to the facilitation observed for non-social semantic categories. As there is yet no consensus concerning which ERP component should be more sensitive to affective priming, as in previous studies, we expected to find both an N400 (Zhang et al., 2006, 2010) and an LPC effect (Herring et al., 2011). Additionally, we expected these effects to be greater towards social category targets relative to nonsocial categories, reflecting their hypothesized stronger reliance on the affective features being primed.

2. Method

2.1 Participants

Twenty-six healthy right-handed participants, all Italian native speakers, were recruited from the local population in exchange of monetary compensation, and all gave written consent prior to participation. Because of the inability to obtain sufficient EEG data in three participants due to the presence of artefacts coming from excessive movement, these were excluded from the analysis, bringing the final sample to 23 participants (13 female, age range 19-29). The study was approved by the Ethics Committee of the International School for Advanced Studies (SISSA, Trieste) in accordance with the Declaration of Helsinky.

2.2 Materials

A total of 108 plural nouns of social categories (N = 54) and nonsocial categories (N = 54) were used as target stimuli after being selected from a larger database rated by a different sample of 12 subjects (see Appendix 2). Each semantic category was divided in a positive and negative subset of 27 elements each, differing significantly in their valence ratings [Social positive vs. Social negative, $t(52) = 20.66$, $p < .001$; Nonsocial positive vs. Nonsocial negative, $t(52) = 22.24$, $p < .001$). Additionally, within each valence subset, social and nonsocial category nouns were matched according to their average ratings of valence [Social negative *vs* Nonsocial negative, $t(52) = 1.32$, $p = .19$; Social positive *vs* Nonsocial positive, $t(52) = 1.34$; $p = .19$), arousal [Social negative *vs* Nonsocial negative, $t(52) = 1.54$, $p = .13$; Social positive *vs* Nonsocial positive, $t(52) = 1.56$; $p = .13$], familiarity [Social negative *vs* Nonsocial negative, $t(52) = 1.22$, $p = .23$; Social positive *vs* Nonsocial positive, $t(52) = .28$; $p = .78$], and length [Social negative *vs* Nonsocial negative, $t(52) = .71$, $p = .48$; Social positive *vs* Nonsocial positive, $t(52) = .98$; $p = .33$] (see Table 1). A separate sample of 18 words to be used as primes was obtained from the ANEW rated database of Italian words (Montefinese, Ambrosini, Fairfield, & Mammarella, 2014). These consisted of 9 positive and 9 negative nouns and adjectives of significantly different valence [$t(18) = 33.23$, $p < .001$], and matched

according to their arousal [$t(18) = .28$, $p = .78$], familiarity [$t(18) = .02$, $p = .98$], length [$t(18) = .86$, $p = .40$], concreteness [$t(18) = .24$, $p = .81$]. The additional concreteness matching was done to account for the different effects of abstract and concrete prime words on affective priming (Yao & Wang, 2013). To this list of words, 9 strings of hash symbols (#) with an equal average length to the positive and negative words (5-8 elements) was added to serve as affectively neutral primes, bringing the total number of primes to 27.

To ensure an equal semantic relationship between the prime words and each target Category/Valence group, we employed the Italian version of the *snaut* open online software (http://meshugga.ugent.be/snaut-italian-2/) to extract their respective semantic distances, calculated based on word co-occurrences in large text corpora via cosine similarities between semantic vectors (Mandera et al., 2017). A mixed ANOVA on the extracted values using Category and Valence as between factors and Congruence as within factor showed, as expected, a significant main effect of Congruence [$F(1, 104) = 104.9$, $p < .001$, $\eta_p^2 = .50$] characterized by a lower semantic distance for Congruent relative to Incongruent primes, and an interaction between Congruence and Valence [$F(1, 104) = 13.05$, $p < .001$, $\eta_p^2 = .11$] with a greater distance between congruent and incongruent primes for negative, relative to positive words.

Category	N	Valence	Arousal	Familiarity	Length	Semantic distance Δ
Social Positive	27	6.24 (.48)	4.11 (.42)	4.02 (1.11)	7.19 (1.24)	.03 (.04)
Nonsocial Positive	27	6.41 (.45)	3.90 (.56)	3.94 (1.13)	6.82 (1.52)	.03 (.04)
Social Negative	27	3.06 (.64)	3.93 (.50)	3.10 (.88)	7.30 (1.54)	.05 (.05)
Nonsocial Negative	27	3.28 (.58)	3.73 (.48)	2.81 (.86)	7.00 (1.52)	.06 (.04)

Table 1. Mean valence, arousal and familiarity ratings (0 = extremely low; 7 = extremely high), average word length, and difference in the semantic distance (cosine) with the congruent and incongruent primes for each target word category separated by valence class. Standard deviations reported within parentheses.

2.3 Procedure

Participants performed the affective priming task in a dark and acoustically insulated room. The task required to evaluate a target word preceded by a prime of the same or opposite valence. Instructions were provided to the participants in both written and oral form by the experimenter. Prior to the task proper, participants completed a set of 20 practice trials containing a different set of primes and targets. To ensure the task had been understood, participants had to undergo 20 additional practice trials in a newly randomized order if they had responded incorrectly to more than practice 5 trials. Each trial began with the presentation of a fixation cross (500 ms), after a 200 ms interval, the prime (150 ms) and the target (300 ms) appeared, separated by a 100 ms ISI (see Figure 1). After the offset of the target, participants were presented with a blank screen during which they had an additional 1500 ms to give their evaluation via button press. Following this period, or after pressing a button, the blank-screen intertrial interval was jittered between 800 and 1200 ms (in 100 ms intervals). To respond, participants pressed either the "f" or the "j" key on a QWERTY keyboard, with "positive" and "negative" response mapping counterbalanced across participants. Key-choice was driven by the presence of dash-like bumps allowing for easier identification in the dark room. To help participants differentiate primes from targets, the primes were always presented in uppercase and targets in lowercase format. Additionally, to ensure words with equal length subsided an equal amount of visual angle, all word stimuli were presented in Lucida Sans Typewriter monospaced font, thus making the amount of horizontal space occupied by each character the same.

Participants completed 324 test trials split in 12 equal blocks. The breaks between blocks were self-paced. The 108 targets were presented three times, each time paired with a prime of different valence. To avoid two same targets appearing too close to one-another, prime-target pairs were presented in a pseudo-random order. Only after all targets had been presented once in a random order, they were presented for the second time in a new order, and the same was done for the third presentation. Within each of the 3 resulting cycles of 108 trials, the order of the appearance of each

prime-target pair was randomized. The task was programmed in Python using the Psychopy library (Peirce, 2007), and lasted approximately 25 minutes.
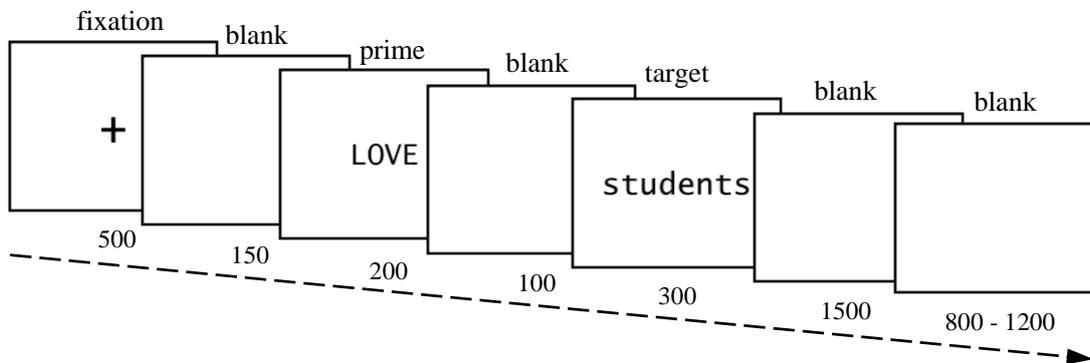

Figure 1. Schematic representation of the affective priming task depicting an affectively congruent trial. Stimuli succession is represented left to right, with the respective screen presentation time expressed in milliseconds below each stimulus. The last blank display represents the intertrial interval with a jittered duration between 800 and 1200 ms in 100 ms intervals.

2.4   *Behavioral data analysis*

Behavioral data were preprocessed and analyzed in R (R Core Team, 2016). Subjects whose overall accuracy was lower than 80% were excluded from further analyses ($n = 3$). On the remaining 20 subjects, the average accuracy score was subjected to a 2 (Category: Social vs. Nonsocial) x 2 (Valence: Positive vs. Negative) x 3 (Congruence: Congruent vs. Incongruent vs. Neutral) repeated measures ANOVA. The same model was then applied to the log-transformed RTs following the removal of incorrect trials and trials in which RTs deviated by more than 2 SD from each subject's average (14% of the total trials). As a measure of effect size we used partial eta squared. Post-hoc tests for significant higher order effects were calculated as planned contrasts on priming effects applying the Bonferroni correction.

2.5 *EEG recordings and data analysis*

EEG was recorded with a set of 64 Ag/AgCl active electrodes mounted on an elastic cap based on the International 10-20 system (Klem, Lüders, Jasper, & Elger, 1999). Signal amplification was achieved through a BioSemi Active-Two amplifier system. During recording data were visualized and stored by means of ActiView acquisition software (ActiView 707, Biosemi, Amsterdam, The Netherlands). Electrode offsets were kept between ±20 mV, and the signal was sampled at a rate of 1024 Hz, with a 24-bit digitization resolution. A common mode voltage based on the ActiveTwo's CMS/DRL feedback loop was used for analog-to-digital conversion of recorded voltages for each electrode (cf. to https://www.biosemi.com/faq/cms%26drl.htm). Data were band-pass filtered between 0.01–100 Hz during data acquisition. To monitor eye movements and blinks, horizontal and vertical electrooculogram (EOG) was recorded via four additional electrodes placed at the outer canthi of both eyes, and below and above the left eye.

EEG data preprocessing was performed using the Brainstorm software (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011) and subsequently analyzed in R (R Core Team, 2016) using the erpR package (Arcara & Petrova, 2014). First, the EEG recordings were resampled offline at 250 Hz and band-pass filtered (0.05 – 40 Hz). Bad electrode channels were removed upon visual inspection, as well as movement artefacts. Eye blinks artifacts were removed through independent components analysis (ICA; Makeig, Bell, Jung, & Sejnowski, 1996). Epochs containing artefacts other than eye movements were removed after visual inspection. Average ERPs were computed separately for each participant, electrode, target category/valence combination from -200 to +1500 ms relative to the onset of the target, and baseline-corrected using the 200 ms pre-stimulus period.

To identify the time windows for the N400 and the LPC components and avoid the issue of circularity in windows selection, we used the *collapsed localizer* approach (Luck & Gaspelin, 2017) and the examination of existing literature. Both for the N400 (480 – 680 ms) and the LPC (700 - 900 ms), the waves collapsed across all conditions were used. For the respective components, the time

regions surrounding the greatest peak amplitudes were selected. Four symmetrical regions of interest (ROIs; see Fig. 2) were then created by averaging the amplitude of the respective electrodes selected based on a visual inspection of the scalp topography of collapsed conditions in the relevant time windows previously obtained. Each of the four ROIs represented a combination of Caudality (frontal vs. posterior) and Laterality (left vs. right).

The resulting ERPs were subjected to two 2 (Category: Social vs. Nonsocial) x 2 (Valence: positive vs. negative) x 2 (Congruence: congruent vs. incongruent) x 2 (Caudality: frontal vs. posterior) x 2 (Laterality: left vs. right) repeated measures ANOVAs, one for each component's time window. Follow-ups to significant interactions consisted of Bonferroni-corrected contrasts of priming effects. Interactions that involved spatial factors (Caudality or Laterality) were followed up within each level. The aim of our study was to assess the effects of affective congruence across categories, thus, consideration is exclusively given to the contrasts concerning the interactions between affective Congruence and target Category. Only interactions yielding significant follow-up contrasts are reported.
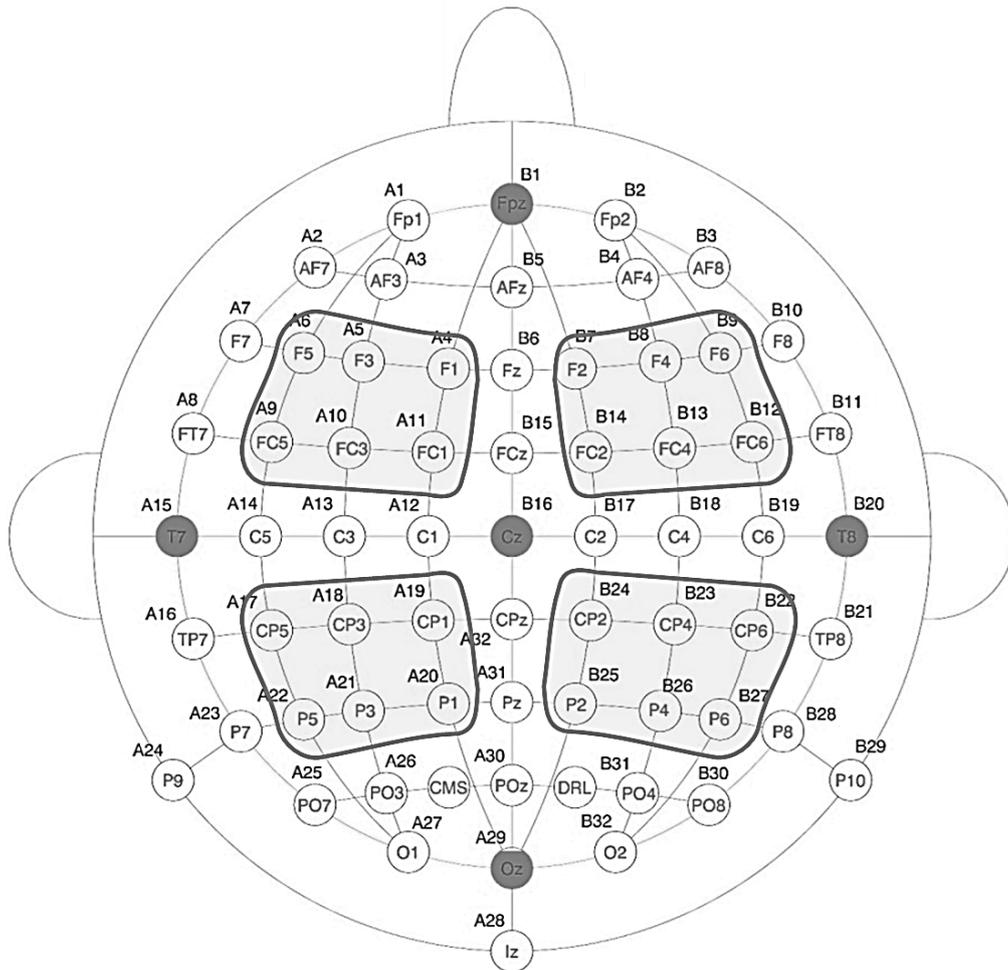
Figure 2. Scalp regions used for electrode grouping for statistical analyses of event-related potentials.

3. **Results**

3.1 *Behavioral results*

The analysis of the RTs showed a significant main effect of Congruence [$F(2, 38) = 8.76$, $p < .001$, $\eta_p^2 = .32$] and a significant interaction between Category and Congruence [$F(2, 38) = 3.88$, $p = .03$, $\eta_p^2 = .17$]. All other main effects and interactions were not significant ($p > .1$). Follow-up analyses on the Congruence effect revealed a significant difference between the Congruent and Neutral condition [$t(38) = 3.48$, $p < .01$], and between the Congruent and Incongruent condition [$t(38) = 3.76$, $p < .01$], whereas between the Incongruent and Neutral conditions there was no significant difference [$t(38) = -.30$, $p > .05$]. In the interaction, a significant difference between Congruent and Incongruent trials

(i.e., the priming effect) was observed for Social groups [$t(38) = 4.65$, $p < .001$] but not for Nonsocial categories [$t(38) = 1.87$, $p = .38$]. This difference was significantly greater for Social groups relative to Nonsocial categories [$t(38) = -2.78$, $p = .05$].
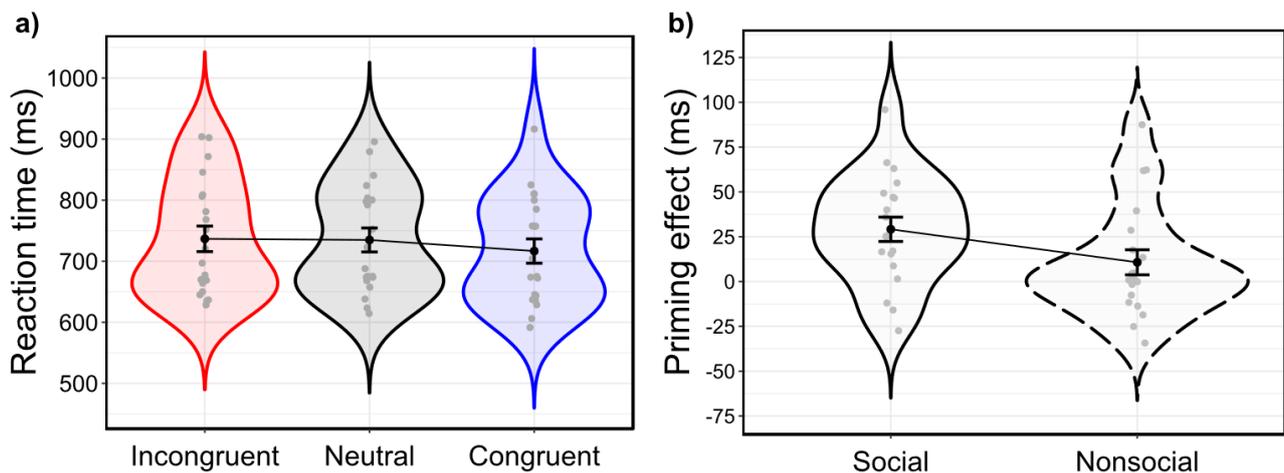
Figure 3. a) Violin plots with reaction times (ms) to target words based on the affective congruence with the preceding prime. Congruent trials are faster relative to Incongruent and Neutral trials, indicating the presence of a facilitation of the affectively congruent primes. b) Priming effect, calculated as the difference between Incongruent and Congruent trials, for Social and Nonsocial targets. Social groups showed a greater affective priming effect following the same primes. Error bars represent ±1 SE.

3.2   ERP results

480 – 680 ms. This time interval encompassed a negative-going wave, peaking around 600 ms, interpreted by previous authors using an analogous paradigm as reflecting the N400 (Zhang et al., 2006). Although the results revealed a significant interaction between the factors of Category, Valence, and Congruence [$F(1, 19) = 6.85$, $p = .017$, $\eta_p^2 = .26$], post-hoc analyses did not show significant categorical differences based on Congruence for targets of either Valence (both ps > .05).

700 – 900 ms. The analysis of the later time-window also revealed a significant 3-way interaction between Category, Valence and Congruence [$F(1, 19) = 7.25$, $p = .01$, $\eta_p^2 = .28$] which in

turn interacted also with Caudality [$F(1, 19) = 6.57$, $p = .02$, $\eta_p^2 = .26$]. Post-hoc Bonferroni-corrected contrasts showed a higher LPC for Congruent as compared to Incongruent trials for Positive Social group targets in the anterior electrode clusters [$t(89) = -3.09$, $p = .008$], which was not significant in the posterior clusters [$t(89) = -2.21$, $p = .09$]. Only in the anterior electrodes, this difference between Congruent and Incongruent trials for Social group targets was also significantly greater than the one for Nonsocial targets [$t(89) = 2.70$, $p = .03$]. No other LPC effects were observed (all ps > .05). An exploratory analysis of the anterior clusters motivated by the visual inspection of the scalp distribution of the LPC effect for Social groups (see Figure 4b) has shown that a marginally significant effect over the left electrodes [$t(104) = -2.43$, $p = .05$], and a significant effect over the right electrodes [$t(104) = -3.50$, $p = .002$], leading to a significant difference between Social and Nonsocial categories only over the right anterior cluster [$t(48) = 2.87$, $p = .02$]. Again, no LPC effect was found for Nonsocial targets (all ps > .05).

a)

b)

Social positive   Nonsocial positive

Social negative   Nonsocial negative
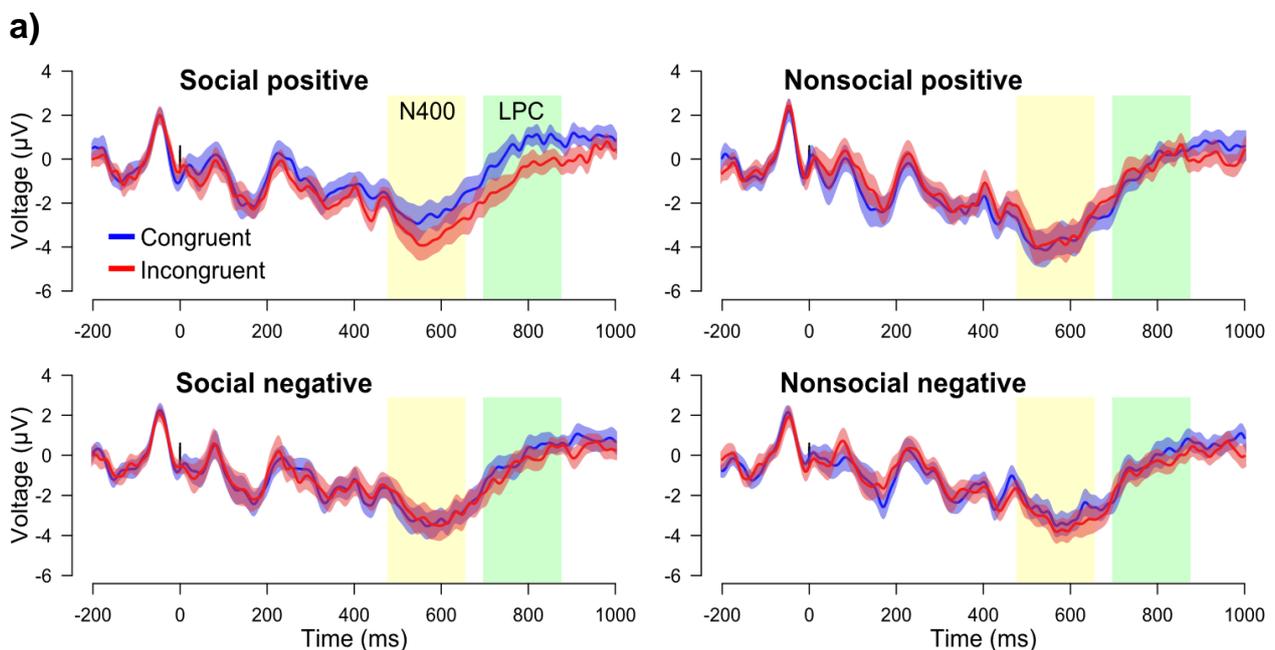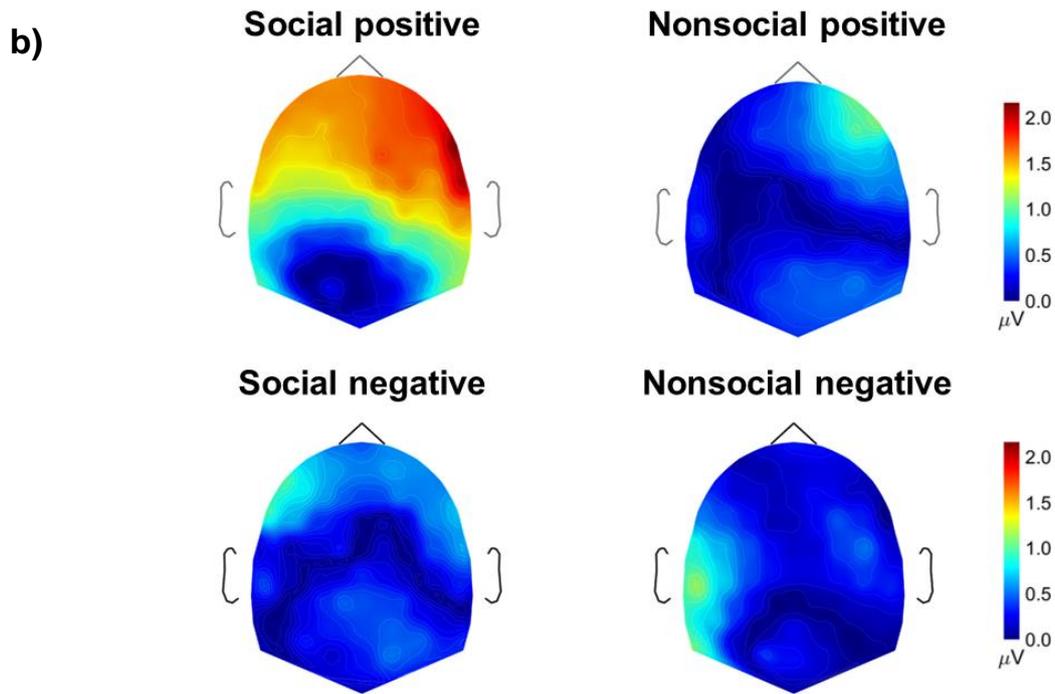
Figure 4. a) Grand average ERP waveforms of the right anterior electrode cluster divided by congruent (blue lines) and incongruent trials (red lines) for social groups (left column) and objects (right column) of positive valence (top row) and negative valence (bottom row). Only social groups of positive valence presented a significant priming effect in the LPC time window. Colored bands represent ±1 SE of the respective grand averages. b) Following the same disposition, topographic maps of the difference in amplitude between congruent and incongruent trials in the 700-900 ms time window for all conditions.

4. Discussion

Processing affective features has been proposed to contribute to the semantic representation of conspecifics at the level of social groups (Piretti et al., 2015; Rumiati et al., 2014). The present study aimed at testing this claim by assessing whether priming affective information facilitates an evaluative response towards social groups relative to objects. To achieve this aim, we tested the hypothesis that deciding whether the name of a social group is positive or negative, compared to that of an object, is faster following the presentation of semantically unrelated but affectively congruent positive and negative primes. The behavioral data supported the hypothesis that participants would

show a significant affective priming effect for social group names regardless of their valence, but not for objects. The priming effect was mostly driven by faster evaluative responses to affectively congruent targets (i.e., positive priming), compared with incongruent and neutral conditions, while no difference was observed between these two.

As argued in previous priming studies, the facilitation found could be due to congruent primes inducing a greater motor readiness to press the corresponding response key (Damian, 2001), rather than being the result of a spreading of activation within an affective/semantic network (De Houwer & Randell, 2002; Fazio, 2001; Spruyt, Hermans, Houwer, & Eelen, 2002). However, the presence of a categorical difference in the magnitude of the affective priming effect allows us to dismiss this interpretation of the findings. As the response priming might have partially driven the main priming effect by facilitating congruent motor responses, we argue that the categorical difference is likely to have emerged from an unequal spreading of activation in the affective/semantic network, from which social groups received a greater benefit. In addition, our findings allow us to argue that the spreading of activation in such network is not the same for all concepts, but it applies to some concepts more than to others, and might depend on the relevance of the affective features for their representation: this argument would explain why social group targets benefit from this spreading of activation more than the nonsocial category targets do. On the other hand, the lack of an affective priming effect for nonsocial category targets is in accordance with previous reports failing to find such effect with concrete words, mostly denoting object names (Yao & Wang, 2013; but see also De Houwer & Randeil, 2002).

As to the ERP findings, we observed a significant category- and valence-specific effect for positive social groups in the 700-900 ms time window, with congruent trials eliciting a greater positivity relative to incongruent trials in the right-frontal region. This finding is however not in line with previous affective priming studies, reporting a greater amplitude in the incongruent condition from an LPC, and associating it with affective integration (Herring et al., 2011). In particular, the

scalp distribution of the effect suggests a it may reflect a different late right frontal component. Taking also into account its latency, the effect may indicate a sustained post-retrieval processing, indeed previously attributed to a greater motivational relevance of emotional stimuli (Ally & Budson, 2007; Van Strien, Langeslag, Strekalova, Gootjes, & Franken, 2009). This increased processing for positive social groups may have thus been triggered by the presentation of the preceding positive primes. To this end, the effect could also be interpreted as an anticipated late peak for primed as opposed to unprimed positive social groups. Both interpretations would be in line with previous findings suggesting that positive primes facilitate associative processes by increasing accessibility to associations (Storbeck & Clore, 2005). Social groups might thus receive a greater benefit from such increase in the affective semantic domain.

In line with recent studies showing that the absence of a strong modulation of the semantic relationship between primes and targets does not elicit N400 effects (Herring et al., 2011), we failed to find a significant affective priming in the N400 time window. Although the congruent primes were semantically closer to the targets compared to incongruent primes, this difference may have indeed not been enough pronounced, suggesting that the results mostly reflect a modulation of the affective relationship between targets and primes, rather than a semantic one. This claim is also supported by the presence of a significantly stronger effect for social relative to nonsocial targets, despite the two categories not differing in their semantic distance from the congruent and incongruent primes. It is also worthy to mention that the lack of such significant effect could be simply related to statistical reasons. Even if an effect is true, as an N400 in affective priming may be, it is not expected to be replicated in every experiment (Lakens & Etz, 2017).

An additional explanation as to why we found no effect for negative social groups in both ERP components, despite the presence of a behavioral effect, could be associated to differences in familiarity between positive and negative targets. Since the target semantic category matchings were conducted within each valence, it is possible that the expected effects for negative social categories

did not emerge because of the overall lower familiarity of negative nouns (see Table 1). Previous studies indeed showed that ERP effects are influenced by familiarity in priming paradigms (Voss, Lucas, & Paller, 2010). Even if this was the reason for the lack of an effect for negative social groups, the lack of an equivalent LPC effect for positive nonsocial categories still allows us to argue in favor of a greater reliance on affective features by positive social groups.

The greater influence of affective priming towards social groups could be also explained by their relatively greater ratings of abstractness. Indeed, the same claim about a greater reliance on affective features has also been attributed to more abstract words in general because of the lack of a sensory component (Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011; Vigliocco et al., 2014). Although social groups are on average rated as relatively more abstract than objects, first, there is no indication of the presence of a direct correlation with a word's abstractness and its reliance on affective processes, and second, they do not qualify as abstract words, as they do not lack sensory properties. Indeed, social group names as concrete stimuli are also found in several studies assessing the differences between the affective processing of abstract and concrete concepts (e.g., Newcombe, Campbell, Siakaluk, & Pexman, 2012; Yao et al., 2016; Yao & Wang, 2013). Moreover, if the greater reliance on affective features would be explained by an increased abstractness, we would expect patients showing deficits for abstract concepts to also be impaired in processing social groups, and vice versa, however, we are not aware of studies examining such dissociation. To clarify this issue, we suggest that future studies should employ equally representative pictures depicting social groups and objects as target stimuli, and test whether the same differences in affective priming emerge.

In conclusion, by showing a facilitated access to the affective component of their semantic representation following affective priming, we can support the claim that representation of conspecifics relies to a greater extent on the processing of affective features (Piretti, 2017; Rumiati et al., 2014). These results contribute to explaining previous findings associating the damage and stimulation of areas commonly involved in processing affective information with the specific

representation of social groups (Piretti et al., 2015; Suran, Rumiati, et al., 2019). Although the reason and the nature for such reliance is still debated (Meteyard et al., 2012), we propose that the greater association with affective processing when representing social groups could derive from the repeated engagement in empathetic (Barsalou, 2008) and evaluative responses (Piretti et al., 2015) when the meaning of these concepts is acquired and rehearsed.

CHAPTER 4

Study 3: Temporal dynamics of cross-modal affective representations in social and nonsocial categories

Abstract

Affective features have been suggested to play a predominant role in social group representations. Although studies identified overlaps between affective processing regions and social groups representations, not much is known about the temporal dynamics of their interaction. In the present study, I used multivariate decoding of electroencephalography (EEG) data from an evaluative task to track affective representations as a function of stimulus semantic category. Pattern classifiers were trained at distinguishing between positive and negative valence concepts presented as words and then tested on the same concepts presented as pictures. In contrasting performance between social groups and objects, the results revealed a stronger affective decoding for social groups in both an early and late time windows, coinciding with delayed evaluative responses. The present findings provide initial evidence of the presence of categorical differences in affective temporal dynamics, and point to an increased complexity in social group affective/semantic representations.

1. Introduction

Human lesion studies suggest that social categories such as social groups might be represented in an independent brain network from non-social categories (Piretti et al., 2015; Rumiati et al., 2014). Moreover, differently from nonsocial categories, the representation of social groups seems to give affective features a significantly greater weight. In a TMS study, we recently provided evidence that

stimulating the inferior frontal gyrus, commonly associated with the processing of negative affective features, speeds up the categorization of negative social group names. That is, when participants were presented with negative social group names such as "butchers", as opposed to the positive category of "musicians", TMS speeded up categorization responses to the former, thus suggesting a link between affective processing and semantic processing (Suran, Rumiati, et al., 2019). Moreover, using an affective priming paradigm, we found greater facilitation in evaluating social groups relative to objects following the presentation of semantically unrelated but affectively congruent primes. This suggests that the elaboration of affective information enhances the subsequent processing of social groups (Suran, Arcara, Piretti, & Rumiati, 2019). This set of findings is consistent with theories of semantic memory positing a central role of affective features in the representation of social groups (Lambon Ralph et al., 2016; Mahon & Caramazza, 2011; Simmons & Barsalou, 2003). As for how this central role of affect may be implemented anatomically, previous neuroimaging studies identified activation in regions selectively responding only to affective information related to social groups, as well as areas responding to both social groups and nonsocial categories but showing an increased response to the former (Norris et al., 2004). Even within regions processing both types of concepts, it is possible that the greater weight of affective information for social groups may be also reflected in the temporal dynamics of its processing, an aspect that fMRI, given its relatively low temporal resolution may have been unable to capture. To overcome such limitation, in the present study we used the temporal resolution of electroencephalography (EEG) to study whether differences between social and non-social are present in the temporal dynamics of processing their affective features. Additionally, to control for possible confounds of modality-specific mechanisms, we investigated the neural correlates of affective processes common to both visual and lexical inputs.

According to one conceptualization of how semantic knowledge might be organized in the brain, affective features are independent from features that rely on other modalities (Lambon Ralph et al., 2016). A first characteristic of affective features is that they do not possess a dedicated input

modality. This implies that, to study affect, one needs to rely on stimuli, such as words or pictures, which, despite eliciting the same affective processes, are also expected to rely on distinct modality-specific brain regions. For this reason, if one were to identify differences between two categories in affective features processing, it is often not possible to rule out the possibility that the difference is due to other properties of the stimulus, such as visual features in pictures that cannot be fully matched (Fairhall & Caramazza, 2013; Leonardelli, Fait, & Fairhall, 2019). By jointly using the information from two distinct input modalities, multivariate pattern analysis (MVPA) of neural information allows to overcome this limitation by individuating cognitive processes that are shared by different input modalities via cross-decoding (Grootswagers, Wardle, & Carlson, 2017; King & Dehaene, 2014). To this end, for each time point from the target word onset, a classifier was trained to distinguish valence-dependent neural patterns, and then tested on recognizing them in pictures representing the same concepts (King & Dehaene, 2014). Each classifier trained on words was tested on all time intervals from picture presentation, lasting up to 1 second post stimulus-onset. The resulting time generalization analysis allowed us to study the changes in valence-specific brain patterns, and to directly compare them between word and picture targets in the temporal domain (King & Dehaene, 2014). To this end, we asked participants to complete an evaluative task in which they were required to explicitly focus on the affective content of the stimuli while their brain responses were being recorded.

Affect-driven effects common to both words and pictures have been reported before in early and late time windows in studies adopting univariate analysis approaches. For example, singling-out negative information (associated with reactions to threat) can occur as early as in the 165-195 ms time window after target word onset (D. Zhang et al., 2014), while a later time window (~450 ms post-stimulus) has been argued to reflect the automatic processing of the polarity of affectively charged words (D. Zhang et al., 2014), the presence of an affective content of greater arousal (Hinojosa, Carretié, Valcárcel, Méndez-Bértolo, & Pozo, 2009), as well as the start of a more

controlled, context-dependent processing of affective information (Cunningham, Espinet, Deyoung, & Zelazo, 2005). In studies using picture stimuli, the presence of valenced as opposed to affectively neutral content is found in an early (150-180 ms) time window, whereas the affective polarity of such stimuli, as for words, is evidenced by a later component (~450 ms, Zhu et al., 2015). If such temporal and activation similarities reflect the involvement of shared underlying processes, we expected to record modality-independent valence effects both in early (~150 ms) and later time windows (~450 ms). Based on the hypothesized interaction between the semantic category and affective processing, we expected an earlier and better modality-independent decoding when processing social compared to nonsocial categories at all valence processing stages.

2. **Method**

2.1 *Participants*

Twenty-one healthy participants (16 females, age range: 21 – 33 years) took part to the study for monetary compensation. The inclusion criteria consisted in speaking Italian as first language, self-reported right-handedness, and a minimum of 95% conformity in evaluating the target word stimuli according to the valence manipulation based on an online survey. The study protocol was approved and carried out in accordance with the recommendations of the local Ethics Committee, and in accordance with the Declaration of Helsinki. All subjects gave written informed consent prior to participating.

2.2 *Materials*

A total of 40 nouns of social groups (N = 20) and objects (N = 20) was selected from a larger database rated by a different sample of 12 subjects and used as target word stimuli. Each semantic category was equally split in a positive and negative subset of 10 elements each, differing significantly in their

valence ratings (see Appendix 3). Within the positive and the negative valence stimuli, social group and object nouns were matched according to their average ratings of valence, arousal, familiarity, and length (see Table 1).

To exclude the possibility of categorical differences in RTs being due to differences in valence ambivalence (Cunningham et al., 2004) - as participants might take longer at considering both positive and negative features - we also administered an additional online survey to 40 participants (33 females, age range: 21-35). The survey presented participants with the target stimuli in a random order and asked to rate on a 9-point Likert scale the degree of both the positivity and negativity of each concept (0 = not at all, 9 = very much).To calculate ambivalence from the survey data, we computed for each concept an index of ambivalence using the equation of the Gradual Threshold Model (Priester & Petty, 1996) and subjected it to a 2 (Category: Social vs. Nonsocial) x 2 (Valence: Positive vs. Negative) ANOVA. The analysis of the resulting ambivalence scores revealed no effect of Category [$F(1, 36) = .60$, $p = .44$, $\eta_p^2 = .02$], a main effect of Valence [$F(1, 36) = 44.71$, $p < .001$, $\eta_p^2 = .55$], with negative concepts being overall more ambivalent, and no significant interaction [$F(1, 36) = 3.03$, $p = .09$, $\eta_p^2 = .08$].

A sample of 40 pictures representing the same concepts expressed by the selected words was retrieved from Google Images and matched across category and valence polarity in their representativeness ratings, collected from a different sample of 14 participants (see Table 1). Where applicable, the faces of the depicted individuals were manually blurred using the GIMP 2 software (Kimball, Mattis, Natterer, & Neumann, 2013). Additionally, pictures were also matched in low-level visual features (all ps > .05). These included average luminance, amount of red, green and blue, and average spatial frequency, extracted through a custom Matlab script (adapted from Blechert, Meule, Busch, & Ohla, 2014).

Variables	Valence	Arousal	Familiarity	Length	Ambivalence	Pic. represent.
Social Positive	6.14 ± .71	3.91 ± .62	3.46 ± 1.02	7.4 ± 1.17	1.82 ± 2.06	6.69 ± .26
Object Positive	6.42 ± .47	4.00 ± .69	3.8 ± 1.04	6.8 ± 1.48	0.45 ± 1.01	6.64 ± .16
Social Negative	3.30 ± .81	3.88 ± .57	3.63 ± .75	7.9 ± .52	4.50 ± 2.10	6.64 ± .27
Object Negative	3.07 ± .48	3.81 ± .57	2.94 ± 1.13	7.1 ± 1.52	5.03 ± 1.44	6.70 ± .20
ANOVA	$p < .001$	$p = .92$	$p = .26$	$p = .37$	$p = .09$	$p = .87$

Table 1. Descriptive statistics for selected words and pictures, divided the combination of semantic category and valence polarity. Means of word length and ratings of valence (1, unpleasant to 7, pleasant), arousal (1, not arousing to 7, very arousing), familiarity (1, unfamiliar to 7, familiar), ambivalence, and of picture representativeness (1, not representative to 7, very representative). On the bottom row, *p*-values of one-way ANOVAs on each factor, using as independent variable the combination of category and valence polarity.

To ensure an equal speed in semantic access between categories, we collected a measure of accessibility by devising a simple categorisation task. In this task, an independent sample of 20 participants (11 females, age range: 25-35) was asked to provide speeded responses to the selected stimuli. The task required to categorize the targets based on whether they represented people or objects, using the same timings of the main task (see Fig. 1). The analysis of the reaction times showed no main effect of category [$F(1, 19) = .53$, $p = .48$, $\eta_p^2 = .03$], for which participants took on average the same time to categorize social groups and objects, regardless of modality and valence (both $ps > .05$).

2.3 Procedure

Following the montage of the EEG cap, participants were seated in an acoustically isolated room and asked to fixate an "X" in the center of the screen for 3 minutes to record resting-state EEG activity.

After this period, they completed an evaluative requiring to indicate via button press whether the concept represented by the target stimulus was associated with a pleasant or unpleasant feeling. A PC running PsychoPy (Peirce, 2007) controlled the presentation of the stimuli and the recording of responses. Stimuli were projected on a gray background via a 19" LCD monitor with resolution of 1280*1024 pixels and a screen refresh rate of 60 Hz. During each trial, a fixation cross was presented for 700 ms, followed by a 200 ms blank screen and by the presentation of the target stimulus (300 ms). The target was then replaced by another blank screen that lasted for 1700 ms, giving participants a total of 2000 ms to respond from target onset (see Figure 1). The evaluation was given via button press by using the index finger of each hand placed over the 'f' and 'j' QUERTY keyboard buttons. The intertrial interval (ITI) was jittered between 800 and 1200 ms at 100 ms intervals, and presented a white fixation cross on a gray background. Participants were instructed to try to restrict their blinking to the ITI period to reduce the number of artefacts. The target stimuli consisted of images or words representing social groups or objects of positive or negative valence, presented in a random order. Participants first completed a block of 10 practice trials, followed by 12 test blocks of 40 trials each. Single blocks contained either pictures of words in an alternating, counterbalanced order, and were separated by self-paced breaks.

The experiment consisted of a 2 (Category: social group vs. object) x 2 (Valence: positive vs. negative) x 2 (Modality: picture vs. word) within-subjects design, with response time (RT), accuracy, and EEG voltage as the dependent variables.
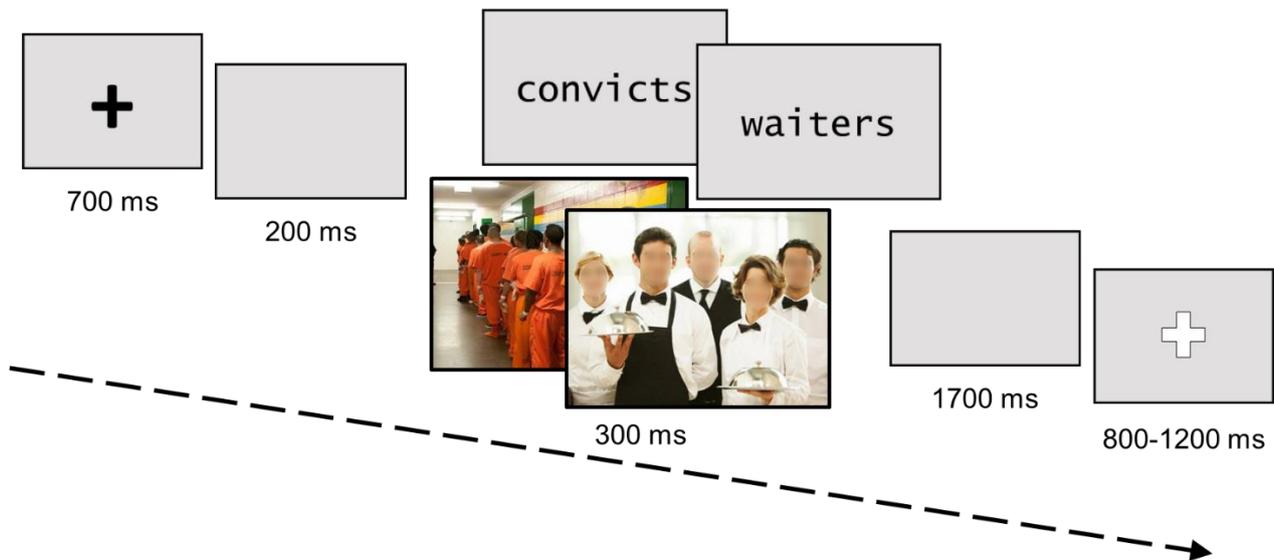
Figure 1. Temporal progression an evaluation trial with examples of social groups of negative and positive valence, represented as words or pictures (only one of the stimuli was presented each time). Participants were instructed to evaluate the target starting from its onset, and to try not to blink until the appearance of the white cross (ITI).

2.4 *Electrophysiological recordings*

A set of 64 Ag/AgCl active electrodes connected to a BioSemi Active-Two amplifier system were mounted on an elastic cap according to the International 10/20 system to record the continuous neural signal by means of ActiView acquisition software (Biosemi, Amsterdam, Netherlands). Electrode offsets were kept between ±20 mV, while the signal was sampled at a rate of 1024 Hz with a 24-bit resolution. A common mode voltage based on the ActiveTwo's CMS/DRL feedback loop was used for analog-to-digital conversion of recorded voltages for each electrode (cf. to https://www.biosemi.com/faq/cms%26drl.htm). Anti-aliasing filters were used and data were band-pass filtered between 0.01–100 Hz during data acquisition.

EEG data preprocessing was performed using the Brainstorm software (Tadel et al., 2011). First, the EEG recordings were downsampled offline at 125 Hz and band-pass filtered (0.05 – 40 Hz). Bad electrode channels were removed upon visual inspection, as well as movement artefacts. Eye

blinks artefacts were removed through independent components analysis (ICA; (Makeig et al., 1996).

Epochs containing artefacts other than eye movements were removed after visual inspection. One

subject (female, age 24) was excluded from all further analyses due to the presence of excessive

movement artefacts. The data were then epoched from -200 to +1500 ms relative to the onset of the

target and baseline-corrected from -200 ms to target onset.

2.5 *Behavioral data analysis*

Behavioral data were preprocessed and analyzed in R (R Core Team, 2016). Data from correct trials

within 2 SD from each participant's average were then log-transformaed to reduce the skew of the

distribution and subjected to a 2 (Category: Social vs. Nonsocial) x 2 (Valence: Positive vs. Negative)

x 2 (Modality: names vs. pictures) repeated-measures ANOVA. Follow-ups to significant interactions

consisted of Bonferroni-corrected contrasts. Only results involving interactions with the Category

factor and yielding significant post-hoc contrasts are reported.

2.6 *ERP data analysis*

The time windows for the main components were identified from the existing literature

(Hinojosa, Carretié, Valcárcel, et al., 2009; D. Zhang et al., 2014; Zhu et al., 2015) and the visual

inspection of average peak amplitudes. In order not to lose statistical power in quantifying effects

over several electrodes, eleven regions of interest (ROIs; see Figure 2) were then created by averaging

the amplitude of the respective electrodes (as in Hinojosa, Carretié, Valcárcel, Méndez-Bértolo, &

Pozo, 2009). The resulting ERPs were subjected to two 2 (Category: Social vs. Nonsocial) x 2

(Valence: positive vs. negative) x 2 (Modality: word vs. picture) x 11 (Cluster: [OC, LP, RP, LC,

RC, LF, RF, FP, MF, MC, MP]) repeated measures ANOVAs, one for each component's time

window (150 - 200 ms, 400 - 700 ms and 700 - 1000 ms). Where necessary, the degrees of freedom

of the *F* ratios were adjusted using the Greenhouse–Geisser epsilon correction. Follow-ups to significant interactions including Valence consisted of Bonferroni-corrected contrasts of Valence effects. Only interactions yielding significant follow-up contrasts are reported.
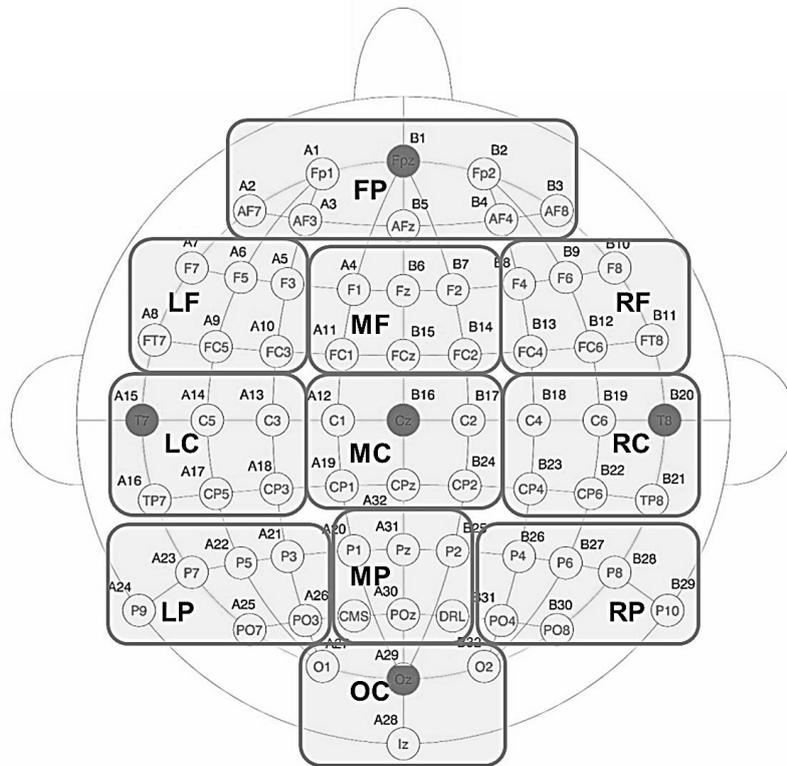
Figure 2. Clusters of electrodes grouped for statistical analysis of ERPs. OC, occipital, LP, left posterior, RP, right posterior, LC, left central, RC, right central, LF, left frontal, RF, right frontal, FP, frontopolar, MF, middle frontal, MC, middle central, MP, middle posterior.

2.7 Decoding analysis

Multivariate classification analyses were performed using the CoSMoMVPA analysis package (www.cosmomvpa.org) (Oosterhof, Connolly, & Haxby, 2016) implemented in MATLAB. Classification was performed separately for every 8 ms time bin using linear discriminant analysis (LDA) classifiers. These were trained to discriminate the patterns of activation across EEG sensors for the two valence conditions in one subset of the data, and tested on another.

Cross-decoding was conducted separately for Social and Object categories. Training data consisted of trials from the Words condition, after which classifiers were tested on trials from the Pictures condition. In this way, the decoding performed by the classifier on pictures derived from the identification of the same patterns of valence differences that had been learned on word stimuli. The correctly identified patterns were thus common to both modalities. To increase the reliability signal-to-noise ratio and the of the data for the classifier, for each participant and experimental condition separately, five averaged trials were created. As the number of correct trials differed between participants and within conditions, the averages containing roughly the trials corresponding to each experimental block/run, with the constraint that no average was derived from more than one trial more than the other averages.

Given the consistency of the response mappings within participants, to avoid the confound of decoding motor responses rather than valence differences (Grootswagers et al., 2017), classifiers were trained and tested on the combined data from pairs of participants with opposite mappings. The classification accuracy of each participant was calculated by averaging the performance of all the classifications containing that participant's data (i.e., 10 for each subject). The percentage of correct predictions of the classifier was used as index of classification accuracy. The classification was generalized across train and test times, for which it was repeated for their every possible combination, leading to a classification accuracy map of 125 x 125 points (i.e., 1000 ms x 1000 ms with 125 Hz resolution) for every comparison in each participant. Individual maps were smoothed with an averaging box filter of the size of 3 x 3 time points (i.e., 24 ms in both training and testing time).

2.8 Statistical testing

To identify time-periods presenting above chance classification accuracy, we used threshold-free cluster-estimation procedure (Smith & Nichols, 2009) with default parameters from the CoSMoMVPA package (Oosterhof et al., 2016), using multiple comparison correction based on a sign-permutation test (with null distributions created from 10,000 bootstrapping iterations). To reveal

where decoding performance was significant, the threshold on the statistical maps was set at $Z > 2.57$ (i.e., $p < .005$). The procedure was first applied within each category against the mean of 0.5 decoding accuracy and then used to check for significant differences in contrasting the two.

3. Results

3.1 Behavioral results

Reaction times analysis showed a main effect of category [$F(1, 19) = 128.83, p < .001, \eta_p^2 = .87$] with faster responses to non-social than to social categories. No other significant main effects were present. Significant first order interactions emerged between category and valence [$F(1, 19) = 18.24, p < .001, \eta_p^2 = .49$], and between category and modality [$F(1, 19) = 47.94, p < .001, \eta_p^2 = .72$], with the former indicating how the difference in RTs between social groups and objects was significantly greater in the positive relative to the negative valence targets ($p = .001$), and in pictures relative to words ($p < .001$). Last, a post-hoc analysis of the significant second-order interaction between category, valence and modality [$F(1, 19) = 8.37, p = .009, \eta_p^2 = .31$], revealed that social groups were evaluated slower relative to objects in all valence and modality combinations (all $ps < .001$) but the negative word one, where there was no significant categorical difference ($p = .60$). (see Figure 3).
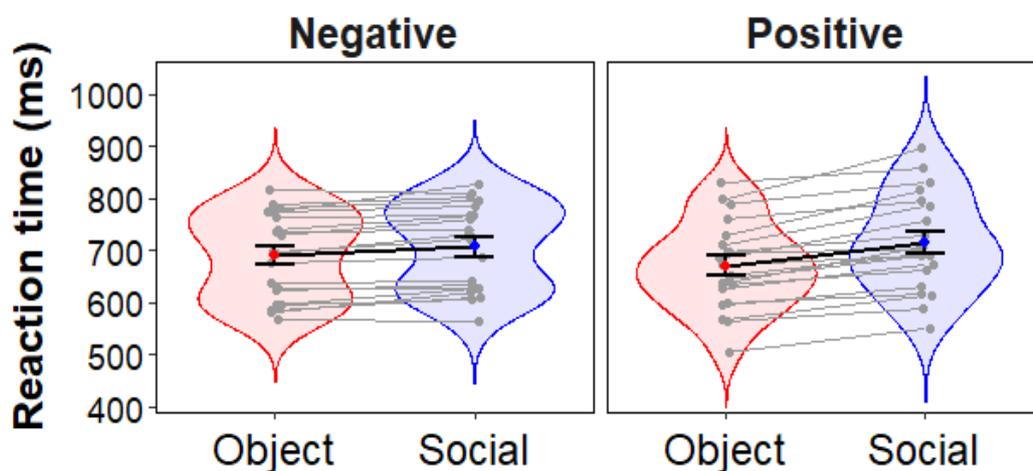

Figure 3. Violin plots displaying single participant mean RT distribution and group mean RTs for Object and Social category targets divided by valence. Error bars represent ± 1 SE.

3.2   ERP results

150 – 200 ms. The analysis of the P1 time window resulted in a significant three-way interaction between Valence, Category and Cluster [$F(3.19, 60.53) = 3.28$, $p = .025$, $\eta_p^2 = .15$], and a significant four-way interaction between Valence, Category, Modality and Cluster [$F(2.61, 49.61) = 7.54$, $p < .001$, $\eta_p^2 = .28$]. Follow-up analyses for the word modality evidenced a significant Valence effect in the right frontal cluster ($p < .05$), with a greater amplitude for negative as opposed to positive stimuli ($p < .05$). The effect was only significant for Social groups ($p < .05$), while the effect for and difference from Objects was not significant (both ps $> .1$). No effects were present for pictures in the same cluster. For pictures, a significant Valence effect was present for Objects in the occipital, left posterior and left anterior clusters (ps $< .05$), with no effect for Social groups (ps $> .05$), and a significant difference between categories (ps $< .05$). A Valence effect for both categories was present in the right posterior and the mid frontal clusters (ps $< .05$), presenting also a significant difference between categories ($p < .05$). This difference was characterized by a greater amplitude for negative vs positive stimuli for Objects and a greater amplitude for positive vs negative stimuli for Social groups.

400 – 700 ms. A significant interaction was present between Valence, Category and Cluster [$F(3.71, 70.49) = 3.27$, $p = .02$, $\eta_p^2 = .15$]. At the single cluster level, a marginally significant Valence effect, with a greater amplitude for positive stimuli, was only present for words in the central right electrodes ($p = .07$). In the picture modality, the same cluster also presented a significant Valence effect for Objects ($p < .05$), where the difference with Social groups was marginally significant ($p = .08$).

700 – 1000 ms. Significant interactions between Valence, Modality and Cluster [$F(3.84, 72.98) = 2.54$, $p = .049$, $\eta_p^2 = .12$], Valence, Category and Cluster [$F(4.36, 82.75) = 2.62$, $p = .036$, $\eta_p^2 = .12$], Valence, Category and Modality [$F(1, 19) = 6.87$, $p = .017$, $\eta_p^2 = .27$] were present in the latest time window. These were accompanied by a marginally significant four-way interaction involving

Valence, Category, Modality and Cluster [$F(3.24, 61.57) = 2.51$, $p = .063$, $\eta_p^2 = .12$]. Follow-up analyses showed a significant Valence effect in the right posterior cluster for words ($p < .05$) but not for pictures ($p > .1$). Within this cluster, the effect of Valence was significant for Social groups ($p < .05$) but not for Objects ($p > .1$), with the difference between the two categories also being significant ($p < .05$). A marginally significant Valence effect in the left central cluster for words only ($p = .07$), with a greater amplitude for negative stimuli, was driven by a significant effect for Social groups ($p < .05$), with no effect for Objects and no significant difference between categories (both ps $> .1$). In the right central cluster, the main effect of Valence ($p < .05$) was accompanied by two marginally significant effects in both categories (ps $= .07$), in all cases displaying a greater amplitude for positive relative to negative word stimuli. For pictures, only a marginally significant Valence effect emerged for Social groups in the left posterior region ($p = .09$).

3.3   *Cross-modal valence decoding in social groups and object categories*

Valence cross-decoding results for social groups evidenced a first significant decoder performance period at 100-150 ms, followed by a second interval going from ~450 to 1000 ms (see Figure 3A & 3C). In the case of object categories, only one significant window was present between ~400 and 700 ms (see Figure 3B & 3C). When contrasting decoder performance between categories, a significantly greater decoding performance was found for social groups in two time windows - an early window going from 50 to 150 ms and a late window starting at 750 ms up to 1000 ms (see Figure 3D).

Time-generalisation results evidenced how decoding did not differ based on modality, as all significant regions laid on the diagonal (see Figures 4A & 4B). Additionally, a sustained pattern of neural activity across modalities was present for social groups in the ~450-1000 ms time window (King & Dehaene, 2014), as evidenced by the square-shaped region of significance (see Figure 4A). This pattern was unique to social groups and was not present for objects (see Figure 4B).
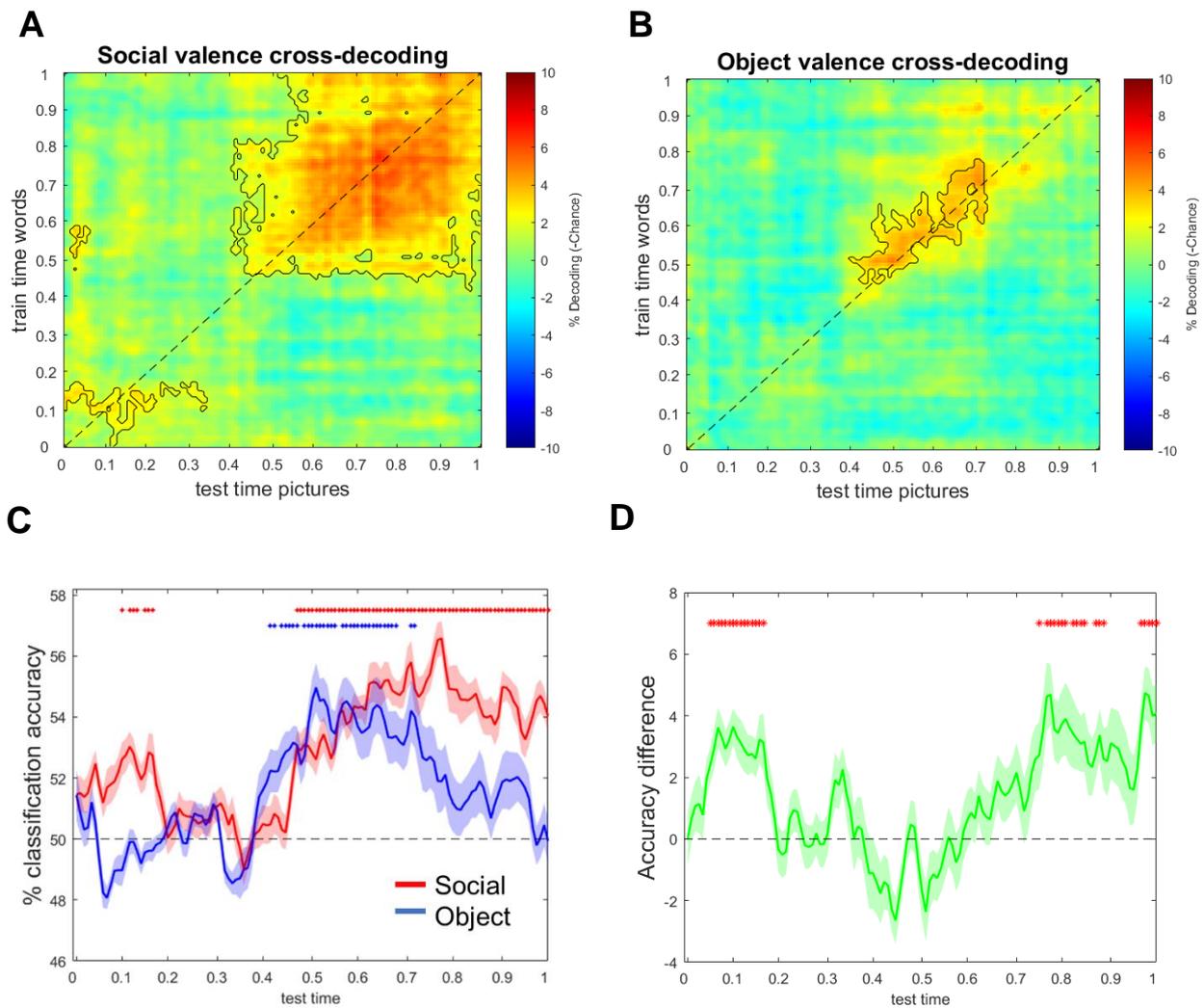
Figure 4. A) Time-generalization plots resulting from cross-decoding valence of social category and object category targets **(B)**. Black contours represent above chance decoding accuracy (*p* < .001). **C)** Classification accuracy across the time-generalization diagonal for social and object category targets, and **(D)** their difference. Asterisks indicate above chance decoding accuracy time intervals (*p* < .001).

4. Discussion

With the present study we aimed at identifying the distinctive patterns of brain activity associated with the affective processing of social groups and compare it to the one of nonsocial categories. Following previous neuropsychological studies (Piretti et al., 2015; Rumiati et al., 2014; Suran, Rumiati, et al., 2019) and theoretical propositions suggesting a greater relevance of affective features

in representing social categories (Lambon Ralph et al., 2016; Mahon & Caramazza, 2011; Simmons & Barsalou, 2003), and following studies showing both common and distinct anatomical bases for processing social and non-social affective information (Norris et al., 2004), we expected to find categorical differences in the temporal dynamics of affective processing, prioritizing the decoding of valence in social groups compared to nonsocial categories. To rule out possible confounds generated by the input modality and thus isolate affective evaluation, we cross-decoded the effects between two different input-modalities (words and images) using MVPA by training of classifiers using neural data associated with one modality (e.g. words) and test on data from the other modality (pictures). Additionally, we applied time generalization to study the evolution of the affect-specific neural code while locating its processing in the temporal dimension of the two input modalities. Our results showed significant differences in decoding affective features between social and nonsocial category targets. This difference was present in both early and later time windows, and was also reflected in overall response times as longer responses to social groups, but just when participants were required to make affective judgments.

Significant decoding performance appeared first at 100 ms post-stimulus that was unique to social categories. This earlier decoding might indicate the prioritization of affective processing for Social groups over Objects. This effect might also be linked to the early effects found by previous ERP studies using words (D. Zhang et al., 2014) and pictures (Zhu et al., 2015) which, although here not detected across modalities by univariate analyses, was captured by the greater sensitivity and earlier detection attributed to MVPA (Grootswagers et al., 2017). In the second time window of significant decoding, above-chance accuracy started approximately at ~400 ms for nonsocial categories and at ~450 ms for social groups, reflecting the delay found in the behavioral response. This time interval overlaps with the emergence of the late positive components modulated by the affective content of both words and pictures (D. Zhang et al., 2014; Zhu et al., 2015), and it represents the only time interval in which a significant difference between negative and positive stimuli was

present across modalities in the same scalp region when analysing ERPs. While for Objects significant decoding in the second time window lasted until ~700 ms, significant affective decoding in Social groups extended to 1000 ms post-stimulus and was significantly greater than the one for Object categories from ~750 ms on. In previous ERP research, effects occurring in this time window are associated with a greater sustained processing of emotional stimuli (Citron, 2012) and an enhanced motivational significance (Y. Liu, Huang, McGinnis-Deweese, Keil, & Ding, 2012; Schupp et al., 2000; Tempel et al., 2013), which might be intrinsically greater in social groups due to the presence of conspecifics. Although more valence effects were identified with the univariate analysis for Social groups relative to Objects, none of these were captured across modalities and in the same regions. The claim of a more sustained processing of Social groups is also supported by the results of the time generalization, presenting a square-shaped pattern of significance, indeed associated with sustained patterns of brain activity in the decoded dimension (King & Dehaene, 2014).

Reaction times to social groups evidenced a slower evaluative response in comparison to nonsocial categories. The lack of such differences in ambivalence and categorization times of the same targets allows us to exclude the possibility that the disparity was due to social group stimuli presenting stronger conflicting (i.e., nondominant) affective associations (Cunningham, Johnson, Gatenby, Gore, & Banaji, 2003; Cunningham et al., 2004), or having a delayed semantic access, respectively. As suggested by previous neuroimaging studies evidencing slower responses in categorizing phrases describing social interactions versus nonsocial actions (Wood, Romero, Makale, & Grafman, 2003), we speculate that it is possible that participants activated more complex representations for social relative to nonsocial categories to make affective decisions. Evaluative processes generally involve a more complex selection of information in comparison to categorization (Cunningham et al., 2003), thus task demands might interact with the sociality of the semantic categories of target stimuli. For example, the present data might reflect how the evaluation of social groups requires accessing all relevant representations of the behaviors commonly associated with

them, while in the case of nonsocial categories, just the representation of their common usage may have sufficed.

In conclusion, our results suggest that processing of affective features comes with distinct temporal dynamics for social groups and nonsocial categories, and represent further evidence of the interaction between semantic and affective information in the temporal domain. An early decoding of affective features for social groups however was not associated with faster responses, as the behavioural findings have shown that subjects took longer in providing a behavioural response when evaluating social groups as opposed to objects. Interestingly, this effect seems to be independent from the timing of access to the more basic person/object categorical representations. It is indeed possible that a slower evaluative response to social groups was triggered by self-presentation concerns when required to provide judgments about other people following access to the categorical semantic information (Nosek, 2005), and thus requiring additional affective processing stages as opposed to objects. The same sustained processing of affective information of social groups was also evidenced at a neural level, with longer lasting classifier decoding accuracy in a late time window, going beyond the behavioural response. We suggest that the delayed response and longer decoding of affective features in social groups, possibly associated with more sustained post-semantic attentional processes due to more complex representations, supports the hypothesis of a greater relevance of their affective features compared to nonsocial categories (Rumiati et al., 2014). As our conclusions are derived from a setting explicitly requiring subjects to focus only on affective content, it is possible that its increased weight was specific to the present task demands. By directly comparing the neural signatures resulting from different tasks, future studies may focus on elucidating which aspects of affective feature processing in social groups interact with experimental demands and which ones occur independently from it, being thus more intrinsic to their processing.

CHAPTER 5

General discussion

In the present thesis, I investigated the relationship between the representation of social groups and the processing of their affective features. I carried out this investigation by comparing social and non-social category stimuli in three separate studies. The general hypothesis, common to all three studies, was that affective features play significantly more weight in social group representations, in thus differing from non-social representations. More specifically, I expected this prominent role of affective features in social groups representations to be reflected in faster behavioural responses following the stimulation of an affective cortical region (Study 1), higher susceptibility to affective priming (Study 2), and a better decoding of affective valence from neural activity (Study 3).

Neuropsychological studies first suggested that the representation of social groups might be distinct from that of other semantic categories by reporting patients suffering from neurodegenerative disease (Rumiati et al., 2014). Although a specificity for conspecifics at the level of individuals was already known (e.g., Miceli et al., 2000; Thompson et al., 2004), by bringing the evidence to the level of social groups, the findings allowed for a more direct comparison with other semantic categories previously found to present distinct representational bases (for reviews, see Capitani, Laiacona, Mahon, & Caramazza, 2003, and Thompson-Schill, 2003). A possible explanation to the emergence of category-specificity for social groups has since been attributed to the presence of a special link with the processing of affective features, which would make up a significant part of social group semantic representations (Rumiati et al., 2014). Further evidence in support of such claim came from the study of the neural correlates of social group representations. By studying the brain areas commonly damaged in a cohort of patients showing lexical-semantic deficits for social groups, Piretti and colleagues (Piretti et al., 2015) identified the possible extent of the underlying neural network.

The authors found a co-occurrence of semantic deficits for social categories and widespread brain damage affecting to areas like the amygdala, the insula, and the inferior frontal gyrus, all previously associated with affective processes (Piretti et al., 2015). From a theoretical perspective, several models presented an argument associating social semantic representations with affective processes (Lambon Ralph et al., 2016; Mahon & Caramazza, 2011; Simmons & Barsalou, 2003), also calling for more evidence about the existence of such connection and its neural correlates (Lambon Ralph et al., 2016).

To fill this gap, in my dissertation I carried our three studies including computer-based tasks in conjunction with neuroscientific techniques. More specifically, tasks required subjects to categorize (Study 1) or to evaluate (Studies 2-3) lexical and picture stimuli of social groups and non-social categories. Each behavioural paradigm was paired with either brain stimulation (TMS; Study 1) or neural activity recordings (EEG; Studies 2-3). In the latter case, two analytical approaches were used to answer specific research hypotheses, one univariate (ERPs; Study 2), measuring differences in activation at specific neural components, and one multivariate (MVPA; Study 3), tapping into the study of the brain's temporal dynamics.

The results of the Study 1 show how the stimulation of the *pars opercularis* of the left inferior frontal gyrus (LIFGop), previously linked to affective processing (Kensinger & Schacter, 2006; but see also Leclerc & Kensinger, 2011), might influence how social groups are represented in the brain. This influence was indexed by participants' faster person/object categorization responses to social group names, without altering the speed of the behavioural responses to non-social categories. Additionally, these effects were confined to negative social categories, in line with the documented association between LIFGop activity and the processing of negative word stimuli (García-Pacios et al., 2017; Kensinger & Schacter, 2006; Leclerc & Kensinger, 2011; Van der Cruyssen et al., 2014). While my data do not allow us to fully map all the regions causally involved in processing affective and semantic content, however they do provide initial evidence on the contribution of an area not

previously considered in the domain of social/affective processing. Thus, by demonstrating that the effect of LIFGop stimulation is category and valence-specific, my results extend the *hub-and-spokes* model (Lambon Ralph et al., 2016), which placed valence features processing in the orbitofrontal cortex (OFC), as they identify an additional potential area contributing to the affective semantic representations of social groups. More in general, as the behavioural effect of the stimulation was limited to the negative valence social groups, the results also support the claim of a separability between areas processing positive and negative valence (Cacioppo et al., 1999) occurring at the third, semantic level of affective processing as described in Panksepp's theory (Panksepp, 1998, 2012).

In the Study 2, I found that when healthy participants engage in the processing of positive and negative affective content via affective priming, affectively congruent social group names presented shortly after to be evaluated are processed significantly faster than non-social category names. Moreover, this effect was present despite non-social category names being matched for valence ratings. This behavioural finding had a neural counterpart in a late right frontal positive component, suggesting that social group representations and affective information processing possibly have overlapping neural correlates. The priming effect fits the spreading of activation theories of affective priming, supporting the existence of an affective semantic node activated by the primes (Fazio, 2001; Fazio et al., 1995, 1986). Additionally, the categorical difference in the priming effect extends these theories, demonstrating how the strength of such spreading might vary as a function of the target word category. As an equal spreading of activation from an affective node (elicited by affective priming) to all affectively congruent concepts has been deemed unlikely (Klauer & Musch, 2003), my findings suggest that the spreading of activation might be a more selective phenomenon than previously thought. When controlling for differences in other word properties and for differences in the semantic distance with the primes, as done in Study 2, the strength of the affective priming itself could be adopted as an index of reliability on affective features by different concepts.

Unlike in Study 1, in Study 2 the greater affective priming response to social groups (vs objects) occurred regardless of their valence. Nevertheless, at the level of brain responses, the neural correlates of affective priming emerged only for positive social groups. The effect was recorded in a later time window of the ERPs, and was strongest in the frontal right electrodes. The distribution of the response, its latency, and the greater amplitude for congruent trials together may indicate the presence of sustained processes for positive social groups following positive primes, potentially related to an increased post-retrieval motivated attention. Although this effect was valence-specific, the presence of differences in familiarity between positive and negative words overall, shown having an influence in later ERPs (Voss et al., 2010), does not allow to confidently exclude the possibility that negative social group targets might have elicited the same effect under matched conditions. Indeed, in Study 2, the familiarity matching was only done between categories, and, in order to maintain a greater amount of stimuli, not between positive and negative words.

In Study 3 I used both lexical and pictorial stimuli to investigate the neural correlates of affective processing independently of the input modality. The approach allowed me to avoid the limitations of Studies 1-2, in which affective processing of social group concepts was investigating only with lexical inputs. As conceptual processing of valenced stimuli was previously argued to activate an affect processing node in the semantic network, this should occur independently of the input modality. The results revealed how a significant decoding of affective valence occurs for social groups but not for non-social categories at an early post-stimulus time interval (100-200 ms). Whereas an intermediate time window (450-750 ms) revealed an overlap in above chance decoding performance of the two categories, above chance performance was present only for social groups in a later time window (750-1000 ms). Such pattern suggests that, compared to non-social categories, the neural decoding of affective information of social groups might be prioritized in early processing stages, going through a similar processing of non-social categories during intermediate stages, an then persisting in later stages. This finding might be due to the involvement of areas, such as the amygdala,

in which the processing of emotion has been shown to be boosted when originating from a social stimulus, and other areas, such as the medial prefrontal cortex (mPFC), selectively responding to emotional social stimuli (Norris et al., 2004), and the dorsolateral prefrontal cortex (DLPFC), activated during social judgments and impression updating (Hughes, Zaki, & Ambady, 2017). As the activity of the amygdala is regarded as automatic and thus occurring at early stages, it is likely that the initial decoding of social groups was driven by its activity. As for the later decoding, it is likely that more controlled evaluative processes came into play, in line with the reflective role of prefrontal areas during evaluations (Cunningham & Zelazo, 2007). These results are consistent with those of Study 2, where the priming effect for social groups was evident in the same late time interval, generated from electrodes situated approximately above the right DLPFC. It might thus be possible that both findings reflect a sustained attention to (Ally & Budson, 2007; Van Strien et al., 2009) or further updating of affective semantic content (Cunningham & Zelazo, 2007) in social groups, likely extending temporally beyond the forced-choice behavioural response. Unexpectedly, behavioural responses showed that participants' reaction times were relatively longer when evaluating social groups, whereas in the semantic categorization task I found no difference with non-social categories. As to this latter finding, it is possible that the representations of social categories possess significantly more complex affective information accessed by the cognitive system in evaluative contexts (Wood et al., 2003), although additional studies will be needed to empirically test such interpretation.

Future research should be also carried out to address a series of limitations present in my studies. First, as the brain stimulation might well affect additional areas connected to the one directly stimulated (Massimini et al., 2005), I cannot conclude that the LIFGop is the *only* area responsible for processing affective semantic features. To reach this conclusion a more extended mapping is required. As a starting step, given that the effects of the stimulation have been shown to extend to homologue areas on the opposite hemisphere (Fitzgerald, Fountain, & Daskalakis, 2006), it would be relevant to study whether the stimulation of the right IFGop leads to similar findings. As the brain

responses to primed social categories were located in a right frontal region above the DLPFC, which comprises the RIFGop, its stimulation may affect the processing of social categories of positive affective valence.

Furthermore, as in Study 1 I used only names as stimuli, it is not possible to know if the role of the LIFGop in processing affective information is specific to the lexical domain or whether its processing is amodal, and thus independent of the input modality (Fairhall & Caramazza, 2013; Leonardelli et al., 2019). The same observation applies to Study 2, as only one modality is taken into account. While Study 3 shed light on the correlates of affective processes common to two modalities, and captured effects in early and late processing stages, I cannot argue that the recorded positive/negative affective decoding occurs automatically; neither can I argue that it generalizes to other tasks, as it might only emerge in a setting requiring a forced binary evaluative response.

Studies 2-3, where participants performed an evaluative task, leave both unanswered the question about the automaticity of the recorded neural components. A first distinction can be made between automatic and controlled processes, respectively referring to spreading of activation, or early access, and task-dependent evaluation and updating(Cunningham & Zelazo, 2007). To disentangle these processes, future studies could adopt the same paradigms while manipulating the task requirements. Such manipulation could require participants to focus on the semantic category of the targets in a categorization task, or on more linguistic, 'shallow' aspects of the stimuli (lexical decision or naming). Although the effects of the primes may decrease in magnitude, it might be still possible to record how brain responses are affected by the interaction between the category of the targets and their congruence with the primes. The same approach can be taken in furthering the study of the temporal dynamics of affective valence processing, to assess whether the same decoder performance differences between social and non-social categories emerge during other tasks. Although an earlier decoding might persist, later stages may not be engaged as much as in the evaluative task used in the third study, as suggested by the lack of categorical differences in the behavioural control experiment.

Although only briefly introduced for exploratory purposes, the study of affective ambivalence may also be a promising venue for future research on categorical affective semantic representations (Cunningham et al., 2004). Despite the categorical differences in the bahavioral responses of the third study could not be related to higher levels of ambivalence for social groups relative to objects, the possibility that the consideration given to nondominant affective associations varies from one category to the other should not be excluded. Previous research has shown that affective ambiguity is more likely to affect reflective stages of processing, particluarly in tasks requiring a forced-choice affective evaluation, and correlating with activity in the ventrolateral prefrontal cortex (Cunningham et al., 2003). A comparion between social and nonsocial concepts based on the relationship between ambivalence and brain activity at single concept level may shed light on this possibility.

It should be noted that all studies have been primarily concerned with comparing social groups and objects, and a comparison with other categories for which selective deficits have been reported, like fruits/vegetables (Hart et al., 1983; Samson & Pillon, 2003) and animals (Blundo et al., 2006; Caramazza & Shelton, 1998), is missing. While the former could be excluded a priori, emotional responses elicited by animals may have a comparable magnitude to that of social groups. Nevertheless, even with such comparison, I would expect the same affect-driven category specific effects for social groups to emerge.

While the study of conspecifics at group level was justified by previous research (e.g., Rumiati et al., 2014), future studies could also further investigate the extent to which the affective processing of social groups differs from that of single individuals in terms of temporal dynamics and underlying areas. For instance, as access to social category knowledge has been shown to precede that of individual knowledge (Cloutier et al., 2005; Mason & Macrae, 2004), if a direct comparison between these two levels of specificity was made using a procedure similar to that of Study 3, I would expect valence effects be decoded faster for social groups. However, as in such case one would have to look

at a face to recognize the individual, the emotionality of the displayed facial expression might introduce an additional level of complexity that will have to be considered.

As a last note, while the framework I adopted focused on the positive/negative valence dimension, it is possible that the study of single emotions may uncover unique processing areas and neural components when representing social groups. For example, previous research on social cognition has shown that distinct brain areas are activated when social groups elicit emotions that are exclusively social (e.g., pity or envy) as compared to non-exclusively social (e.g., disgust; Harris, McClure, van den Bos, Cohen, & Fiske, 2007). Accounting for this social exclusivity factor may thus provide more insights about the degrees of separation between processing social and non-social affective knowledge.

In conclusion, my thesis provides novel evidence of a separability between social and non-social categorical knowledge in the affective domain. My results call for future replications to assess their robustness and also to overcome possible limitations with the use of multiple input modalities, the study of affective processes beyond the dichotomous valence dimension, and the use of multiple tasks to isolate distinct cognitive processes.

REFERENCES

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., & Damasio, A. R. (2000). A role for romatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *The Journal of Neuroscience*, *20*(7), 2683–2690. https://doi.org/10.1523/jneurosci.20-07-02683.2000

Ally, B. A., & Budson, A. E. (2007). The worth of pictures: Using high density event-related potentials to understand the memorial power of pictures and the dynamics of recognition memory. *NeuroImage*, *35*(1), 378–395. https://doi.org/10.1016/j.neuroimage.2006.11.023

Amedi, A., Jacobson, G., Hendler, T., Malach, R., & Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral Cortex (New York, N.Y. : 1991)*.

Amedi, A., Raz, N., Azulay, H., Malach, R., & Zohary, E. (2010). Cortical activity during tactile exploration of objects in blind and sighted humans. *Restorative Neurology and Neuroscience*. https://doi.org/10.3233/RNN-2010-0503

Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn3800

Amodio, D. M., & Ratner, K. G. (2011). A memory systems model of implicit social cognition. *Current Directions in Psychological Science*. https://doi.org/10.1177/0963721411408562

Arcara, G., & Petrova, A. (2014). erpR: ERP analysis, graphics and utility functions. *R Package Version 0.2.0.*

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology*, *62*(6), 893–912.

https://doi.org/10.1037/0022-3514.62.6.893

Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The Automatic Evaluation Effect: Unconditional Automatic Attitude Activation with a Pronunciation Task. *Journal of Experimental Social Psychology*, *32*(1), 104–128. https://doi.org/10.1006/jesp.1996.0005

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2010). Grounded Cognition: Past, Present, and Future. *Topics in Cognitive Science*, *2*(4), 716–724. https://doi.org/10.1111/j.1756-8765.2010.01115.x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Blechert, J., Meule, A., Busch, N. A., & Ohla, K. (2014). Food-pics: An image database for experimental research on eating and appetite. *Frontiers in Psychology*, *5*(JUN). https://doi.org/10.3389/fpsyg.2014.00617

Buxbaum, L. J., & Saffran, E. M. (2002). Knowledge of object manipulation and object function: Dissociations in apraxic and nonapraxic subjects. *Brain and Language*. https://doi.org/10.1016/S0093-934X(02)00014-7

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.76.5.839

Capitani, E., Laiacona, M., Mahon, B., & Caramazza, A. (2003). What are the facts of semantic

category-specific deficits? A critical review of the clinical evidence. *Cognitive Neuropsychology*, *20*(3–6), 213–261. https://doi.org/10.1080/02643290244000266

Caramazza, A., Anzellotti, S., Strnad, L., & Lingnau, A. (2014). Embodied Cognition and Mirror Neurons: A Critical Assessment. *Annual Review of Neuroscience*, *37*(1), 1–15. https://doi.org/10.1146/annurev-neuro-071013-013950

Caramazza, A., & Shelton, J. R. (1998). Domain-Specific Knowledge Systems in the Brain : The Animate-Inanimate Distinction. *Journal of Cognitive Neuroscience*, *10*(1), 1–34. https://doi.org/10.1162/089892998563752

Carnaghi, A., Silveri, M. C., & Rumiati, R. I. (2015). On the relationship between semantic knowledge and prejudice about social groups in patients with dementia. *Cognitive and Behavioral Neurology*, *28*(2), 71–79. https://doi.org/10.1097/WNN.0000000000000059

Carroll, J. M., Yik, M. S. M., Russell, J. A., & Barrett, L. F. (1999). On the Psychometric Principles of Affect. *Review of General Psychology*. https://doi.org/10.1037/1089-2680.3.1.14

Castelli, L., Zogmaister, C., & Tomelleri, S. (2009). The Transmission of Racial Attitudes Within the Family. *Developmental Psychology*. https://doi.org/10.1037/a0014619

Cattaneo, Z., Devlin, J. T., Salvini, F., Vecchi, T., & Silvanto, J. (2010). The causal role of category-specific neuronal representations in the left ventral premotor cortex (PMv) in semantic processing. *NeuroImage*, *49*(3), 2728–2734. https://doi.org/10.1016/j.neuroimage.2009.10.048

Cattaneo, Z., Rota, F., Vecchi, T., & Silvanto, J. (2008). Using state-dependency of transcranial magnetic stimulation (TMS) to investigate letter selectivity in the left posterior parietal cortex: A comparison of TMS-priming and TMS-adaptation paradigms. *European Journal of Neuroscience*, *28*(9), 1924–1929. https://doi.org/10.1111/j.1460-9568.2008.06466.x

Chedid, G., Wilson, M. A., Provost, J. S., Joubert, S., Rouleau, I., & Brambati, S. M. (2016).

Differential involvement of the anterior temporal lobes in famous people semantics. *Frontiers in Psychology*, *7*, 1333. https://doi.org/10.3389/fpsyg.2016.01333

Chen, R., Wassermann, E. M., Canos, M., & Hallett, M. (1997). Impaired inhibition in writer's cramp during voluntary muscle activation. *Neurology*, *49*(4), 1054–1059. https://doi.org/10.1212/WNL.49.4.1054

Citron, F. M. M. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language*. https://doi.org/10.1016/j.bandl.2011.12.007

Cloutier, J., Mason, M. F., & Macrae, C. N. (2005). The perceptual determinants of person construal: Reopening the social-cognitive toolbox. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.88.6.885

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*. https://doi.org/10.1037/0033-295X.82.6.407

Cunningham, W. A., Espinet, S. D., Deyoung, C. G., & Zelazo, P. D. (2005). Attitudes to the right- and left: Frontal ERP asymmetries associated with stimulus valence and processing goals. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2005.04.044

Cunningham, W. A., Johnson, M. K., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2003). Neural components of social evaluation. *Journal of Personality and Social Psychology*, *85*(4), 639–649. https://doi.org/10.1037/0022-3514.85.4.639

Cunningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and Explicit Evaluation: fMRI Correlates of Valence, Emotional Intensity, and Control in the Processing of Attitudes. *Journal of Cognitive Neuroscience*, *16*(10), 1717–1729. https://doi.org/10.1162/0898929042947919

Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: a social cognitive

neuroscience perspective. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2006.12.005

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*. https://doi.org/10.1016/0010-0277(89)90005-X

Damian, M. F. (2001). Congruity effects evoked by subliminally presented primes: Automaticity rather than semantic processing. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 154–165. https://doi.org/10.1037/0096-1523.27.1.154

Davidson, R. J. (1992). Emotion and Affective Style: Hemispheric Substrates. *Psychological Science*, *3*(1), 39–43. https://doi.org/10.1111/j.1467-9280.1992.tb00254.x

Davidson, R. J. (1995). Cerebral asymmetry, emotion, and affective style. In R. J. Davidson & K. Hugdahl (Eds.), *Brain asymmetry* (pp. 361–387). MIT Press, Cambridge.

Davidson, Richard J., & Irwin, W. (1999). The functional neuroanatomy of emotion and affective style. *Trends in Cognitive Sciences*, *3*(1), 11–21. https://doi.org/10.1016/S1364-6613(98)01265-0

De Houwer, J., Hermans, D., Rothermund, K., & Wentura, D. (2002). Affective priming of semantic categorisation responses. *Cognition and Emotion*. https://doi.org/10.1080/02699930143000419

De Houwer, J., & Randell, T. (2002). Attention to Primes Modulates Affective Priming of Pronunciation Responses. *Experimental Psychology*, *49*(3), 163–170. https://doi.org/10.1026/1618-3169.49.3.163

Decety, J., & Chaminade, T. (2003). Neural correlates of feeling sympathy. *Neuropsychologia*, *41*(2), 127–138. https://doi.org/10.1016/S0028-3932(02)00143-4

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious,

preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences*, *10*(5), 204–211. https://doi.org/10.1016/j.tics.2006.03.007

Fairhall, S. L., & Caramazza, A. (2013). Brain Regions That Represent Amodal Conceptual Knowledge. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.0051-13.2013

Farah, M. J., & McClelland, J. L. (1991). A Computational Model of Semantic Memory Impairment: Modality Specificity and Emergent Category Specificity. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/0096-3445.120.4.339

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion*, *15*(2), 115–141. https://doi.org/10.1080/02699930125908

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in Automatic Activation as an Unobtrusive Measure of Racial Attitudes: A Bona Fide Pipeline? *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.69.6.1013

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*(2), 229–238. https://doi.org/10.1037/0022-3514.50.2.229

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. https://doi.org/10.1093/cercor/bhu186

Fillmore, C. J. (1975). An Alternative to Checklist Theories. *The Annual Proceedings of the First Annual Meeting of the Berkley Linguistics Society*.

Fitzgerald, P., Fountain, S., & Daskalakis, Z. (2006). A comprehensive review of the effects of rTMS on motor cortical excitability and inhibition. *Clinical Neurophysiology*, *117*(12), 2584–2596.

https://doi.org/10.1016/j.clinph.2006.06.712

Fuggetta, G., Rizzo, S., Pobric, G., Lavidor, M., & Walsh, V. (2009). Functional representation of living and nonliving domains across the cerebral hemispheres: a combined event-related potential/transcranial magnetic stimulation study. *Journal of Cognitive Neuroscience*, *21*(2), 403–414. https://doi.org/10.1162/jocn.2008.21030

García-Pacios, J., Garcés, P., del Río, D., & Maestú, F. (2017). Tracking the effect of emotional distraction in working memory brain networks: Evidence from an MEG study. *Psychophysiology*, *54*(11), 1726–1740. https://doi.org/10.1111/psyp.12912

Ghashghaei, H. T., & Barbas, H. (2002). Pathways for emotion: Interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience*. https://doi.org/10.1016/S0306-4522(02)00446-3

Gilbert, S. J., Swencionis, J. K., & Amodio, D. M. (2012). Evaluative vs. trait representation in intergroup social judgments: Distinct roles of anterior temporal lobe and prefrontal cortex. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2012.09.002

Greenwald, A. G., & Banaji, M. R. (1995). Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes. *Psychological Review*. https://doi.org/10.1037/0033-295X.102.1.4

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn_a_01068

Harris, L. T., McClure, S. M., van den Bos, W., Cohen, J. D., & Fiske, S. T. (2007). Regions of the MPFC differentially tuned to social and nonsocial affective evaluation. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(4), 309–316. https://doi.org/10.3758/CABN.7.4.309

Haslam, C., & Sabah, M. (2013). Preservation of person-specific knowledge in semantic memory

disorder: A longitudinal investigation in two cases of dementia. *Journal of Neuropsychology*, *7*(1), 132–138. https://doi.org/10.1111/j.1748-6653.2012.02030.x

Herring, D. R., Taylor, J. H., White, K. R., & Crites, S. L. (2011). Electrophysiological responses to evaluative priming: The LPP is sensitive to incongruity. *Emotion*, *11*(4), 794–806. https://doi.org/10.1037/a0022804

Hilton, J. L., & von Hippel, W. (1996). STEREOTYPES. *Annual Review of Psychology*, *47*(1), 237–271. https://doi.org/10.1146/annurev.psych.47.1.237

Hinojosa, J. A., Carretié, L., Méndez-Bértolo, C., Míguez, A., & Pozo, M. A. (2009). Arousal contributions to affective priming: Electrophysiological correlates. *Emotion*, *9*(2), 164–171. https://doi.org/10.1037/a0014680

Hinojosa, J. A., Carretié, L., Valcárcel, M. A., Méndez-Bértolo, C., & Pozo, M. A. (2009). Electrophysiological differences in the processing of affective information in words and pictures. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(2), 173–189. https://doi.org/10.3758/CABN.9.2.173

Hughes, B. L., Zaki, J., & Ambady, N. (2017). Motivation alters impression formation and related neural systems. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsw147

Itkes, O., Kimchi, R., Haj-Ali, H., Shapiro, A., & Kron, A. (2017). Dissociating affective and semantic valence. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000291

Ito, T. A., & Urland, G. R. (2003). Race and Gender on the Brain: Electrocortical Measures of Attention to the Race and Gender of Multiply Categorizable Individuals. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.85.4.616

Jefferies, E., Patterson, K., Jones, R. W., & Lambon Ralph, M. A. (2009). Comprehension of Concrete and Abstract Words in Semantic Dementia. *Neuropsychology*. https://doi.org/10.1037/a0015452

Kay, J., & Hanley, J. R. (2002). Preservation of memory for people in semantic memory disorder: Further category-specific semantic dissociation. *Cognitive Neuropsychology*, *19*(2), 113–133. https://doi.org/10.1080/02643290143000114

Kensinger, E. A., & Schacter, D. L. (2006). Processing emotional pictures and words: Effects of valence and arousal. *Cognitive, Affective and Behavioral Neuroscience*, *6*(2), 110–126. https://doi.org/10.3758/CABN.6.2.110

Kimball, S., Mattis, P., Natterer, M., & Neumann, S. (2013). Gimp: Gnu image manipulation program. Retrieved from www.gimp.org

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2014.01.002

Kissler, J., Assadollahi, R., & Herbert, C. (2006). Emotional and semantic networks in visual word processing: insights from ERP studies. *Progress in Brain Research*, *156*, 147–183. https://doi.org/10.1016/S0079-6123(06)56008-X

Klauer, K. C., & Musch, J. (2003). Affective priming: findings and theories. In *The psychology of evaluation*.

Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, *52*(3), 3–6.

Klinger, M. R., Burton, P. C., & Pitts, G. S. (2000). Mechanisms of unconscious priming: I. Response competition, not spreading activation. *Journal of Experimental Psychology: Learning, Memory,*

and Cognition, *26*(2), 441–455. https://doi.org/10.1037/0278-7393.26.2.441

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The Representation of Abstract Words: Why Emotion Matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34. https://doi.org/10.1037/a0021446

Kuhlmann, M., Hofmann, M. J., & Jacobs, A. M. (2017). If you don't have valence, ask your neighbor: Evaluation of neutral words as a function of affective semantic associates. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2017.00343

Labeye, E., Oker, A., Badard, G., & Versace, R. (2008). Activation and integration of motor components in a short-term priming paradigm. *Acta Psychologica*, *129*(1), 108–111. https://doi.org/10.1016/j.actpsy.2008.04.010

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lakens, D., & Etz, A. J. (2017). Too True to be Bad. *Social Psychological and Personality Science*, *8*(8), 875–881. https://doi.org/10.1177/1948550617693058

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2016). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Lebrecht, S., Bar, M., Barrett, L. F., & Tarr, M. J. (2012). Micro-valences: Perceiving affective valence in everyday objects. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2012.00107

Leclerc, C. M., & Kensinger, E. A. (2011). Neural processing of emotional pictures and words: A comparison of young and older adults. *Developmental Neuropsychology*, *36*(4), 519–538. https://doi.org/10.1080/87565641.2010.549864

Leonardelli, E., Fait, E., & Fairhall, S. L. (2019). Temporal dynamics of access to amodal representations of category-level conceptual information. *Scientific Reports*, *9*(1), 239. https://doi.org/10.1038/s41598-018-37429-2

Leshinskaya, A., & Caramazza, A. (2016). For a cognitive neuroscience of concepts: Moving beyond the grounding issue. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-015-0870-z

Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2016). The Brain Basis of Positive and Negative Affect: Evidence from a Meta-Analysis of the Human Neuroimaging Literature. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhv001

Liu, H., Hu, Z., Peng, D., Yang, Y., & Li, K. (2010). Common and segregated neural substrates for automatic conceptual and affective priming as revealed by event-related functional magnetic resonance imaging. *Brain and Language*. https://doi.org/10.1016/j.bandl.2009.11.001

Liu, Y., Huang, H., McGinnis-Deweese, M., Keil, A., & Ding, M. (2012). Neural Substrate of the Late Positive Potential in Emotional Processing. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.3109-12.2012

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157. https://doi.org/10.1111/psyp.12639

Luo, Q., Peng, D., Jin, Z., Xu, D., Xiao, L., & Ding, G. (2004). Emotional valence of words modulates the subliminal repetition priming effect in the left fusiform gyrus: An event-related fMRI study. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2003.09.048

Lyons, F., Kay, J., Hanley, J. R., & Haslam, C. (2006). Selective preservation of memory for people in the context of semantic memory disorder: Patterns of association and dissociation. *Neuropsychologia*, *44*(14), 2887–2898. https://doi.org/10.1016/j.neuropsychologia.2006.06.005

Mahon, B. Z., & Caramazza, A. (2009). Concepts and Categories: A Cognitive Neuropsychological Perspective. *Annual Review of Psychology*. https://doi.org/10.1146/annurev.psych.60.110707.163532

Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain? *Trends in Cognitive Sciences*, *15*(3), 97–103. https://doi.org/10.1016/j.tics.2011.01.004

Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. J. (1996). Independent Component Analysis of Electroencephalographic Data. *Advances in Neural Information Processing Systems*, *8*, 145–151. https://doi.org/10.1109/ICOSP.2002.1180091

Man, V., Nohlen, H. U., Melo, H., & Cunningham, W. A. (2017). Hierarchical Brain Systems Support Multiple Representations of Valence and Mixed Affect. *Emotion Review*. https://doi.org/10.1177/1754073916667237

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://doi.org/10.1016/j.jml.2016.04.001

Martin, A. (2016). GRAPES—Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. *Psychonomic Bulletin and Review*, *23*(4), 979–990. https://doi.org/10.3758/s13423-015-0842-3

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*. https://doi.org/10.1126/science.270.5233.102

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*. https://doi.org/10.1038/379649a0

Mason, M. F., & Macrae, C. N. (2004). Categorizing and individuating others: The neural substrates of person perception. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/0898929042947801

Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., & Tononi, G. (2005). Breakdown of Cortical Effective Connectivity During Sleep. *Science*, *309*(5744), 2228–2232. https://doi.org/10.1126/science.1117256

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Meteyard, L., & Vigliocco, G. (2008). The Role of Sensory and Motor Information in Semantic Representation: A Review. In *Handbook of Cognitive Science*. https://doi.org/10.1016/B978-0-08-046616-3.00015-3

Miceli, G., Capasso, R., Daniele, a, Esposito, T., Magarelli, M., & Tomaiuolo, F. (2000). Selective deficit for people's names following left temporal damage: an impairment of domain-specific conceptual knowledge. *Cognitive Neuropsychology*, *17*(6), 489–516. https://doi.org/10.1080/02643290050110629

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). The adaptation of the affective norms for english words (ANEW) for Italian. *Behavior Research Methods*, *46*(3).

Moss, H. E., Abdallah, S., Fletcher, P., Bright, P., Pilgrim, L., Acres, K., & Tyler, L. K. (2005). Selecting among competing alternatives: Selection and retrieval in the left inferior frontal gyrus. *Cerebral Cortex*, *15*(11), 1723–1735. https://doi.org/10.1093/cercor/bhi049

Myung, J., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: manipulation knowledge of objects. *Cognition*, *98*(3), 223–243. https://doi.org/10.1016/j.cognition.2004.11.010

Neely, J. H. (2012). Semantic Priming Effects in Visual Word Recognition; A Selective Review of Current Findings and Theories. In *Basic processes in reading*.

Newcombe, P. I., Campbell, C., Siakaluk, P. D., & Pexman, P. M. (2012). Effects of Emotional and Sensorimotor Knowledge in Semantic Processing of Concrete and Abstract Nouns. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00275

Noonan, K. A., Jefferies, E., Visser, M., & Lambon Ralph, M. A. (2013). Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *Journal of Cognitive Neuroscience*, *25*(11), 1824–1850. https://doi.org/10.1162/jocn_a_00442

Norris, C. J., Chen, E. E., Zhu, D. C., Small, S. L., & Cacioppo, J. T. (2004). The Interaction of Social and Emotional Processes in the Brain. *Journal of Cognitive Neuroscience*, *16*(10), 1818–1829. https://doi.org/10.1162/0898929042947847

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/0096-3445.134.4.565

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*. https://doi.org/10.3389/fninf.2016.00027

Opie, G. M., Ridding, M. C., & Semmler, J. G. (2015). Task-related changes in intracortical inhibition assessed with paired- and triple-pulse transcranial magnetic stimulation. *Journal of Neurophysiology*, *113*(5), 1470–1479. https://doi.org/10.1152/jn.00651.2014

Oshio, R., Tanaka, S., Sadato, N., Sokabe, M., Hanakawa, T., & Honda, M. (2010). Differential effect of double-pulse TMS applied to dorsal premotor cortex and precuneus during internal operation of visuospatial information. *NeuroImage*, *49*(1), 1108–1115. https://doi.org/10.1016/j.neuroimage.2009.07.034

Panksepp, J. (1998). *Affective neuroscience: the foundations of human and animal*. *New York: Oxford University Press*.

Panksepp, J. (2012). The Philosophical Implications of Affective Neuroscience. *Journal of Consciousness Studies*, *19*(3–4), 6–48.

Papeo, L., Wurm, M. F., Oosterhof, N. N., & Caramazza, A. (2017). The neural representation of human versus nonhuman bipeds and quadrupeds. *Scientific Reports*. https://doi.org/10.1038/s41598-017-14424-7

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987. https://doi.org/10.1038/nrn2277

Patterson, K., Ralph, M. A. L., Jefferies, E., Woollams, A., Jones, R., Hodges, J. R., & Rogers, T. T. (2006). "Presemantic" cognition in semantic dementia: Six deficits in search of an explanation. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn.2006.18.2.169

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, *9*(2), 148–158. https://doi.org/10.1038/nrn2317

Piretti, L. (2017). The semantic representation of social groups and its neural substrate (Doctoral dissertation).

Piretti, L., Carnaghi, A., Campanella, F., Ambron, E., Skrap, M., & Rumiati, R. I. (2015). The neural network associated with lexical-semantic knowledge about social groups. *Cortex*, *70*, 155–168. https://doi.org/10.1016/j.cortex.2015.06.024

Pridmore, S., Fernandes Filho, J. A., Nahas, Z., Liberatos, C., & George, M. S. (1998). Motor

Threshold in Transcranial Magnetic Stimulation. *The Journal of ECT*, *14*(1), 25–27. https://doi.org/10.1097/00124509-199803000-00004

Priester, J. R., & Petty, R. E. (1996). The Gradual Threshold Model of Ambivalence: Relating the Positive and Negative Bases of Attitudes to Subjective Ambivalence. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.71.3.431

R Core Team. (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and Deterioration of Semantic Memory: A Neuropsychological and Computational Investigation. *Psychological Review*, *111*(1), 205–235. https://doi.org/10.1037/0033-295X.111.1.205

Rogers, T. T., Patterson, K., Jefferies, E., & Lambon Ralph, M. A. (2015). Disorders of representation and control in semantic cognition: Effects of familiarity, typicality, and specificity. *Neuropsychologia*, *76*, 220–239. https://doi.org/10.1016/j.neuropsychologia.2015.04.015

Rosenthal, R. (1986). Meta-Analytic Procedures for Social Science Research. *Educational Researcher*, *15*(8), 18. https://doi.org/10.2307/1175262

Rossi, S., Hallett, M., Rossini, P. M., & Pascual-Leone, A. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology*, *120*(12), 2008–2039. https://doi.org/10.1016/j.clinph.2009.08.016

Rossi, S., Hallett, M., Rossini, P. M., & Pascual-Leone, A. (2011). Screening questionnaire before TMS: An update. *Clinical Neurophysiology*, *122*(8), 1686. https://doi.org/10.1016/j.clinph.2010.12.037

Rumiati, R. I., Carnaghi, A., Improta, E., Diez, A. L., & Silveri, M. C. (2014). Social groups have a representation of their own: Clues from neuropsychology. *Cognitive Neuroscience*, *5*(2), 85–96. https://doi.org/10.1080/17588928.2013.876981

Russell, J. A., & Carroll, J. M. (1999). On the Bipolarity of Positive and Negative Affect. *Psychological Bulletin*. https://doi.org/10.1037/0033-2909.125.1.3

Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Tiffany, I., & Lang, P. J. (2000). Affective picture processing: The late positive potential is modulated by motivational relevance. *Psychophysiology*. https://doi.org/10.1017/S0048577200001530

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, *132*(3), 617–627. https://doi.org/10.1093/brain/awn279

Shirota, Y., Sommer, M., & Paulus, W. (2016). Strength-Duration Relationship in Paired-pulse Transcranial Magnetic Stimulation (TMS) and Its Implications for Repetitive TMS. *Brain Stimulation*, *9*(5), 755–761. https://doi.org/10.1016/j.brs.2016.04.019

Silvanto, J., Muggleton, N., & Walsh, V. (2008). State-dependency in brain stimulation studies of perception and cognition. *Trends in Cognitive Sciences*, *12*(12), 447–454. https://doi.org/10.1016/j.tics.2008.09.004

Simmons, W. K., & Barsalou, L. W. (2003). The similarity-in-topography principle: reconciling theories of conceptual deficits. *Cognitive Neuropsychology*, *20*(3–6), 451–486. https://doi.org/10.1080/02643290342000032

Simmons, W. K., Hamann, S. B., Harenski, C. L., Hu, X. P., & Barsalou, L. W. (2008). fMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology Paris*. https://doi.org/10.1016/j.jphysparis.2008.03.014

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2008.03.061

Spruyt, A., Hermans, D., Houwer, J. De, & Eelen, P. (2002). On The Nature of the Affective Priming Effect: Affective Priming of Naming Responses. *Social Cognition*, *20*(3), 227–256. https://doi.org/10.1521/soco.20.3.227.21106

Spruyt, A., Houwer, J. De, & Hermans, D. (2009). Modulation of automatic semantic priming by feature-specific attention allocation. *Journal of Memory and Language*, *61*(1), 37–54. https://doi.org/10.1016/j.jml.2009.03.004

Storbeck, J., & Clore, G. L. (2005). With sadness comes accuracy; with happiness, false memory: Mood and the false memory effect. *Psychological Science*, *16*(10), 785–791. https://doi.org/10.1111/j.1467-9280.2005.01615.x

Storbeck, J., & Robinson, M. D. (2004). Preferences and Inferences in Encoding Visual Objects: A Systematic Comparison of Semantic and Affective Priming. *Personality and Social Psychology Bulletin*. https://doi.org/10.1177/0146167203258855

Suran, T., Arcara, G., Piretti, L., & Rumiati, R. I. (2019). Electrophysiological Correlates of Social Group Representations in Affective Priming. https://doi.org/10.31234/osf.io/kewu5

Suran, T., Rumiati, R. I., & Piretti, L. (2019). The contribution of the left inferior frontal gyrus in affective processing of social groups. *Cognitive Neuroscience*, 1–10. https://doi.org/10.1080/17588928.2019.1593127

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Computational Intelligence and Neuroscience*, *2011*, 1–13. https://doi.org/10.1155/2011/879716

Tempel, K., Kuchinke, L., Urton, K., Schlochtermeier, L. H., Kappelhoff, H., & Jacobs, A. M. (2013). Effects of positive pictograms and words: Anemotional word superiority effect? *Journal of Neurolinguistics*. https://doi.org/10.1016/j.jneuroling.2013.05.002

Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: Inferring "how" from "where." *Neuropsychologia*, *41*(3), 280–292. https://doi.org/10.1016/S0028-3932(02)00161-6

Thompson, S. A., Graham, K. S., Williams, G., Patterson, K., Kapur, N., & Hodges, J. R. (2004). Dissociating person-specific from general semantic knowledge: Roles of the left and right temporal lobes. *Neuropsychologia*, *42*(3), 359–370. https://doi.org/10.1016/j.neuropsychologia.2003.08.004

Tukachinsky, R., Mastro, D., & Yarchi, M. (2015). Documenting Portrayals of Race/Ethnicity on Primetime Television over a 20-Year Span and Their Association with National-Level Racial/Ethnic Attitudes. *Journal of Social Issues*, *71*(1), 17–38. https://doi.org/10.1111/josi.12094

Van der Cruyssen, L., Heleven, E., Ma, N., Vandekerckhove, M., & Van Overwalle, F. (2014). Distinct neural correlates of social categories and personality traits. *NeuroImage*, *104*, 336–346. https://doi.org/10.1016/j.neuroimage.2014.09.022

Van Strien, J. W., Langeslag, S. J. E., Strekalova, N. J., Gootjes, L., & Franken, I. H. A. (2009). Valence interacts with the early ERP old/new effect and arousal with the sustained ERP old/new effect for affective pictures. *Brain Research*, *1251*, 223–235. https://doi.org/10.1016/j.brainres.2008.11.027

Vigliocco, G., Kousta, S. T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: The role of emotion. *Cerebral Cortex*, *24*(7), 1767–1777. https://doi.org/10.1093/cercor/bht025

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic

representation. *Language and Cognition*, *1*(02), 219–247. https://doi.org/10.1515/LANGCOG.2009.011

Voss, J. L., Lucas, H. D., & Paller, K. A. (2010). Conceptual priming and familiarity: Different expressions of memory during recognition testing with distinct neurophysiological correlates. *Journal of Cognitive Neuroscience*, *22*(11), 2638–2651. https://doi.org/10.1162/jocn.2009.21341

Warrington, E. K. (1975). The selective impairment of semantic memory. *The Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/14640747508400525

Warrington, E. K., & Mccarthy, R. (1983). Category specific access dysphasia. *Brain*. https://doi.org/10.1093/brain/106.4.859

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, *107*(3), 829–853. https://doi.org/10.1093/brain/107.3.829

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03196322

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*(1), 101–126. https://doi.org/10.1037/0033-295X.107.1.101

Wood, J. N., Romero, S. G., Makale, M., & Grafman, J. (2003). Category-Specific Representations of Social and Nonsocial Knowledge in the Human Prefrontal Cortex. *Journal of Cognitive Neuroscience*, *15*(2), 236–248. https://doi.org/10.1162/089892903321208178

Yao, Z., & Wang, Z. (2013). The effects of the concreteness of differently valenced words on affective priming. *Acta Psychologica*, *143*(3), 269–276. https://doi.org/10.1016/j.actpsy.2013.04.008

Yao, Z., & Wang, Z. (2014). Concreteness of positive word contributions to affective priming: An ERP study. *International Journal of Psychophysiology*, *93*(3), 275–282.

https://doi.org/10.1016/j.ijpsycho.2014.06.005

Yao, Z., Yu, D., Wang, L., Zhu, X., Guo, J., & Wang, Z. (2016). Effects of valence and arousal on emotional word processing are modulated by concreteness : Behavioral and ERP evidence from a lexical decision task. *International Journal of Psychophysiology*, *110*, 231–242. https://doi.org/10.1016/j.ijpsycho.2016.07.499

Zhang, D., He, W., Wang, T., Luo, W., Zhu, X., Gu, R., … Luo, Y. (2014). Three stages of emotional word processing: an ERP study with rapid serial visual presentation. *Social Cognitive and Affective Neuroscience*, *9*(12), 1897–1903. https://doi.org/10.1093/scan/nst188

Zhang, Q., Lawson, A., Guo, C., & Jiang, Y. (2006). Electrophysiological correlates of visual affective priming. *Brain Research Bulletin*, *71*(1–3), 316–323. https://doi.org/10.1016/j.brainresbull.2006.09.023

Zhang, Q., Li, X., Gold, B. T., & Jiang, Y. (2010). Neural correlates of cross-domain affective priming. *Brain Research*, *1329*, 142–151. https://doi.org/10.1016/j.brainres.2010.03.021

Zhu, C., He, W., Qi, Z., Wang, L., Song, D., Zhan, L., … Luo, W. (2015). The time course of emotional picture processing: an event-related potential study using a rapid serial visual presentation paradigm. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2015.00954

APPENDIX 1

Target stimuli from Study 1 [English translation]

Positive		Negative	
Social categories	**Objects**	**Social categories**	**Objects**
agricoltori [farmers]	anelli [rings]	alcolisti [alcoholics]	aghi [needles]
ballerini [dancers]	aquiloni [kites]	anziani [elderly]	bare [coffins]
dottori [doctors]	bambole [dolls]	arabi [Arabs]	barili [barrels]
ebrei [Jews]	barche [boats]	barboni [homeless]	bisturi [scalpels]
infermiere [nurses]	carrozze [coaches]	dentisti [dentists]	coltelli [knives]
medici [doctors]	chiavi [keys]	immigrati [immigrants]	fucili [rifles]
modelle [female models]	corone [crowns]	macellai [butchers]	lamette [blades]
musicisti [musicians]	elicotteri [helicopters]	militari [military]	martelli [hammers]
operai [workers]	gioielli [jewels]	negri [Blacks, der.]	pistole [guns]
pescatori [fishermen]	palloni [balloons]	obesi [obese]	posacenere [ashtrays]
pittori [painters]	porte [doors]	poliziotti [policemen]	pugnali [daggers]
professori [professors]	rossetti [lipsticks]	poveri [poor]	rasoi [razors]
scrittori [writers]	scooter [scooters]	preti [priests]	sigarette [cigarettes]
studenti [students]	televisori [televisions]	prostitute [prostitutes]	siringhe [syringes]
vigili [traffic policemen]	trofei [trophies]	suore [nuns]	stampelle [crutches]

APPENDIX 2

Target stimuli from Study 2 [English translation]

Social categories		Nonsocial categories	
Negative	**Positive**	**Negative**	**Positive**
aguzzini [tormentors]	anziani [elderly]	acari [mites]	anelli [rings]
alcolisti [alcoholics]	artisti [artists]	acidi [acids]	aquiloni [kites]
barboni [homeless]	attori [actors]	bare [coffins]	barche [boats]
becchini [gravediggers]	bagnini [lifeguards]	batteri [bacteria]	biciclette [bicycles]
bulimiche [bulimics]	ballerini [dancers]	bisturi [scalpels]	birre [beers]
bulli [bullies]	camerieri [waiters]	bombe [bombs]	borse [purses]
detenuti [convicts]	capitani [captains]	cannoni [cannons]	caramelle [candies]
dittatori [dictators]	donne [women]	cateteri [catheters]	chitarre [guitars]
drogati [drug addicts]	giovani [youth]	coltelli [knives]	cuscini [pillows]
estremisti [extremists]	insegnanti [teachers]	droghe [drugs]	diamanti [diamonds]
fascisti [fascists]	maestre [teachers, fem.]	fucili [rifles]	divani [couches]
fumatori [smokers]	maestri [masters]	granate [grenades]	finestre [windows]
grassi [fat people]	marinai [sailors]	lapidi [tombstones]	foglie [leaves]
infermi [sick]	medici [medics]	pallottole [bullets]	forchette [forks]
ladri [thieves]	modelle [female models]	pasticche [tablets]	giocattoli [toys]
militari [military]	mogli [wives]	pistole [guns]	gioielli [jewels]
negri [Blacks, derog.]	operai [workers]	proiettili [bullets]	libri [books]
obesi [obese]	pittori [painters]	protesi [protheses]	maglie [sweaters]
orfani [orphans]	poeti [poets]	pugnali [daggers]	monete [coins]
perdenti [losers]	pompieri [firefighters]	sigarette [cigarettes]	palloni [balloons]
poveri [poor]	ragazze [girls]	siringhe [syringes]	perle [pearls]
psicopatici [psychopaths]	ragazzi [boys]	tombe [tombs]	regali [gifts]
sadici [sadists]	scolari [pupils]	tossine [toxins]	slitte [sleighs]
terroni [southerners, derog.]	scrittori [writers]	trapani [drills]	soldi [money]
tossici [toxics]	scultori [sculptors]	veleni [poisons]	tesori [treasures]
vandali [vandals]	sportive [sportswomen]	verruche [warts]	trofei [trophies]
zingari [gypsies]	studenti [students]	vesciche [blisters]	violini [violins]

Prime stimuli from Study 2 [English translation]

Negative	Positive
PAURA [FEAR]	AMORE [LOVE]
LUTTO [MOURNING]	GIOIA [JOY]
MORTE [DEATH]	FELICE [HAPPY]
DOLORE [PAIN]	VIVACE [LIVELY]
FERITE [WOUNDS]	SORRISO [SMILE]
INCUBO [NIGHTMARE]	TRIONFO [TRIUMPH]
TRADIRE [BETRAY]	VINCERE [TO WIN]
ANSIOSO [ANXIOUS]	SUCCESSO [SUCCESS]
PERICOLO [DANGER]	PASSIONE [PASSION]

Valence polarity	Valence	Arousal	Concreteness	Familiarity	Length
Positive	6.33 (.35)	5.47 (.31)	4.12 (.82)	5.28 (.71)	6.56 (1.13)
Negative	1.78 (.22)	5.50 (.19)	4.21 (.70)	5.30 (.44)	6.11 (1.05)
Neutral/#	-	-	-	-	6.33 (.94)

Mean valence, arousal, concreteness and familiarity ratings (0 = extremely low; 7 = extremely high) and average number of characters for each prime valence category. Standard deviations reported within parentheses.

APPENDIX 3

Target word stimuli from Study 3 [English translation]

Social groups		Objects	
Negative	**Positive**	**Negative**	**Positive**
alcolisti [alcoholics]	bambini [children]	bare [coffins]	anelli [rings]
barboni [homeless]	anziani [elderly]	coltelli [knives]	biciclette [bicycles]
detenuti [convicts]	suore [nuns]	droghe [drugs]	chitarre [guitars]
drogati [junkies]	camerieri [waiters]	fucili [rifles]	cuscini [pillows]
migrant [migrants]	marina [sailors]	granate [grenades]	libri [books]
militari [military]	ballerina [dancers]	manette [handcuffs]	matite [pencils]
obesi [obese]	bagnini [lifeguards]	pistole [guns]	orology [watches]
prostitute [prostitutes]	pompieri [firefighters]	sigarette [cigarettes]	palloni [balloons]
terroristi [terrorists]	scultori [sculptors]	siringhe [syringes]	perle [pearls]
zingari [gypsies]	pittori [painters]	stampelle [crutches]	violini [violins]