Dynamic Interactive theory as a domain-general account of social perception

Jonathan B. Freeman

New York University

Ryan M. Stolier

Columbia University

Jeffrey A. Brooks

New York University

Corresponding author:

Jonathan B. Freeman, Ph.D.
Department of Psychology
New York University
6 Washington Place
New York, NY 10003
Email: jon.freeman@nyu.edu

Abstract

The perception of social categories, emotions, and personality traits from others′ faces each have been studied extensively but in relative isolation. We synthesize emerging findings suggesting that, in each of these domains of social perception, both a variety of bottom-up facial features and top-down social cognitive processes play a part in driving initial perceptions. Among such top-down processes, social-conceptual knowledge in particular can have a fundamental structuring role in how we perceive others′ faces. Extending the Dynamic Interactive framework (Freeman & Ambady, 2011), we outline a perspective whereby the perception of social categories, emotions, and traits from faces can all be conceived as emerging from an integrated system relying on domain-general cognitive properties. Such an account of social perception would envision perceptions to be a rapid, but gradual, process of negotiation between the variety of visual cues inherent to a person and the social cognitive knowledge an individual perceiver brings to the perceptual process. We describe growing evidence in support of this perspective as well as its theoretical implications for social psychology.

Dynamic Interactive theory as a domain-general account of social perception

Although often warned not to judge a book by its cover, we cannot help but render any number of judgments on encountering the people around us. From facial features alone, seemingly immediately we perceive the social categories to which they belong (e.g., gender, race), their current emotional state (e.g., sad), and the personality characteristics they likely possess (e.g., trustworthy, intelligent). The field of social psychology has taken great interest in these judgments, as the outcomes of each type of judgment have wide-ranging consequences for social interaction and society at large. Social category judgments tend to spontaneously activate related stereotypes, attitudes, and goals and can bear a number of cognitive, affective, and behavioral consequences, such as providing a basis for subsequent prejudice and discrimination ADDIN EN.CITE (Brewer, 1988; Fiske & Neuberg, 1990; Macrae & Bodenhausen, 2000). Emotion judgments, of course, have long been noted to drive nonverbal communication and provide critical signals for upcoming behaviors ADDIN EN.CITE (Darwin, 1872; Ekman & Friesen, 1971; Ekman, Sorensen, & Friesen, 1969). Finally, trait judgments from facial features alone occur spontaneously and outside awareness, and they can impact a range of evaluations, behavior, and real-world outcomes including political elections, financial success, and criminal-sentencing decisions (for review, Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015).

Given their implications, early work in the field of social psychology focused on the products of these judgments and the variety of downstream effects that ensue. At the same time, research in the cognitive, neural, and vision sciences aimed to characterize the underlying visual cues and basic mechanisms driving face perception. Recently, a unified 'social vision' approach has formed (Adams, Ambady, Nakayama, & Shimojo, 2011; Balcetis & Lassiter, 2010; Freeman & Ambady, 2011), in which the process of social perception is integrated with the products that

follow. This approach stands in contrast to a more traditional divide wherein these levels of analysis are studied by fairly separate disciplines.

Although social categorization, emotion recognition, and trait attribution have each been studied in relatively isolated literatures, traditionally these literatures have shared what could be called a feed-forward emphasis (although with exceptions). In a feed-forward approach, perceptual cues activate an internal representation (e.g., social category, emotion, trait), which in turn drives subsequent cognitive, affective, motivational, and behavioral processes. For instance, classic and influential models of social categorization treated a fully formed categorization (e.g., man, Black person) as the initial starting point ADDIN EN.CITE (Brewer, 1988; Fiske & Neuberg, 1990; Macrae & Bodenhausen, 2000); prominent basic emotion theories treated emotion percepts (e.g., angry) as if directly "read out" from specific combinations of facial action units in a universal, genetically determined fashion (Ekman, 1993); and popular models of face-based trait impressions have tended to focus on specific sets of facial features that produce specific impressions in a bottom-up fashion (Oosterhof & Todorov, 2008; Zebrowitz & Montepare, 2008). In all cases, the face itself directly conveys a social judgment, and little attention was paid to processes harbored within perceivers that might also shape perception.

There has been an increasing recognition of such processes and the important role that top-down social cognitive factors, such as stereotypes, attitudes, and goals, play in "initial" social perceptions ADDIN EN.CITE (Freeman & Johnson, 2016; Hehman, Stolier, Freeman, Flake, & Xie, 2019; Huang & Sherman, 2018; Kawakami, Amodio, & Hugenberg, 2017). In the context of perceiving social categories and its interplay with stereotype processes, the Dynamic Interactive (DI) theory provided a framework and computational model to understand the mutual interplay of bottom-up visual cues and top-down social cognitive factors in driving perceptions

(Freeman & Ambady, 2011). Here we extend the DI theory to provide an understanding of similar mutual interplay in the context of perceiving emotions and personality traits as well. We aim to show that an integrated system relying on domain-general cognitive principles may provide a helpful model of visually-based social perception, broadly construed.

1. Dynamic Interactive (DI) Theory

The DI framework (Freeman & Ambady, 2011) uses domain-general cognitive and computational principles, such as recurrent processing and mutual constraint satisfaction, in order to argue that an initial social perception (e.g., Male, Black, Happy) is a rapid, yet gradual, process of negotiation between the multiple visual features inherent to a person (e.g., facial and bodily cues) and what social cognitive processes a perceiver brings to the perceptual process (e.g., stereotypes, attitudes, goals). Accordingly, initial categorizations are not discrete "read outs" of facial features; they evolve over hundreds of milliseconds – in competition with other partially-active perceptions – and may be dynamically shaped by context and one's stereotypes, attitudes, and goals.

Why might this be the case? At the neural level, the representation of a social category would be reflected by a pattern of activity distributed across a large population of neurons. Thus, activating a social category representation would involve continuous changes in a pattern of neuronal activity (Smith & Ratcliff, 2004; Spivey & Dale, 2006; Usher & McClelland, 2001). Neuronal recordings in nonhuman primates have shown that almost 50% of a face's visual information rapidly accumulates in the brain's perceptual system within 80 ms, while the remaining 50% gradually accumulates over the following hundreds of milliseconds (Rolls & Tovee, 1995). As such, during early moments of perception when only a "gist" is available, the

transient interpretation of a face is partially consistent with multiple interpretations (e.g., both Male or Female). As information accumulates and representations become more fine-grained, the pattern of neuronal activity dynamically sharpens into an increasingly stable and complete representation (e.g., Male), while other, competing representations (e.g., Female) are pushed out ADDIN EN.CITE (Freeman & Ambady, 2011; Freeman, Ambady, Midgley, & Holcomb, 2011; Freeman, Stolier, Brooks, & Stillerman, 2018).

The DI theory emphasizes the importance of dynamic competition inherent to the perceptual process because it allows the perceptual system to take the natural diversity in a face′s visual cues (e.g., masculine features on a woman′s face) and slot it into stable categories. A central premise to the theory is that, during the hundreds of milliseconds it takes for the neuronal activity to achieve a stable pattern (~100% Male or ~100% Female), top-down factors, such as attitudes, goals, and most notably stereotypes or social-conceptual knowledge more generally, can also exert an influence, thereby partly determining the pattern that the system settles into (Freeman & Ambady, 2011). Accordingly, social category perception is rendered a compromise between the perceptual cues "actually" there and the social cognitive factors and pre-existing assumptions perceivers bring to the perceptual process. Just as expectations from prior knowledge are effortlessly and rapidly brought to bear on perceiving an ambiguous letter to align with one′s assumptions (e.g. A or H, see Fig. 1), so too may social expectations and social-conceptual knowledge shape face perception. Indeed, simulations with the computational model derived from the DI theory (described later) have been shown to account well for a wide range of phenomena ADDIN EN.CITE (e.g., Freeman & Ambady, 2011; Freeman, Pauker, & Sanchez, 2016; Freeman, Penner, Saperstein, Scheutz, & Ambady, 2011).

Figure 1. Perceivers readily and involuntarily perceive the words "CAT" and "THE", rather than "CHT" and "TAE", due to top-down knowledge of such words, even though the middle A/H letter is identical. Figure taken from Gaspelin and Luck (2018).

Of all social cognitive processes that may shape initial perceptions, the top-down impact of stereotypes has perhaps the strongest support in terms of the theoretical mechanism at play. It also connects social perception to a wider literature on the interplay of conceptual knowledge and visual perception. Stereotypes, after all, are merely conceptual knowledge related to social categories – semantic associations activated by social category representations. A central argument of the DI theory is that stereotypes are semantic associations that, when activated, can become implicit expectations during perception, and that they thereby take on the ability to influence perception ADDIN EN.CITE (Freeman & Ambady, 2011; Freeman, Penner, et al., 2011; Johnson, Freeman, & Pauker, 2012) (see Fig. 2). However, we believe that the theoretical and computational basis of understanding the interplay of facial features and stereotypes in social category perception (which was the initial focus of the DI theory) sets the stage for understanding the interplay of visual features and conceptual knowledge in driving social perception more broadly (e.g. emotion perceptions, trait impressions).

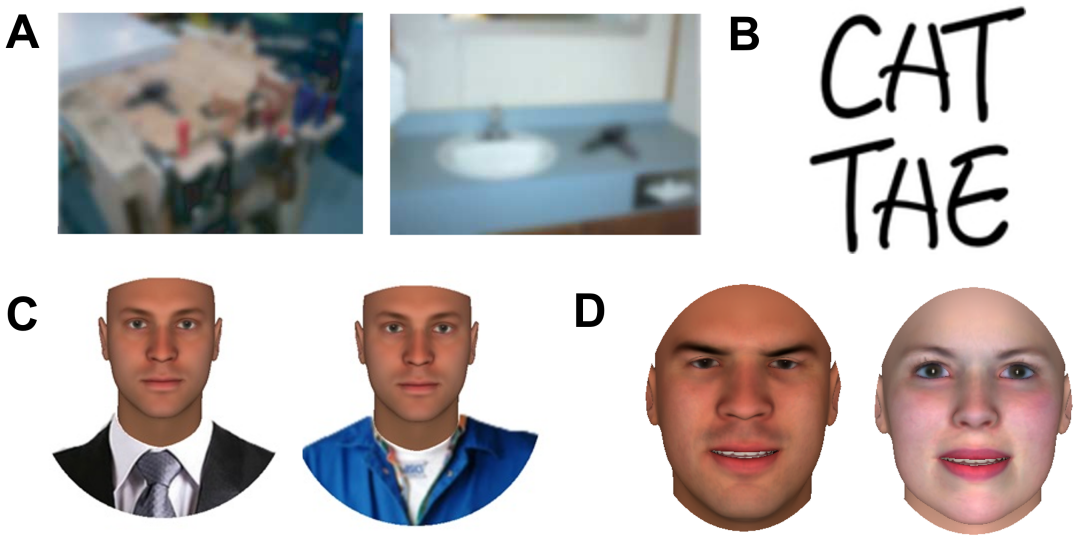

Figure 2. The impact of social-conceptual knowledge on face perception shares a fundamental similarity with more general top-down impacts of conceptual associations in perception. **(A)** Conceptual knowledge about hairdryers and drills and about garages and bathrooms leads an ambiguous object to be readily disambiguated by the context (Bar, 2004). **(B)** The "CAT" and "THE" example from Fig. 1, where stored representations of "CAT" and "THE" lead to opposite interpretations of the same letter. **(C)** Contextual attire cues bias perception of a racially-ambiguous face to be White when surrounded by high-status attire but to be Black when surrounded by low-status attire, due to stereotypic associations between race and social status (Freeman, Penner, et al., 2011). **(D)** An emotionally ambiguous face is perceived to be angry when male but happy when female, due to stereotypic associations linking men to anger and women to joy (Hess, Adams, & Kleck, 2004). Adapted from Freeman and Johnson (2016).

1.1 Conceptual knowledge in visual perception

Intuitively, we might expect that our perception of a visual stimulus such as a face would be immune to conceptual knowledge (e.g., stereotypes) and other top-down factors, instead reflecting a veridical representation of the perceptual information before our eyes (Marr, 1982). This was long argued to be the case (Fodor, 1983; Pylyshyn, 1984) and is still an assumption of many popular feed-forward models of object recognition (Riesenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007). An important exception historically was the 'New Look' perspective

that emerged over a half-century ago, arguing that motives can impact perception (i.e., we see what we want to see) and providing evidence that, for example, poor children overestimate the size of coins (Bruner & Goodman, 1947). However, the perspective soon lost favor. Today, many researchers view perception as an active and constructive process, where context and prior knowledge adaptively constrain perception. As such, few are likely to refute top-down influences on perceptual decision-making generally, but debate continues as to whether these influences would operate at the level of perception itself, or merely on attentional or post-perceptual decision processes (Firestone & Scholl, 2015; Pylyshyn, 1999). In our view, top-down influences are likely to manifest at multiple levels of perceptual processing itself, and continued arguments for the cognitive impenetrability of perception are difficult to reconcile with swaths of empirical findings and a modern understanding of the neuroscience of perception (see Vinson et al., 2016).

Indeed, numerous findings now support the notion that top-down conceptual knowledge plays an important role in visual perception. And while initially the DI theory incorporated such insights to focus on stereotypes′ impact on face perception, we aim to show here that the theory and conceptually situated nature of perception can be extended to understand other domains of social perception more generally. Evidence for the conceptual scaffolding of perception is now quite vast (for review, Collins & Olson, 2014). Large-scale neural oscillations across the brain allow visual perception to arise from both bottom-up feed-forward and top-down feedback influences (Engel, Fries, & Singer, 2001; Gilbert & Sigman, 2007), and even the earliest of responses in primary visual cortex (V1-V4) are sensitive to learning and altered by top-down knowledge (Damaraju, Huang, Barrett, & Pessoa, 2009; Li, Piëch, & Gilbert, 2004).

With respect to conceptual knowledge, learning about a novel category has consistently been shown to facilitate the recognition of objects ADDIN EN.CITE (Collins & Curby, 2013; Curby, Hayward, & Gauthier, 2004; Gauthier, James, Curby, & Tarr, 2003) and changes the discriminability of faces′ category-specifying features (Goldstone, Lippa, & Shiffrin, 2001). Detailed semantic knowledge, such as stories about a stimulus, can facilitate the recognition of objects and faces, and such influences manifest as early 100 ms after visual exposure (Abdel-Rahman & Sommer, 2008; Abdel-Rahman & Sommer, 2012). A brain region central to object and face perception, the fusiform gyrus (FG), is sensitive to such knowledge and learning (Tarr & Gauthier, 2000) and readily modulated by perceptual ′priors′ and top-down expectation signals from ventral-frontal regions, notably the orbitofrontal cortex (OFC) ADDIN EN.CITE (Bar, 2004; Bar, Kassam, et al., 2006; Freeman et al., 2015; Summerfield & Egner, 2009) (see Fig. 3). For instance, when participants have an expectation about a face, top-down effective connectivity from the OFC to the FG is enhanced, suggesting that expectation signals available in the OFC may play a role in modulating FG visual representations (Summerfield & Egner, 2009; Summerfield et al., 2006). Moreover, when presented with objects, activity related to successful object recognition is present in the OFC 50-85 ms earlier than in regions involved in object perception, again suggesting a role for OFC expectation signals that may affect FG perceptual processing (Kveraga, Boshyan, & Bar, 2007).

We have shown that, when participants view faces, the representational structure of activity patterns in the FG (involved in face perception) partly reflects stereotypical expectations (Stolier & Freeman, 2016) and emotion concepts (Brooks, Chikazoe, Sadato, & Freeman, 2019). Such findings are consistent with growing evidence that perceptual representations in object-

processing brain regions do not reflect processing of visual cues alone, but additionally reflect abstract semantic relationships between perceptual categories ADDIN EN.CITE (Jozwik, Kriegeskorte, Storrs, & Mur, 2017; Khaligh-Razavi & Kriegeskorte, 2014). More generally, the FG has been shown to be sensitive to a variety of other top-down social cognitive processes, such as goals (Kaul, Ratner, & Van Bavel, 2013) and intergroup processes ADDIN EN.CITE (Brosch, Bar-David, & Phelps, 2013; Kaul, Ratner, & Van Bavel, 2014; Van Bavel, Packer, & Cunningham, 2008).

1.2 An extended DI model

How could we account for such findings and understand the conceptual scaffolding of perceiving social categories, emotions, and traits? Regarding the underlying representations involved, early models in social perception took an information-processing approach ADDIN EN.CITE (e.g., Brewer, 1988; Fiske & Neuberg, 1990; Hamilton, Katz, & Leirer, 1980; Smith, 1984; Srull & Wyer, 1989), viewing representations as discrete symbolic units manipulated through propositions and logical rules in what can be described as a 'physical symbol system' (Newell, 1980). This included the popular 'spreading activation' associative networks, which are highly valuable in understanding phenomena such as stereotype activation and priming ADDIN EN.CITE (e.g., Blair & Banaji, 1996; Dovidio, Evans, & Tyler, 1986).

Distributed neural-network models (and localist approximations) in social cognition ADDIN EN.CITE (Freeman & Ambady, 2011; Kunda & Thagard, 1996; Read & Miller, 1998; Smith & DeCoster, 1998), including the DI model, assume that representations are not encapsulated by any single static unit, but instead reflect a unique pattern distributed over a population of units. It is the distributed pattern, dynamically re-instantiated in every new

instance, that serves as the unique ʻcodeʼ for a given social category, stereotype, trait, or memory. Such models have considerably higher neural plausibility (Rumelhart, Hinton, & McClelland, 1986; Smolensky, 1989), as multi-cell recordings have now made clear it is the communal activity of a population of neurons – a specific pattern of firing rates – that provides the ʻcodeʼ for various kinds of sensory and abstract cognitive information (i.e., a 'population code'; Averbeck, Latham, & Pouget, 2006).

The DI model (Freeman & Ambady, 2011) is a recurrent connectionist network with stochastic activation ADDIN EN.CITE (McClelland, 1991; McClelland & Rumelhart, 1981; Rogers & McClelland, 2004). When a face is presented to the network, facial feature detectors in the *cue* level activate social categories in the *category* level, which in turn activate stereotype attributes in the *stereotype* level; in parallel, top-down attentional processes activate task demands in the *higher-order* level, which amplify or attenuate certain pools of social categories in the *category* level. At every moment in time, a node has a transient activation level, which can be interpreted as the strength of a tentative hypothesis that the node is represented in the input. After the system is initially stimulated by bottom-up and top-down inputs (e.g., a face and a given task demand), activation flows among all nodes at the same time (as a function of their particular connection weights).

Because processing is recurrent and nodes are all bi-directionally connected, this results in a dynamic back-and-forth flow of activation among many nodes in the system, leading them gradually to readjust each otherʼs activation as they mutually constrain one another over time. This causes the system to gradually stabilize over time onto an overall pattern of activation that best fits the inputs and maximally satisfies the systemʼs constraints (the inputs and the relationships among nodes). The model thereby captures the notion that social category

perceptions dynamically evolve over fractions of a second, emerging from the interaction between bottom-up sensory cues and top-down social cognitive factors. As such, perceptions of social categories are a gradual process of negotiation between visual cues and perceiver knowledge. Although many neural systems would be involved, recent extensions of the DI framework (Freeman & Johnson, 2016) propose that this integration of bottom-up and top-down information in initial social perception centers on the interplay of the FG (involved in face perception), OFC (involved in top-down expectation signals), and anterior temporal lobe (ATL; involved in the storage and retrieval of semantic associations; Olson, McCoy, Klobusicky, & Ross, 2012) (see Fig. 3).

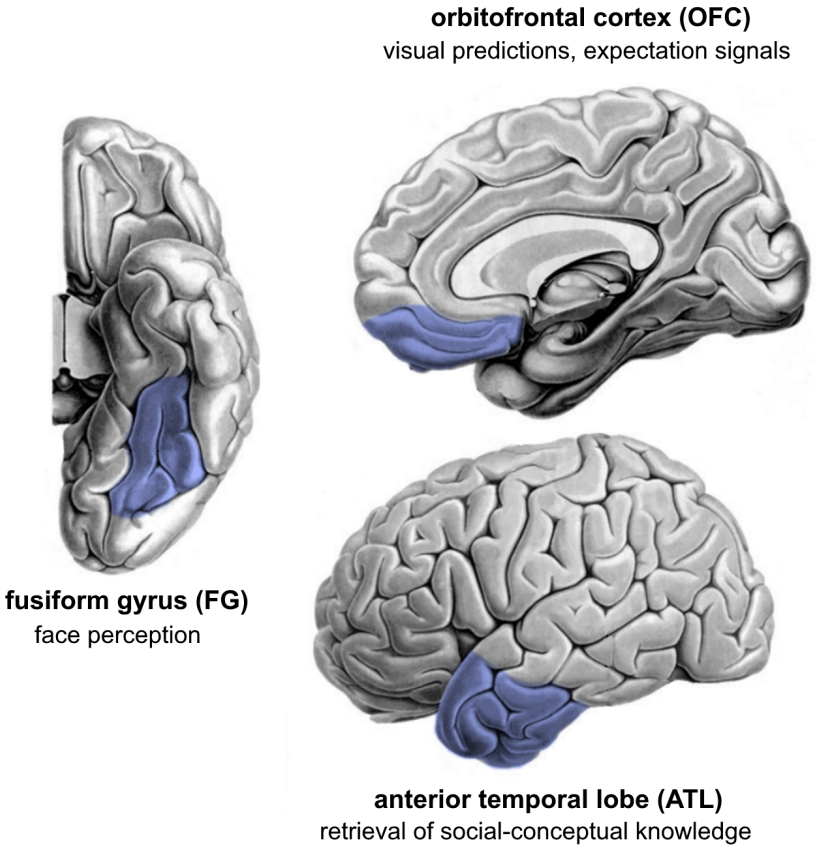

Figure 3. Freeman and Johnson (2016) posited that the fusiform gyrus (FG), orbitofrontal cortex (OFC), and anterior temporal lobe (ATL) together play an important role in the coordination of

sensory and social processes during perception. The FG is centrally involved in visual processing of faces, the ATL broadly involved in semantic storage and retrieval processes, and the OFC involved in visual predictions and top-down expectation signals. In this perspective, when perceiving another person′s face, evolving representations in the FG lead the ATL to retrieve social-conceptual associations related to tentatively perceived characteristics. This social-conceptual information available in the ATL, in turn, is used by the OFC to implement top-down visual predictions (e.g., based on social-conceptual knowledge) that can flexibility modulate FG representations of faces more in line with those predictions. Such a network would support a flexible integration of bottom-up facial cues and higher-order social cognitive processes. Adapted from Freeman and Johnson (2016).

In order to extend from social category perception to a more comprehensive system for emotion perception[1] and trait perception as well, we can conceive of the *category* and *stereotype* levels as a single level (Fig. 4). While separate levels in the original model, the *category* and *stereotype* levels both reflect knowledge structures or attributes; combining them into a single level is nearly functionally equivalent from the perspective of the model. As in Fig. 4, this single level of an extended DI model would include categories (e.g., Male, Asian), emotions (e.g., Happy, Angry), and traits (e.g., Trustworthy, Dominant). Stereotype attributes (e.g., Aggressive, Caring) in this case are equivalent to traits (also see Kunda & Thagard, 1996). As in the original model, nodes that are associatively consistent (e.g., Male – Aggressive, Trustworthy – Likeable, Happy – Trustworthy) have positive excitatory connections, and those that are inconsistent (e.g., Female – Aggressive) have negative inhibitory connections; those unassociated have no connection. Based on task instructions in a particular context, higher-order task demand nodes (e.g., Race Task Demand, Emotion Task Demand, Dominance Task Demand) will excite the pool of nodes relevant for the task (i.e., the response set) and inhibit those irrelevant for the task,

[1] The DI model (Freeman & Ambady, 2011) does account for the perception of emotion categories, but the simulations and discussion in the original work was focused on emotion categories′ interaction with gender and racial stereotypes. In the current work, we describe how the model can be leveraged to understand the role of conceptual knowledge in emotion perception more comprehensively.

thereby allowing certain sets of social categories, emotions, or traits to dominate attention in service of the context or judgment currently at hand (see Fig. 4).

1.3 Social-conceptual structure becomes perceptual structure

A central implication of the DI model and its extension here is that social-conceptual structure is always in an intimate exchange with perceptual structure, and thus how we think about social groups (i.e., stereotypes), emotions, or personality traits helps determine how we visually perceive them in other people. One well-studied example is "race is gendered" effects, whereby perceptual judgments of Black faces are biased toward Male judgments and Asian faces biased toward Female judgments. These perceptual effects have been demonstrated using a variety of paradigms and have been related to individual differences in the strength of overlapping stereotype associations between Black and Male stereotypes (e.g., aggressive, hostile) and Asian and Female stereotypes (e.g., docile, communal) (Johnson et al., 2012). Recently such social-conceptual biasing of perceiving gender and race was shown to be reflected in neural-representational patterns of the FG region (involved in face perception) while viewing faces, showing that it is reflected in the basic perceptual processing of those faces (Stolier & Freeman, 2016). Simulations with the DI model showed that such effects naturally arise out of the recurrent interactions between cue, category, and stereotype representations inherent to the system (Freeman & Ambady, 2011; Phenomenon 3). As facial cues (e.g., larger jaw) activate categories (e.g., Male) that in turn activate stereotypes (e.g., aggressive) during perception, all conceptually related attributes (i.e., stereotypes) begin feeding back excitatory and inhibitory pressures to category representations (e.g., Male and Female) and lower-level cue representations (e.g., larger jaw and round face). This has the effect of causing the conceptual similarity of any

two categories (or by extension, any two emotions or two traits) to scaffold that pair's perceptual similarity.

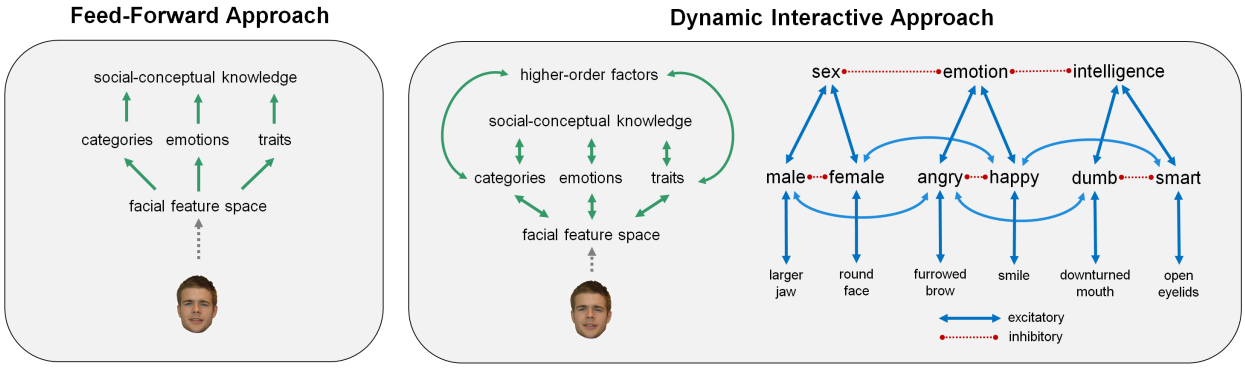

Figure 4. In a Feed-Forward Approach, facial features are represented in a facial feature space, which in turn activates social categories, emotions, and traits, thereafter activating related social-conceptual knowledge and impacting subsequent processing and behavior. In an extended DI framework, during perception, as facial features begin activating categories (e.g., Male), emotions (e.g., Angry), and/or traits (e.g., Smart), related social-conceptual attributes will be activated as well, but they will also feed excitatory and inhibitory pressures back on the earlier activated representations. The continuous, recurrent flow of activation among all internal representations of categories, emotions, traits, and social-conceptual attributes (here all organized into a single processing level) leads social-conceptual knowledge to have a structuring effect on perceptions and even featural representation. During this process, higher-order task demands (e.g., sex, emotion, intelligence) amplify and attenuate representations so as to bring task-relevant attributes to the fore for the specific task context at hand. Note that this depiction is highly simplified; a limitless number of other attributes and their connections could be included, and a number of excitatory and inhibitory connections are omitted here for simplicity. Also note that facial feature space be modeled using a range of approaches from simplified sets of facial features, as seen here, to more complex computational approaches based on the brain's visual-processing stream (Riesenhuber & Poggio, 1999); multiple levels of visual processing could be included and not all levels of visual processing need be bidirectional.

Our recent research has tested this social-conceptual scaffolding of perceiving faces in the context of perceiving social categories, emotions, and traits more comprehensively using a technique called representational similarity analysis (RSA). Using RSA, we can examine how representational structure (i.e., the pairwise similarities among representations) is conserved across conceptual, perceptual, and/or physical representational spaces to test whether conceptual

structure is reflected in perceptual structure, even when acknowledging the contribution of physical structure in faces themselves (for more on the approach, Freeman et al., 2018). Indeed, in one set of studies examining social category perception, we found that for any given pair of gender, race, or emotion categories (e.g., Black and Male, Female and Happy), a greater biased similarity in stereotype knowledge between the two categories was associated with a greater bias to perceive faces belonging to those categories more similarly (Stolier & Freeman, 2016) (see Fig. 5A).

To assess perceptual similarity in these studies, we used computer mouse-tracking, which provides a window into the real-time dynamics leading up to social perceptual judgments (or any other kind of forced-choice response; Freeman, 2018). Specifically, by examining how a participant′s hand settles into a response over time, and may be partially pulled toward other potential responses along the way, numerous studies have leveraged mouse-tracking to chart out the real-time dynamics through which social categories, emotions, stereotypes, attitudes, and traits activate and resolve over hundreds of milliseconds ADDIN EN.CITE (for reviews, Freeman, 2018; Freeman, Dale, & Farmer, 2011; Stillman, Shen, & Ferguson, 2018). During two -choice tasks (e.g., Male vs. Female), deviation in a subject′s hand trajectory toward each category response provides an indirect measure of the degree to which that category was activated during perception. If conceptual knowledge links one category to another (e.g., Black to Male), subjects′ perceptions should be biased toward that category and, consequently, their hand trajectories should deviate toward that category response. Thus, a greater deviation in hand movement toward the opposite response serves as a measure of the degree to which the opposite response category was co-activated with the chosen response, and in turn, how similarly the current stimulus is perceived as consistent with the opposite response – even if not explicitly

selected as such (see Fig. 5).

We recently conducted a similar comprehensive test of the conceptual scaffolding of emotion. In a series of studies examining the six commonly studied emotion categories Anger, Fear, Disgust, Happiness, Sadness, and Surprise, we demonstrated that when subjects held any two emotion categories more conceptually similar, they showed a tendency to perceive the categories′ corresponding facial expressions more similarly (Brooks et al., 2019; Brooks & Freeman, 2018). In some studies, perceptual similarity was assessed using mouse-tracking, such as when a perceiver′s greater conceptual similarity between Anger and Disgust leads to a greater attraction to the ″Disgusted″ response for an angry face (or vice-versa) (Fig. 5B). Another study used a reverse correlation technique, which allows a visual estimation of the cues that individuals expect to see for a given face category ADDIN EN.CITE (Dotsch, Wigboldus, Langner, & van Knippenberg, 2008; Todorov, Dotsch, Wigboldus, & Said, 2011). By superimposing random noise patterns over a single base face and having subjects select across many trials which of two noise-altered face images appear to be convey Anger vs. Disgust, for example, averaging the noise patterns can reveal an estimate of what Anger or Disgust appears in the mind′s eye of the subject. Using this technique, we were able to visualize perceivers′ visual prototypes for the six emotions. The results converged with the mouse-tracking findings, revealing that any pair of emotions deemed conceptually more similar in the mind of a perceiver yielded more physically similar visual prototypes (as measured through independent ratings and the physical similarity of the prototype images themselves).

Additional research found that, for both social category and emotion perception, this conceptual shaping of perceptual structure was evident in neural patterns of regions important for face perception (FG) when perceivers viewed faces. Further, the correlation of conceptual

structure and perceptual structure held above and beyond any inherent physical resemblances in the face stimuli themselves ADDIN EN.CITE (Brooks et al., 2019; Stolier & Freeman, 2016). Such findings suggest that the locus of conceptual shaping of perceptual structure is at relatively early perceptual stages of processing, rather than reflecting a mere response bias or post-perceptual decision processes. Finally, an additional set of studies tested the influence of conceptual similarity on perceptual similarity in face-based trait impressions as well. Using multiple techniques, including perceptual ratings and reverse correlation, here again we found that that an increased tendency to believe two traits (e.g., openness and agreeableness) are more similar conceptually predicted a greater similarity in the actual facial features used to make inferences about those traits, e.g., what makes a face appear open or agreeable to a perceiver (Stolier, Hehman, Keller, Walker, & Freeman, 2018) (Fig. 5C).

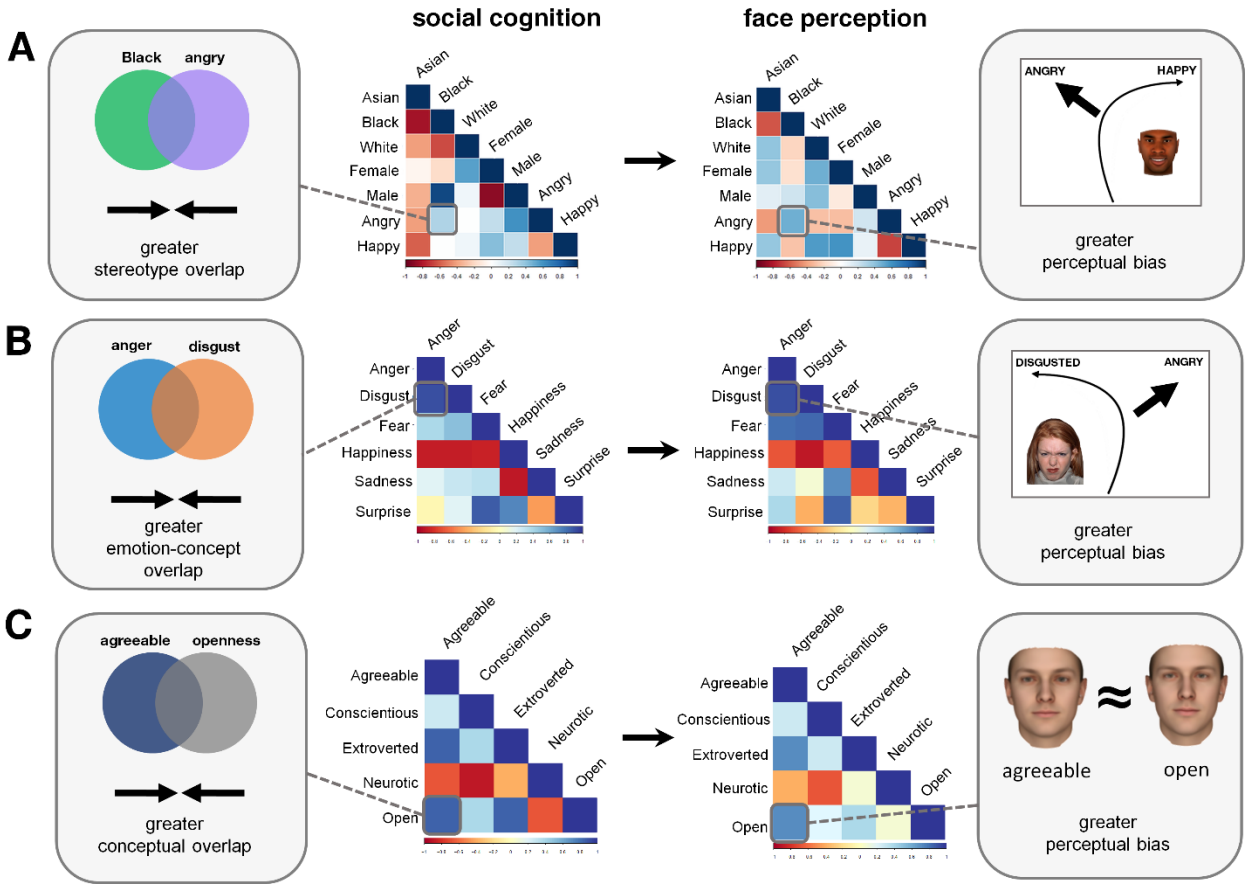

Figure 5. Social-conceptual structure shapes face perception. Dissimilarity matrices (DMs) comprise all pairwise similarities/dissimilarities and are estimated for both conceptual knowledge and perceptual judgments. Unique values under the diagonal are vectorized, with each vector reflecting the structure of the representational space, and a correlation or regression then tested the vectors′ relationship. **(A)** Participants′ stereotype DM (stereotype content task) predicted their perceptual DM (mouse-tracking), showing that a biased similarity between two social categories in stereotype knowledge was associated with a bias to see faces belonging to those categories more similarly, which in turn was reflected in FG neural-pattern structure (Stolier & Freeman, 2016). **(B)** Participants′ emotion-concept DM (emotion ratings task) predicted their perceptual DM (mouse-tracking), showing that an increased similarity between two emotion categories in emotion -concept knowledge was associated with a tendency to perceive those facial expressions more similarly (Brooks & Freeman, 2018), which was also reflected in FG pattern structure (Brooks et al., 2019). **(C)** Participants′ conceptual DM (trait ratings task) predicted their perceptual DM (reverse correlation task), showing that an increased tendency to believe two traits are conceptually more similar is associated with using more similar facial features to make inferences about those traits (Stolier, Hehman, Keller, et al., 2018). Figure adapted from Freeman et al. (2018).

The DI framework could parsimoniously account for such findings through a single recurrent system wherein perceptions of social categories, emotions, and traits all emerge out of the basic interactions among cues, social cognitive representations, and higher-order cognitive states (see Fig. 4). Below, we contextualize this perspective by reviewing in greater depth recent research on perceiving social categories, emotions, and traits, including the role that social-conceptual knowledge and other social cognitive processes play. Surely, the phenomena of social categorization, emotion perception, and trait inference have important differences; at the same time, the DI approach argues that theoretical and empirical advances may be gained by conceiving of these as emerging from a single recurrent system for social perception that relies on domain-general cognitive properties (at least certainly insofar as these phenomena operate as social perceptual judgments). Perceptions of social categories, emotions, and traits are all scaffolded by social-conceptual knowledge in similar fashion because they emerge out of basic domain-general interactions among cues, social cognitive representations, and higher-order cognitive states.

2. Perceiving Social Categories

Given the complexity of navigating the social world, people streamline mental processing by placing others into social categories. Perceivers maintain conceptual categories of others, each tied to rich sets of information that streamline our ability to predict behavior. These categories span any dimension along which we divide one another, such as demographic categories including race, gender, and age (Macrae & Bodenhausen, 2000), abstract in- and out-groups (Tajfel, 1981), and cultural and occupational groups (Fiske, Cuddy, Glick, & Xu, 2002). Seminal

work by Allport (1954) argued that individuals perceive others via spontaneous, perhaps inevitable, category-based impressions that are highly efficient and designed to economize on mental resources. As described earlier, since then, a vast array of studies has demonstrated that such category-based impressions bring about a host of cognitive, affective, and behavioral outcomes, changing how we think and feel about others and behave towards them, often in ways that may operate non-consciously ADDIN EN.CITE (e.g., Bargh & Chartrand, 1999; Brewer, 1988; Devine, 1989; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio, Jackson, Dunton, & Williams, 1995; Fiske, Lin, & Neuberg, 1999; Gilbert & Hixon, 1991). A traditional emphasis has therefore been to document the downstream implications of person categorization and its myriad outcomes for social interaction.

About 15 years ago, social psychologists began to examine the perceptual determinants of social categorization, such as how processing of stimulus features maps onto higher level stages of the social categorization pipeline. For example, one series of studies showed that perceivers more efficiently extract facial category vs. identity cues, which was interpreted as perhaps an important factor setting the stage for categorical thinking at later stages of person perception (Cloutier, Mason, & Macrae, 2005). The downstream consequences of perceiving category cues were further evidenced by findings showing that such cues can function independently of category membership itself in automatic evaluations (Livingston & Brewer, 2002) and stereotypic attributions ADDIN EN.CITE (Blair, 2002; Blair, Judd, Sadler, & Jenkins, 2002). Moreover, category-relevant features in isolation (e.g., hair) were shown to even automatically trigger category activation (Martin & Macrae, 2007), and even within-category variation in the prototypicality of race-related cues ADDIN EN.CITE (Blair, Judd, & Fallman, 2004; Blair et al., 2002; Freeman, Pauker, Apfelbaum, & Ambady, 2010) or sex-related cues

(Freeman, Ambady, Rule, & Johnson, 2008) have been shown to powerfully shape perceptions.

One consequence of such within-category variation – the natural diversity in the category cues of our social world – is that it often leads multiple categories to become simultaneously active during initial perceptions (Freeman & Ambady, 2011; Freeman & Johnson, 2016). Moreover, such social category co-activations, often indexed using the mouse-tracking technique described earlier, may not just be innocuous peculiarities of the perceptual system, but instead consequential social perceptual phenomena with tangible downstream impacts. These partial and parallel co-activations of categories, not observed in explicit responses, lead to differences in stereotyping ADDIN EN.CITE (Freeman & Ambady, 2009; Mason, Cloutier, & Macrae, 2006) and social evaluation ADDIN EN.CITE (Johnson, Lick, & Carpinella, 2015; Livingston & Brewer, 2002). For instance, Black individuals with more prototypically Black faces tend to receive harsher criminal sentences (Blair et al., 2004a, 2004b) including capital punishment (Johnson, Eberhardt, Davies, & Purdie-Vaughns, 2006). Similarly, American female politicians with less prototypically female facial features (i.e. more masculine cues) are less likely to be elected in conservative American states (Carpinella, Hehman, Freeman, & Johnson, 2016; Hehman, Carpinella, Johnson, Leitner, & Freeman, 2014), and this effect is predicted by the perceptual biasing effect that occurs when individuals categorize the politicians by sex (i.e. co-activation of the Male category; Hehman, Carpinella, et al., 2014).

In addition to within-category variation, social perception is also sensitive to a number of other forms of extraneous perceptual input in the environment. For example, race categorization shows sensitivity to context such that targets are more likely to be categorized as White or Asian if they are seen in a culturally-congruent visual context (Freeman, Ma, Han, & Ambady, 2013). Even cues inherent to the individual (e.g., hair and clothing) can supply a source of expectation

and prediction that may impact face processing. For example, clothing can bias race categorization by exerting a contextual cue to the social status of an individual, eliciting visual predictions about the person′s race. One study presented subjects with faces morphed along a Black–White continuum, each with low-status attire (e.g., a janitor uniform) or high-status attire (e.g., a business suit). Subjects categorized the faces as White or Black while their mouse trajectories were recorded. The study found that low-status attire biased perceptions toward the Black category while high-status attire biased perceptions toward the White category. When race and status were stereotypically incongruent (e.g., a White face with low-status attire or a Black face with high-status attire), participants′ mouse movements showed a continuous attraction to the opposite category, suggesting that the social status associated with clothing exerted a top-down influence on race categorization (Freeman et al., 2011).

In addition to external cues in the environment, social perception also shows sensitivity to inputs from the perceiver. These include motivations and expectations that bias the processing of novel stimuli, as well as preexisting perceptual heuristics used to make sense of ongoing sensory input. One such abstract top-down factor that can impact social perception is a perceiver′s own goals and motivations, which can bear weight on perception even when they reside outside of conscious awareness. In this sense, perception is chronically "motivated" to pick up on whatever aspects of the environment are most relevant or useful to current processing goals. For example, transient sexual desire can increase the speed and accuracy of sex categorization (Brinsmead-Stockham, Johnston, Miles, & Macrae, 2008). Notable and consequential effects of motivated social categorization occur in the case of race perception. For example, situations of economic scarcity lead White subjects to rate Black faces as more Black and to more often rate mixed-race faces as Black (Krosch & Amodio, 2014). Studies have found that subjects are more likely to

identify an impoverished image of a gun as a gun when primed with a Black face, due to stereotypical associations (Eberhardt, Goff, Purdie-Vaughns, & Davies, 2004). Similar effects emerge for group identity, which produces a strong chronic motivational state to perceptually categorize others differently based on their in- or out-group status (Xiao, Coppin, & Van Bavel, 2016a, 2016b). For example, political group identity leads subjects to represent biracial candidates as lighter or darker in skin-tone if they are in the same or different political group, respectively (Caruso, Mead, & Balcetis, 2009).

2.1. Conceptual influences in social categorization

Stereotypes are merely conceptual knowledge related to social categories, and they have been extensively studied in social psychology and traditionally considered to be triggered after categorizing a target person ADDIN EN.CITE (Allport, 1954; Brewer, 1988; Fiske & Neuberg, 1990; Macrae & Bodenhausen, 2000). As described earlier, only fairly recently have approaches considered the influence that stereotypes can have on a visual percept before it has fully stabilized ADDIN EN.CITE (Freeman & Ambady, 2011; Freeman & Johnson, 2016; MacLin & Malpass, 2001). One set of studies demonstrated that prior race labels alter the perceived lightness of a face, such as knowledge that a person is Black making a face′s skin tone appear darker (Levin & Banaji, 2006). Another set of studies found that racially ambiguous faces were more likely to be categorized as Black and judged to have Afrocentric facial features, if they had a stereotypically Black hairstyle (MacLin & Malpass, 2001). As suggested by the DI model′s simulations with the analogous status stereotype effects on race perception via attire cues above (Freeman, Penner, et al., 2011), such effects of hairstyle cues are likely driven by conceptual stereotype associations. But in considering such findings of stereotypical contexts, it is difficult

to know whether biased perceptual decisions reflect a bias on perception itself or merely at a post-perceptual decision stage. This would suggest stereotypes affect how perceivers think about the targets but not how they "see" them.

While a post-perceptual explanation cannot be entirely ruled out, recent work has been able to more closely investigate stereotype impacts on social category perception by examining how stereotypes bind ostensibly unrelated categories together (Freeman & Ambady, 2011; Freeman & Johnson, 2016). Just as hairstyle or other visual cues can shape social categorization by activating conceptual associations, so can one category (e.g., Black) serve as context for perception of another category (e.g., Male), even one on a seemingly unrelated dimension. This line of work has demonstrated the inherent intersection of race and sex, such that certain pairs of race and sex categories share stereotypes (e.g. the categories Asian and Female sharing conceptual associations with docility and submissiveness; the categories Black and Male sharing conceptual associations with hostility and physical ability) and are biased to be perceived concurrently as a result (Carpinella, Chen, Hamilton, & Johnson, 2015; Johnson, Freeman, & Pauker, 2012). An important consequence is that individuals who do not meet the expected stereotype-congruent combination of social categories (e.g. Asian men and Black women) are the subject of biased stereotypic expectations that can negatively influence their experiences in dating, university life, and the workforce ADDIN EN.CITE (Galinsky, Hall, & Cuddy, 2013).

Providing evidence for top-down conceptual structuring can be difficult when intrinsic physical resemblances are also at play. For example, stereotypes prescribe men as angrier and women as happier, and men's faces are more readily perceived angry and women's faces more readily perceived happy ADDIN EN.CITE (Hess et al., 2004; Hess et al., 2000). However, e volutionary psychologists have suggested this to be driven by intrinsic physical overlap in the

facial features specifying anger and masculinity (e.g., furrowed brow) and joy and femininity (e.g., roundness) (Becker, Kenrick, Neuberg, Blackwell, & Smith, 2007). DI model simulations account for such physical resemblance effects well (Freeman & Ambady, 2011; Phenomenon 4) – regardless of whether they exist due to distal evolutionary pressures (e.g., for men to be dominant and women to be submissive; Becker et al. 2007) or simply arbitrary physical covariation.

Nevertheless, given both potential factors at play, it is difficult to isolate specific top-down stereotypic factors driving the perceptual privileging of male anger and female joy. However, unconstrained, data-driven tasks have been valuable for isolating the effect of stereotypes on binding sex and emotion categories together (Brooks & Freeman, 2018). In one set of studies, we used the reverse correlation technique described earlier to produce each subject′s visual prototype faces for the categories Male, Female, Angry, and Happy (Brooks, Stolier, & Freeman, 2018). We had independent raters judge the prototype Male, Female, Angry, and Happy faces on apparent sex and emotion. We found that the reverse-correlated face prototypes showed a systematic bias in their appearance that was consistent with stereotypes, with Female prototypes biased toward Happy (and vice-versa) and Male prototypes biased toward Angry (and vice-versa). In follow-up studies, we found that this effect was strongly predicted by a given individual′s conceptual associations between those categories. That is, the more that a subject harbored stereotype-congruent knowledge about sex and emotion categories (i.e., high overlap between Female-Happy and Male-Angry), the more likely were they to yield visual prototypes for those categories that were biased in appearance. Importantly, each category is attended to in isolation, making it unlikely that subjects were conceptually primed to produce biased responses in the initial reverse correlation task.

Neuroimaging can be highly valuable in addressing the question of whether top-down effects exist at perceptual vs. post-perceptual processing stages, in that it can identify which levels of neural representation top-down impacts manifest (Freeman et al., 2018; Stolier & Freeman, 2015). In two studies, we measured the overlap of social categories at three levels: in their conceptual structure as related by similar trait stereotypes, measured via explicit surveys; in their visual perception from faces, through a perceptual categorization task; and in their neural representation, by comparing the similarity of the categories′ representational patterns across the brain (Stolier & Freeman, 2016). Conceptual similarity was measured as the similarity in stereotype associations of each category, e.g., where the categories Black and Male may be high in hostility and sociability stereotypes but low in affectionate stereotypes. Perceptual similarity was measured with mouse-tracking, where participants categorized faces along each category in a two-choice task (e.g., Male vs. Female), and similarity was calculated as the degree to which participant mouse trajectories were drawn towards any one category response regardless of their final response (e.g., trajectories drawn towards 'Male' while categorizing Black female faces). Lastly, we measured neural similarity of each category as the similarity in the multi-voxel neural patterns of each category-pair.

Indeed, in both studies, we found that social categories more similar in conceptual knowledge were perceived more similarly. Category-pairs more related in stereotypes were also more interdependent during perceptual categorization. For instance, consistent with prior work, the stereotype content task indicated greater conceptual overlap between the Black and Male categories than Black and Female categories (Johnson et al., 2012). As seen in Fig. 5A, when categorizing faces belonging to these categories (e.g. Black female faces by sex), participants

were more drawn towards the stereotype-consistent category response, regardless of their explicit response (e.g., mouse trajectories were drawn more toward Male en route to the Female response). Moreover, these conceptually entangled category-pairs were also more similar in their multi-voxel neural patterns in regions involved in face perception (FG) and top-down expectation (OFC). These findings survived analyses that controlled for potential physical similarity of the faces themselves. This suggested that, even in regions important for basic face perception, a face′s social categories are shaped by social-conceptual knowledge as well, namely stereotypes about those categories (Stolier & Freeman, 2016).

2.2. Summary

Although social categorization was long treated as a starting point and only its downstream products took theoretical center-stage, the past 15 years have increasingly zoomed in on the categorization process. Such work has found that perceptions of face′s social categories are susceptible to a range of social cognitive factors, such as stereotypes, attitudes, and goals, which are often presumed to operate only downstream of categorization. Stereotypes, i.e., social-conceptual knowledge, can have a pronounced impact in structuring perceptions, and growing findings confirm the close interplay between perceiver knowledge and facial features in driving initial perceptions – a premise central to the framework outlined here.

3. Perceiving Emotions

Humans have the impressive social-perceptual ability to infer someone else′s emotional state from perceptual information on their face: a scowling person looks angry, a frowning person looks sad, a smiling person looks happy. The perceptual operations that lead to

categorizations of others′ emotional states are just as transparent as those that lead to other social categorizations such as gender or standard object categorizations – no effortful deliberation is required to perceive emotion from facial expressions. And yet, perceiving and categorizing emotions in others affords incredibly rich social inferences, allowing us to anticipate others′ future actions and mental states and plan our own behavior accordingly.

Due to the ease and fluency with which we make rich inferences from facial actions, there is long-standing interest in facial expressions and how they are perceived – experimental psychologists have been studying facial emotion since the field′s inception in the 19th century. Early theoretical assumptions and intuitions about facial emotion were largely influenced by those set out in Darwin′s 1872 book *The Expression of the Emotions in Man and Animals*. Darwin viewed the study of facial expressions as a test case for the theory of evolution, and wanted to discover and document potential evolutionary "principles" for the existence of facial expressions. Darwin pioneered a number of methods, and made a number of theoretical assumptions, that persist in the field today. These include aiming to build a taxonomy of facial expressions, examining which facial expressions people can reliably recognize by having them categorize pictures, and assuming that cross-cultural data can address questions of innateness or universality in facial expressions (Darwin, 1872; Gendron & Barrett, 2017). This approach inspired an early body of empirical work which largely studied facial emotion perception by having subjects place static posed images of facial expressions into a fixed set of categories. These studies built taxonomies of facial displays that could be reliably "recognized" as specific emotions, and explored the boundary conditions that influenced perceiver agreement ADDIN EN.CITE (Allport, 1924; Feleky, 1914; for a review of this early period of research, see Gendron & Barrett, 2017).

Darwin′s approach persisted further into the 20th century with the highly influential ″basic emotion″ approach ADDIN EN.CITE (Ekman, 1972; Ekman & Cordaro, 2011; Izard, 1971; Izard, 2011; Tracy & Randles, 2011). Ekman (1972) had a particularly influential approach to studying facial expressions. This involved closely associating facial actions with information about the underlying facial musculature, and delineating the specific combinations of facial actions (′facial action units′) that lead subjects to categorize a face as an emotion like A ngry or Afraid (Friesen & Ekman, 1978). The goal of this research was to build taxonomies of emotions that could be considered psychologically ″basic″ by studying consensus between perceivers in how facial expressions were categorized. Informed by greater study of the facial expressions themselves, studies continued to mainly consist of showing posed facial expressions to subjects who were asked to label them (Ekman, Friesen, & Ellsworth, 2013). A great deal of work using this approach shows that perceivers are typically fast, accurate, and largely consensual in their categorizations of facial expressions associated with a small number of ″basic″ emotions ADDIN EN.CITE (most commonly Anger, Disgust, Fear, Happiness, Sadness, and Surprise; Ekman & Friesen, 1971; Ekman, Sorenson, & Friesen, 1969; Izard, 1971; Tracy & Randles, 2011). In general, this approach assumes that the facial expressions associated with the ″basic″ emotions are so evolutionarily old and motivationally relevant that they trigger a direct ″read-out″ of visual features that should be fairly invariant between individuals (Smith, Cottrell, Gosselin, & Schyns, 2005). However, a growing body of work instead suggests that there are a number of contextual and perceiver-dependent factors that weigh in on how facial expressions are visually perceived.

For instance, research shows that facial emotion perception is extremely sensitive to – and even shaped by – the surrounding context. These contextual factors can be as simple as visual aspects of the person displaying an emotion (e.g. their body), multimodal aspects like the person′s voice, the surrounding scene, or more abstract characteristics of the context like the perceiver′s current goals. This body of work is a major factor motivating more recent theories of emotion to treat facial emotion perception as an embedded and situated phenomenon (e.g., Wilson-Mendenhall, Barrett, Simmons, & Barsalou, 2011), and to consider different ways of studying it as a result.

In one sense, it is entirely unsurprising that facial emotion perception would be heavily influenced by the surrounding context, since most instances of perceiving facial emotion occur in particular social contexts or scenes, alongside vocal and bodily cues that convey a wealth of information. But these findings are a serious challenge to classic views of emotion that heavily emphasize diagnosticity of facial cues due to the surprising strength of the effects. In many cases, aspects of the visual and even auditory context can completely dominate input from the face. For example, when someone′s body posture is incongruent with the emotion ostensibly signaled by their face, the ultimate emotion categorization is often consistent with bodily rather than facial cues ADDIN EN.CITE (for reviews, see de Gelder et al., 2005; Van den Stock, Righart, & de Gelder, 2007). While it is unclear whether this means that body posture really carries more diagnostic or important information about emotional states, it does indicate that perceivers heavily rely on cues from the body. Some work does indicate that bodily motion conveys information about specific emotion categories, since perceivers are highly consensual in their emotion categorizations for point-light displays (Atkinson, Dittrich, Gemmell, & Young,

2004).

Similarly notable effects have emerged for vocal cues, such that stereotypically Sad facial expressions are perceived as Happy when they are accompanied by a Happy voice, even when participants are instructed to disregard the voice (de Gelder & Vroomen, 2000). As with body cues, some researchers suggest that this reflects vocal cues being more diagnostic or informative about emotional states compared to the face (Scherer, 2003). A great deal of evidence also suggests that identical facial expressions are perceived differently depending on the visual scene in which they are encountered (e.g., a neutral context, such as standing in a field, or a fearful context, such as a car crash; Righart & De Gelder, 2008). Similar effects occur when participants are just given prior knowledge about the social context emotional facial expressions were originally displayed in (Carroll & Russell, 1996). Social information immediately present in a scene also can influence emotion perception, such that emotion perception is shaped by the facial expressions of other individuals in a visual scene (Masuda et al., 2008).

This growing body of evidence suggests that perceivers spontaneously make use of any information available to them to categorize someone else′s emotional state, and that the face is just one factor weighing in on these perceptions. This has led some researchers to propose that the face itself is ″inherently ambiguous″ (Hassin, Aviezer, & Bentin, 2013). Indeed, these results are widely consistent with insights from vision science that ambiguous stimuli are particularly subject to expectations and associations guided by the environment (Bar, 2004; Summerfield & Egner, 2009). At the very least, this work suggests that experimental designs using isolated posed facial expressions are not able to capture the full range of processes that weigh in on facial emotion perception.

3.1. Perceiver-dependent theories of emotion perception

Classic theories of emotion, most famously the "basic emotion" approach, assume that facial emotion perception occurs as a direct bottom-up read-out of facial cues that are inherently tied to their relevant emotion categories. For example, experiencing a given emotion such as "anger" yields a reliable and specific combination of facial cues that are able to be automatically extracted by a perceiver and effortlessly recognized as "anger" ADDIN EN.CITE (Ekman et al., 2013; Izard, 2011; Smith et al., 2005). An explicit assumption of these models of emotion is that the "basic" emotions are universally recognized across cultures (Ekman, 1972; Ekman & Friesen, 1971). However, the profound susceptibility of facial emotion perception to context – and the readiness with which perceivers make use of any contextual or associative content available to them in order to categorize facial expressions – has led recent theories of emotion and social perception to consider the idea that individual perceivers may serve as their own form of "context", allowing for substantial interindividual and cross-cultural variability in emotion perception.

The basic idea that aspects of the perceiver can sometimes influence emotion perception is not particularly controversial. A large body of work shows that dispositional factors such as social anxiety (Fraley, Niedenthal, Marks, Brumbaugh, & Vicary, 2006), stigma consciousness (Pinel, 1999), and implicit racial prejudice ADDIN EN.CITE (Hugenberg & Bodenhausen, 2003; Hutchings & Haddock, 2008) can impact visual processing of facial expressions. Recent approaches further argue that perceiver-dependence is a fundamental characteristic of emotion perception rather than an occasional biasing factor ADDIN EN.CITE (Barrett, 2017; Freeman &

Ambady, 2011; Freeman & Johnson, 2016; Lindquist, 2013). For example, the Theory of Constructed Emotion holds that facial displays of emotion can only be placed into a given category such as "anger" or "fear" when conceptual knowledge about those emotion categories is rapidly and implicitly integrated into the perceptual process (Barrett, 2017), which is highly consistent with the DI theory's premise of the conceptual scaffolding of various instances of social perceptual judgments. As discussed earlier, the DI framework predicts that a wealth of contextual and conceptual input implicitly informs perception before a face is placed into a stable response (e.g., category, emotion, or trait judgment), allowing for a great deal of influence from the conceptual structure of emotion categories on the ongoing processing of visual displays of emotion.

A natural consequence of this theoretical approach would be substantial variability between individuals, given the variety of different prior experiences, conceptual associations, and dispositional qualities that reside within each individual. As a result, perceiver-dependent theories place less of an emphasis on specific facial expressions being tied to specific discrete emotion categories. If one assumes variability is the norm in emotion, then taxonomies of "basic" emotions are more of a catalog of consensus judgments linked to particular category labels, rather than a definitive account of universal categories. Indeed, meta-analyses show a remarkable lack of consistency between individuals and studies in the neural representation of emotional experiences and perceptions (Kober et al., 2008; Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012) as well as their physiological signatures (Siegel et al., 2018) and associated facial actions in spontaneous displays (Durán & Fernández-Dols, 2018; Durán, Reisenzein, & Fernández-Dols, 2017). If this degree of variability exists in emotion experience

and expression, the perceptual system would have to be flexible, making rapid use of available contextual factors and cognitive resources to make sense of facial emotion displays. As a result, an emerging body of research has begun to directly investigate the role of conceptual knowledge and other such top-down factors influencing facial emotion perception.

3.2. Conceptual influences on emotion perception

A major thread in recent debates about emotion perception concerns the manner in which conceptual knowledge about emotions is involved in emotion perception. Classic theories would assume that categorizing a face as Angry (through a direct read-out of cues assumed to inherently signal anger) would lead to anger-related conceptual knowledge being subsequently activated in service of predicting an angry individual′s behavior. In contrast, if facial emotion perception were influenced by conceptual knowledge, that would require rapid and implicit access of conceptual associations before a percept has stabilized. Thus, most work on the relationship between conceptual knowledge and emotion perception has focused on manipulating access to conceptual knowledge and measuring how this impacts performance in standard emotion perception tasks. Much of this work involves the ″semantic satiation″ technique (Balota & Black, 1997; Black, 2001), in which target trials require subjects to repeat a word 30 or more times (e.g., in this case, ″angry″), temporarily reducing access to the associated concept, before making a response that the concept is hypothetically necessary for. When subjects have access to emotion concepts reduced this way, they show impaired accuracy in emotion categorization (Lindquist, Barrett, Bliss-Moreau, & Russell, 2006) and emotional facial expressions no longer serve as primes for other face stimuli from their category (Gendron, Lindquist, Barsalou, & Barrett, 2012), indicating a relatively low level perceptual role for conceptual knowledge. On the

other hand, increasing access to emotion concept knowledge increases speed and accuracy in emotion perception tasks (Carroll & Young, 2005; Nook, Lindquist, & Zaki, 2015) and shapes perceptual memory for facial expressions (Doyle & Lindquist, 2018; Fugate, Gendron, Nakashima, & Barrett, 2018). Additionally, semantic dementia patients, who have dramatically reduced access to emotion concept knowledge, seemingly fail to perceive discrete emotion at all, instead categorizing facial expressions by broad valence categories (Lindquist, Gendron, Barrett, & Dickerson, 2014).

Many attempts to measure the influence of conceptual knowledge on facial emotion perception have used language (i.e., emotion category labels like "anger") as a proxy for studying emotion concepts. More generally, language has an important role in constructionist theories of emotion due to the central constructionist concept that language is responsible for the common sense intuition that emotion is organized into discrete categories (Doyle & Lindquist, 2017; Lindquist, 2017). Certainly, language is an important factor in how we perceive and categorize emotion. A recent neuroimaging meta-analysis showed that when emotion category labels like "sadness" and "surprise" are incorporated into experimental tasks, the amygdala is less frequently active (Brooks et al., 2017). This supports the idea that facial expressions are ambiguous to some degree and that category labels provide immediate access to conceptual knowledge, reducing perceptual uncertainty. But since most existing work on conceptually scaffolded emotion perception has explicitly manipulated language or conceptual knowledge in tasks, it has been difficult to capture the implicit influence of conceptual knowledge that theories assume is involved in every instance of emotion perception.

As already discussed, assessing representational geometries using representational similarity analysis (RSA) is one way of globally measuring the overall influence of conceptual

knowledge without directly manipulating it (see Fig. 5). Existing uses of RSA to study emotion perception have been fruitful, suggesting that more abstract conceptual information may affect the perceptual representation of emotion. For example, neuroimaging work has been able to adjudicate between dimensional vs. appraisal models of how emotional situations are represented in the brain (Skerry & Saxe, 2015). This line of work also used RSA to show a correspondence between the neural representations of valence information from perceived human facial expressions and inferences from situations (Skerry & Saxe, 2014). One study measured the representational similarity between emotion categories in their perception from faces and voices (Kuhn, Wydell, Lavan, McGettigan, & Garrido, 2017), showing high correspondence between modalities even when controlling for low-level stimulus features. In general, these studies suggest that the representational structure of emotion perception may be shaped by more abstract conceptual features.

In one set of studies, we used RSA to measure the correspondence between subjects′ conceptual and perceptual representational spaces for commonly studied emotion categories – Anger, Disgust, Fear, Happiness, Sadness, and Surprise (Brooks & Freeman, 2018). All studies measured how conceptually similar subjects found each pair of emotions, and used this idiosyncratic conceptual similarity space to predict their perceptual similarity space. In two studies, we measured perceptual similarity using computer mouse-tracking. On each trial, subjects would see a face stimulus displaying a stereotyped emotional facial expression (e.g. a scowl for Anger) and have to categorize it as one of two emotion categories by clicking on response options on the screen (e.g. ′Anger′, ′Disgust′; one response option always corresponded to the intended/posed emotion display). We used mouse-trajectory deviation toward the unselected category response as a measure of perceptual similarity. We found that

conceptual similarity significantly predicted perceptual similarity, even when statistically controlling for intrinsic physical similarity in the stimuli themselves. Thus, the degree of conceptual similarity a subject attributed to a given pair of emotions (e.g. Anger and Disgust) predicted the degree of co-activation of the two categories during perception, even though there was ostensibly only one emotion being conveyed by the face stimulus (Fig. 5B). Moreover, this effect could not be explained by how similar the two categories are in their associated visual properties.

In an additional study, we repeated this approach, but used the reverse correlation technique described earlier to measure perceptual similarity. A given subject in this study was randomly assigned to an emotion category pair (e.g. Anger-Disgust) and asked to complete the reverse correlation task for these two categories, as well as a task to measure conceptual similarity between the categories. Perceptual similarity was measured by having independent raters judge pairs of images (each coming from the same subject) on how similar they were, as well as measuring the inherent visual similarity of the images themselves (on a pixel-by-pixel basis). We found that, when a subject held two emotion categories to be conceptually more similar, their reverse-correlated visual prototypes for these categories took on a greater physical resemblance. Reverse correlation allowed a less constrained test of the relationship between conceptual and perceptual similarity, since each subject was only attending to one emotion in isolation on the reverse correlation trials. As a result, the reverse correlation results provide strong evidence for conceptually scaffolded emotion perception since it is a data-driven task that did not rely on a particular stimulus set, emotion category labels, or any normative assumptions of how different facial emotion expressions should appear.

To identify at what level of neural representation such conceptual scaffolding of facial emotion manifests, subjects completed an fMRI task in which they passively viewed facial expressions of Anger, Disgust, Fear, Happiness, Sadness, and Surprise. Outside the scanner, we also again used a conceptual ratings task to measure the conceptual similarity of each pair of emotions. We found that neural-representational patterns in the face-processing FG region showed a representational structure that was significantly predicted by idiosyncratic conceptual structure. These findings demonstrate that representations of facial emotion categories in the brain′s perceptual system are organized in a way that partially conforms to how perceivers structure those categories conceptually. Such results are consistent with the stereotype scaffolding of a face′s social categories manifesting in the FG as well (Stolier & Freeman, 2016). These results show conceptual impacts on how the brain represents facial emotion categories at a relatively basic level of visual processing. Overall, this growing line of work suggests that the brain′s representation of facial emotion, or of a face′s social categories, do not reflect facial cues alone – they are also partly shaped by the conceptual meaning of those emotions or social categories.

3.3. Summary

The traditional view has been that there are a certain number of emotion categories that can be reliably and automatically recognized in humans, driven directly by facial features evolutionarily evolved to convey each emotion. However, research increasingly suggests that this approach has ignored idiosyncratic perceiver-dependent factors that shape emotion perception. Facial actions undeniably convey important information about internal states, but

there is little evidence that real-world instances of emotion experience yield specific and discrete facial displays like the ones usually studied in psychological research (Durán et al., 2017). Real-world facial displays of emotion are typically much more subtle and brief ADDIN EN.CITE (Barrett, Mesquita, & Gendron, 2011a; Durán et al., 2017; Russell, Bachorowski, & Fernández-Dols, 2003), and their interpretation requires a myriad of contextual and associative top-down factors to weigh in on visual processing. Growing evidence demonstrates that one such top-down factor is each perceiver's idiosyncratic conceptual knowledge about emotion, leading to a highly flexible process for facial emotion perception that may exhibit substantial variability between individuals. While these findings address broad and fundamental questions about the nature of emotion perception, they also dovetail with the findings outlined above on social category perception, and more broadly with the premises of the extended DI framework. Like social categories, and as predicted by the DI framework, emotion perception exhibits substantial flexibility through sensitivity to contextual and social-cognitive top-down factors. Additionally, it is likely that many of these perceptual mechanisms are not specific to emotion perception, but overlap with those involved in social categorization, trait impressions, and non-social perceptual categorizations alike.

4. Perceiving Traits

While we may take for granted that perceivers track social categories and emotions from faces, a more surprising domain of social perception involves the ability to readily infer someone's personality traits based solely upon their facial appearance. While we might assume that reading someone's disposition from their face constitutes an inaccurate snap judgment, research shows that these inferences are not arbitrary - they tend to be highly correlated across

multiple perceivers, even at brief exposures (Bar, Neta, & Linz, 2006; Willis & Todorov, 2006), and often occur automatically and beyond our conscious control (Zebrowitz & Montepare, 2008). For instance, responses in the amygdala, a subcortical region important for a variety of social and emotional processes, tracks a face′s level of perceived trustworthiness even when it is presented outside of conscious awareness using backward masking (Freeman, Stolier, Ingbretsen, & Hehman, 2014). Despite face-based impressions′ generally limited accuracy (Todorov et al., 2015; Tskhay & Rule, 2013), they can often powerfully guide our interactions with others and predict real-world consequences such as electoral outcomes (Todorov, Mandisodza, Goren, & Hall, 2005) or capital-sentencing decisions (Wilson & Rule, 2015), among others (for review, Todorov et al., 2015).

Outside of face-based impressions, social psychologists have long explored impression formation and trait attribution, dating back to Estes (1938) and Asch (1946). Decades of research explored the cognitive mechanisms involved in making dispositional inferences about others and other forms of social reasoning ADDIN EN.CITE (e.g., Skowronski & Carlston, 1987, 1989; Uleman & Kressel, 2013; Uleman, Newman, & Moskowitz, 1996; Winter & Uleman, 1984; Wyer Jr & Carlston, 2018), and still countless other studies explored "zero-acquaintance" judgments in interpersonal encounters that focused on judgmental accuracy in deducing others′ personality upon first meeting them ADDIN EN.CITE (e.g., Albright et al., 1997; Ambady, Bernieri, & Richeson, 2000; Ambady, Hallahan, & Rosenthal, 1995; Ambady & Rosenthal, 1992; Kenny, 1994; Kenny & La Voie, 1984). However, it was only fairly recently that social psychologists began to investigate more seriously face-based impressions in particular ADDIN EN.CITE (Bodenhausen & Macrae, 2006; Zebrowitz, 2006).

Researchers have now linked a large array of facial features with specific impressions, such as facial width, eye size and eyelid openness, symmetry, emotion, head posture, sexual dimorphism, averageness, and numerous others ADDIN EN.CITE (for reviews, Hehman et al., 2019; Olivola & Todorov, 2017; Todorov et al., 2015; Zebrowitz & Montepare, 2008). Discovering the links between all facial features and all traits is challenging, and computational and data-driven approaches can provide more comprehensive assessments (Adolphs, Nummenmaa, Todorov, & Haxby, 2016). Seminal research by Oosterhof and Todorov (2008) took such an approach to characterize the specific features that underlie a range of face impressions. In this work, participants viewed a large set of randomly varying computer-generated faces and evaluated the faces along different personality traits. Principal component analyses identified two fundamental dimensions: trustworthiness and dominance. These dimensions are consistent with the perspective that perceivers tend to place others along two primary dimensions: their intentions to help or harm (warmth) and their ability to enact those intentions (competence) ADDIN EN.CITE (Fiske, Cuddy, & Glick, 2007; Fiske et al., 2002; Rosenberg, Nelson, & Vivekananthan, 1968).

With respect to the cues underlying these two fundamental dimensions, Oosterhof and Todorov (2008) found that the trustworthiness dimension was characterized by faces varying in their baseline resemblance to traditionally happy vs. angry emotion expressions (even when displaying an ostensibly neutral emotional expression). The dominance dimension roughly corresponded to physical strength and facial maturity cues. Such findings can be partially explained by overgeneralization theory (Zebrowitz & Montepare, 2008), which posits that perceivers utilize functionally adaptive and evolutionarily shaped facial cues (e.g., emotion,

facial maturity) and "overgeneralize" to ostensibly unrelated traits (e.g., trustworthiness) due to the cue's association with that trait (e.g., trustworthiness from happy cues; dominance from age cues) ADDIN EN.CITE (Said, Sebe, & Todorov, 2009; Zebrowitz, Fellous, Mignault, & Andreoletti, 2003). Such research has made important advances in understanding the specific arrangements of facial features that reliably evoke particular trait impressions.

More recently, as with social categorization and emotion perception, researchers have begun to document the myriad factors harbored within perceivers that also help determine face impressions. Remarkably, at least for a number of common face impressions, the bulk of their variance is accounted for by idiosyncratic differences in how perceivers infer about faces ADDIN EN.CITE (Hehman et al., 2019; Hehman, Sutherland, Flake, & Slepian, 2017; Xie, Flake, & Hehman, 2018). Indeed, other research has demonstrated that the fundamental dimensions – trustworthiness and dominance – can shift or disappear entirely depending on perceiver factors. For instance, when judging female targets, dominance cues elicit more negative and untrustworthy evaluations, compared to male targets (Oh, Buck, & Todorov, 2019; Sutherland, Young, Mootz, & Oldmeadow, 2015), likely due stereotypic expectations of women as submissive, i.e. benevolent sexism (Glick & Fiske, 1996). Thus, trustworthiness and dominance dimensions cease being independent. On older adult faces, facial dominance comes to take on new meaning (e.g., wisdom) likely due to stereotypes of older adults' physical frailty, inconsistent with the notion of dominance and hostility (Hehman, Leitner, & Freeman, 2014). Perceptions of trustworthiness depend more or less on typicality or attractiveness facial cues depending on whether the target is from our own or a different culture (Sofer et al., 2017)

Motives and goals, such as the motivated processes by which people wish to view close rather than distant others in a more positive light, also shifts trait inferences. For example,

dominance and trustworthiness are positively correlated when judging close and admired others but negatively correlated when judging unfamiliar and outgroup others ADDIN EN.CITE (Cuddy et al., 2009; Kraft-Todd et al., 2017). Other research suggests that the two-dimensional trustworthiness/dominance trait space does not adequately generalize to trait judgments of close others, perhaps due to more complex representations of familiar personalities (Thornton & Mitchell, 2017). Overall, such findings suggest that face impressions are driven not only by an exquisite sensitivity to specific arrangements of bottom-up facial features but also by a variety of top-down social cognitive factors harbored within perceivers.

4.1. Conceptual influences on trait impressions

As with perceiving social categories and emotions, among such top-down factors, social-conceptual knowledge may have a pronounced impact on face-based trait inferences. Social psychologists have long known a predominant force in non-face-based trait impressions is perceivers′ lay (or implicit) personality theories – how perceivers think others′ personalities function. For instance, a perceiver may conceptually associate the personality traits of kindness and intelligence (e.g., a belief that ′kind people are often intelligent′), then apply these conceptual associations in trait inferences (e.g., perceives kindness in others with features associated with intelligence). Classical research demonstrated how perceivers learn the correlation structure of others′ personalities (Lay & Jackson, 1969), use knowledge of these associations to make trait impressions (Asch, 1946), and noted how the structure of trait conceptual knowledge is reflected in impressions of familiar others (Rosenberg et al., 1968).

Recently, we applied such insights to the study of face impressions (Stolier, Hehman, Keller, et al., 2018). The DI framework′s prediction of a conceptual scaffolding of face impressions (due to domain-general interactions between perceptual processing and conceptual

knowledge) is consistent with other theoretical approaches in this domain. In one sense, it helps to integrate classic implicit personality theory (Schneider, 1973) with overgeneralization theory (Zebrowitz & Montepare, 2008), in that face-based impressions of traits with functionally adaptive features (e.g., anger, from emotion cues) are able to "bleed over" into ostensibly unrelated traits (e.g., trustworthiness) due to the conceptual association between those two traits. For instance, re-adapting an example from classical research (Asch, 1946), if a perceiver believes kind people are intelligent, they may infer kindness from the happiness-resemblance of a face, then intelligence from the kindness impression in part. This account provides a potential explanation for how perceivers easily infer just about any attribute from a face, as nearly any trait concept (e.g., perceived extroversion) can be associated with "lower level" traits more readily inferred from facial features. It also predicts that perceivers will vary in their face impressions to the extent they hold different conceptual knowledge.

As with perceiving social categories and emotions, assessing representational geometry using RSA provides a useful means to comprehensively compare conceptual structure and perceptual structure in perceiving traits from faces as well (Stolier, Hehman, & Freeman, 2018) (see Fig. 5C). In a set of studies, first, we measured the similarity structure of how perceivers thought traits were conceptually associated (e.g., 'how likely is a kind person to be intelligent?'), and how similarly traits were perceived in faces (e.g., how correlated were judgments of face kindness and intelligence). Strikingly, we found that trait concept associations explain roughly 70% of variance in face impressions. Next, we tested whether pairs of traits more conceptually associated were also more correlated in face impressions. We found that participants who believe two traits are more associated also see those traits more similarly in others' faces, and further that they use more similar visual features to judge those traits via a reverse correlation task

(Stolier, Hehman, Keller, et al., 2018) (Fig. 5C).

Social psychology has long noted the conspicuously similar set of dimensions found across contexts of social perception and trait inferences (Fiske et al., 2007). Dimensions alike intention and ability (also known as the Big Two) have appeared in contexts of conceptual knowledge (Lay & Jackson, 1969), face impressions (Oosterhof & Todorov, 2008), impressions of familiar people (Rosenberg et al., 1968), and stereotypes of social groups (Fiske et al., 2002), to name a few. A prominent perspective regarding the reason these similar dimensions emerge across social cognition is that they reflect a universal tendency to track two fundamental and independent dimensions due to their functionally adaptive nature, namely others′ intention (warmth, trustworthiness) and ability (competence, dominance) (Fiske et al., 2007).

But extending the DI framework to trait impressions and the conceptual structuring of those impressions raises a different possibility. The structure of trait impressions across these many contexts (e.g., face impressions, familiar person knowledge, group stereotypes) may be similar not because of their evolutionary relevance but because perceivers apply the same domain-general conceptual knowledge whenever they make an impressions (Stolier, Hehman, & Freeman, invited revision). For instance, a perceiver who believes kind people are likely to be intelligent may infer a kind face, acquaintance, or social group as intelligent alike. This may explain why the common dimensions of trait space are not only found in impressions of unfamiliar others ADDIN EN.CITE (Fiske et al., 2002; Oosterhof & Todorov, 2008; Tamir, Thornton, Contreras, & Mitchell, 2016), but even in impressions of those we know well (Rosenberg et al., 1968). This idea is not incompatible with adaptively significant traits being prioritized and central to our impressions, as those traits could of course help drive the structure of conceptual knowledge. But it suggests that the more proximal mechanism underlying the

structure of social impressions is domain-general conceptual knowledge not a functionally adaptive tracking of specific dimensions. To the extent that conceptual knowledge is different across individuals or cultures, the structure of social impressions should follow suit.

To put this perspective to the test, we conducted a series of studies to measure the relationship between perceiver trait conceptual associations and their trait impressions of targets under several distinct contexts (photographs of unfamiliar faces, names of familiar famous or historical persons, and names of social groups and categories) (Stolier et al., invited revision). First, we asked whether traits more conceptually related are also more correlated in impressions, on average across perceivers. For instance, if ′friendliness′ is more conceptually associated with ′cheerfulness′ than ′adventurousness′, are ′friendly′ impressions more correlated with ′cheerful′ than ′adventurous′ impressions, across impressions of faces, people, and groups? We found that this was indeed the case, replicating and extending the findings described above (Stolier, Hehman, Keller, et al., 2018): conceptual and impression models explained a remarkable proportion of variance in one another across each of these domains. We also found that individual differences in conceptual associations predict individual differences in impressions across contexts, where perceivers who more strongly conceptually associate two traits (e.g., ′friendly′-′intelligent′) infer those traits more similarly. These findings provide correlational evidence of a close tie between conceptual knowledge and impressions that is consistently held across these disparate contexts of social perception − face impressions, familiar person knowledge, and group stereotypes (Stolier et al., invited revision).

Of course, these interpretations were limited by the correlational design of these studies. In Study 4, we set out to manipulate perceiver conceptual knowledge to better test a directional relationship between conceptual knowledge and trait impressions. In the context of face

impressions, we performed a between-subjects experiment manipulating whether participants believed two traits were negatively or positively related (e.g., are ʻfriendlyʼ people more or less likely to be ʻintellectualʼ). Participants first read a faux published scientific research article, which described psychology research finding the two personality traits assigned to a participant to be strongly positively or negatively related in humans. Afterward, participants made impressions of faces along one of the two traits from the article, which was then correlated with the impressions along the second trait as judged by an independent set of raters (participants judged only one trait to reduce the transparency of the research question.) Indeed, as predicted, we found that participants manipulated to believe two traits are negatively related conceptually also perceive those traits less similarly in faces, relative to participants manipulated to believe the two traits are positively related conceptually. Although certainly with limitations (e.g., potential for demand characteristics), these results provide initial evidence for the possibility of a casual impact of perceiver trait conceptual knowledge on how we make impressions of others.

4.2. Summary

From these findings emerges a picture of trait inferences fundamentally shaped by our conceptual knowledge. These observations suggest that conceptual associations scaffold face impressions and the facial features that elicit specific trait inferences. From this process, a trait space emerges in which impressions correlate with one another along the structure of conceptual knowledge on which they are based. Prior perspectives have outlined a trait impression process that is predominately bottom-up and fixed in nature, where perceivers track a key set of traits in targets, such as competence and warmth. While adaptive needs may drive prioritization of certain traits to be inferred, this perspective comes short in addressing recent findings. Extending the DI

framework to trait impressions and the role of domain-general conceptual knowledge in such impressions suggests it may be the structure of that knowledge – which itself is shaped by adaptive needs and prioritization of certain social concepts – to be the more proximal mechanism through which trait impressions occur. In turn, a conceptually structured trait impression process allows impressions to be dynamic in nature and vary to any extent the conceptual knowledge of a perceiver varies. While there will be a central tendency in conceptual trait space across perceivers (Sutherland et al., 2018), perhaps due to perceivers all learning trait space from actual human personality which has a prevalent and largely homogenous structure (Lay & Jackson, 1969; Stolier et al., invited revision), any variance in conceptual knowledge will beget individual differences in impressions. In fact, as discussed earlier, most variance in impressions comes from perceiver characteristics (Hehman et al., 2017), and impressions are important drivers of interpersonal behavior, from workplace decisions (Fruhen, Watkins, & Jones, 2015) to electoral and criminal sentencing outcomes (Todorov et al., 2005; Wilson & Rule, 2015), in which case the consideration of top-down factors in trait impressions may be quite important.

5. Implications and Conclusion

The perspective outlined here is that "initial" social perceptions – as in the perception of a face's social categories, emotions, or traits – are hardly initial at all. Extending the DI framework toward a domain-general account of social perception envisions initial social perceptions as emerging from a single computational system relying on domain-general cognitive properties. In this system, social categories, emotions, and trait perceptions all emerge from the recurrent interactions between visual cues, social cognitive representations, and higher-order cognitive states. As automatic and spontaneous as they may be, they are not mere "read outs" of facial

features in this perspective; instead, they arise out of a rapid negotiation process between bottom -up cues and prior conceptual knowledge and social expectations.

Although top-down perceiver characteristics such as conceptual knowledge are only beginning to be appreciated in face-based trait impressions research ADDIN EN.CITE (Hehman et al., 2019; Hehman et al., 2017; Stolier, Hehman, & Freeman, 2018; Stolier, Hehman, & Freeman, 2018, June 11; Xie et al., 2018), and have only been incorporated into social categorization models in the past few years (Freeman & Johnson, 2016), they have received considerable attention in the affective science literature. Constructionist approaches, such as the Theory of Constructed Emotion and the Conceptual Act Model (Barrett, 2006, 2017), and numerous other researchers have for some time considered the structuring role (and for some, necessary role) that conceptual knowledge and context plays in constructing emotion perception and affective experience ADDIN EN.CITE (Barrett & Kensinger, 2010; Barrett, Mesquita, & Gendron, 2011b; Fugate et al., 2018; Gendron, Lindquist, Barsalou, & Barrett, 2012; Gendron, Mesquita, & Barrett, 2013; Lindquist et al., 2006; Russell, 1997). Our approach is largely consistent with such theoretical perspectives, but aims to integrate emotion perception with perceptions of social categories, traits and other domains of social perception. It also makes a number of new predictions, and if formalized into a model instantiation, would offer a computational means to test specific hypotheses.

One of the biggest advantages of the current perspective is to model social categories (and associated stereotypes), emotions, and traits all as social-cognitive knowledge in a single recurrent system, where these three domains of social perception dynamically interact. Although often studied in relative isolation, it would seem implausible that these processes would live in functionally independent worlds. Indeed, for example and as described earlier, a number of

recent studies have revealed interactions between emotion and gender, race, and age; overgeneralization theory in fact proposes certain traits (e.g., trustworthiness) to be mere overgeneralized forms of specific emotions; and recent studies find trait impressions to shift according to the social categories a target inhabits. Extending the DI framework to encompass these seemingly disparate domains may therefore provide valuable opportunities to better understand the many bridges between them and how they mutually shape one another.

Another novel aspect of this perspective is the DI theory′s focus on real-time dynamics underlying perceptual judgments and the "hidden" impacts that can transpire in those dynamics. Generally, when bottom-up visual information is particularly ambiguous, top-down pressures of social-conceptual knowledge and other factors may have enough strength to bias the representational competition one way or another. In other instances, especially when the bottom-up information is clear-cut, such pressures may not have enough strength to alter responses wholesale. Instead, what often occurs, according to this perspective and supporting evidence, is a stronger partial and parallel activation of a category, emotion, or trait, even though it does not manifest as an explicit and overt perceptual judgment. For instance, as in Fig. 5A, feedback activation from perceivers′ stereotypes may lead perceptions of a smiling, happy Black face to be temporarily biased toward an angry interpretation. Although quickly snuffed out in a few hundred milliseconds, we do know such "hidden" activations can predict downstream social consequences independent of the ultimate perceptual judgment itself (Freeman & Johnson, 2016). Thus, one insight from this perspective is that top-down factors and social-conceptual knowledge may create temporary effects during perception; and although brief, they may in fact have lingering consequences. More generally, whether the top-down shaping of an initial perception manifests only transiently or in the stable percept, we know the powerful effects of

these perceptions on downstream processes and real-world consequences, as described earlier. We also know that the majority of variance for some domains, such as trait impressions, is attributable to perceiver factors. Thus, this perspective could be valuable for understanding how the way we understand our social world shapes initial perceptions of faces in ways that affect downstream outcomes.

It is worth nothing that, while "bottom-up" and "top-down" are helpful terms in thinking about the most proximal influence driving an effect of interest, this perspective assumes perceptions of categories, emotions, and traits arise from complex feedback loops involving many cycles of interaction between visual cues, social cognitive knowledge, and higher-order cognitive states (Freeman & Ambady, 2011). In the original DI model, it was helpful to delineate social-cognitive knowledge in two hierarchical levels, a *stereotype* level and a *category* level; however, together these levels in reality functioned as a single collection of social-cognitive attributes. Certainly in the extended DI model with only a single level for categories, emotions, traits (and stereotypes), the "top-down" effect of social-conceptual knowledge on perception is perhaps better described as a product of recurrence among internal representations.

Of course, the DI model and its extension are only small and early parts of a far larger and more complex person perception system. Its processing is all automatic and associative. Many other social psychological models involve controlled components that use higher-order, resource-dependent processing, and a number of subsequent social cognitive processes including potential control processes are likely triggered to be after an initial perception has crystallized. The DI model, however, focuses on understanding how visual and social cognitive processes rapidly shape initial perceptions; after perception occurs, however, numerous complex social cognitive processes are likely to take place.

An important question for future research is the origins of social-conceptual knowledge. This has been studied most extensively with respect to social categories and their stereotypes, and the process of acquiring stereotype associations is fairly well understood. For emotion-concept knowledge, some recent evidence suggests verbal development explains individual differences in children′s emotion concept knowledge (Nook, Sasse, Lambert, McLaughlin, & Somerville, 2017). But in all three domains, our findings suggest the existence of subtle inter-individual variability in perceivers′ conceptual knowledge about social categories, emotions, and traits, which in turn shapes perceptions. Testing the origins and moderators of such conceptual knowledge will be important for future work. Future research could also consider integrating identity representations into the extended DI framework. Certainly, identity and individuation processes have traditionally been central to person perception models, often contrasted with more categorical forms of processing ADDIN EN.CITE (Brewer, 1988; Fiske & Neuberg, 1990; Kunda & Thagard, 1996), and the current perspective would benefit from integrating face identity perception and individuated knowledge with social categories, emotions, and traits. Finally, the current perspective would need to be formalized into an actual DI model extension, with simulations tested against empirical data. Ultimately, such a model could be additionally advanced by incorporating a fully distributed network with higher neural plausibility, an empirical fitting of connection weights, and learning, to provide a more rigorous theoretical constraints on understanding the interplay of visual and social cognitive processes in perception.

In short, emerging findings suggest that, across various domains of social perception, both a variety of bottom-up facial features and top-down social cognitive processes play a part in driving initial perceptions. We proposed here that the perception of social categories, emotions,

and traits from faces can all be conceived as emerging from an integrated recurrent system. In this system, visual and social cognitive processes are in a close exchange, and initial social perceptions emerge in part out of the structure of social-conceptual knowledge.

Acknowledgements

This work was supported in part by research grants NSF BCS-1654731 and NIH R01-MH112640 to J.B.F.

References

ADDIN EN.REFLIST Abdel-Rahman, R., & Sommer, W. (2008). Seeing what we know and understand: How knowledge shapes perception. *Psychonomic Bulletin & Review, 15*, 1055-1063.

Abdel-Rahman, R., & Sommer, W. (2012). Knowledge scale effects in face recognition: An electrophysiological investigation. *Cognitive, Affective, & Behavioral Neuroscience, 12*, 161-174.

Adams, R. B., Ambady, N., Nakayama, K., & Shimojo, S. (2011). *The Science of Social Vision*. New York: Oxford University Press.

Adolphs, R., Nummenmaa, L., Todorov, A., & Haxby, J. V. (2016). Data-driven approaches in the investigation of social perception. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*:20150367.

Albright, L., Malloy, T. E., Dong, Q., Kenny, D. A., Fang, X., & Winquist, L. (1997). Cross-cultural consensus in personality judgments. *Journal of Personality and Social Psychology, 72*, 558-569.

Allport, G. W. (1924). *Social Psychology*. New York, NY: Houghton Mifflin.

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

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology, 32,* 201-271.

Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology, 69*, 518-529.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological bulletin, 111*(2), 256-274.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology, 41*, 258.

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception, 33*, 717-746.

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience, 7*, 358-366.

Balcetis, E., & Lassiter, D. (2010). *The Social Psychology of Visual Perception*. New York: Psychology Press.

Balota, D. A., & Black, S. (1997). Semantic satiation in healthy young and older adults. *Memory & Cognition, 25*, 190-202.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*, 617-629.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., . . . Rosen, B. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 449-454.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*, 269-278.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54*, 462-479.

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10*, 20-46.

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*, 1-23.

Barrett, L. F., & Kensinger, E. A. (2010). Context is routinely encoded during emotion perception. *Psychological Science, 21*, 595-599.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011a). Context in emotion perception. *Current Directions in Psychological Science, 20*, 286-290.

Barrett, L. F., Mesquita, B., & Gendron, M. (2011b). Context in emotion perception. *Current Directions in Psychological Science, 20*, 286-290.

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology, 92*, 179-190.

Black, S. R. (2001). Semantic satiation and lexical ambiguity resolution. *The American Journal of Psychology, 114*, 493-510.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*, 242-261.

Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology, 70*, 1142-1163.

Blair, I. V., Judd, C. M., & Fallman, J. L. (2004). The Automaticity of Race and Afrocentric Facial Features in Social Judgments. *Journal of Personality and Social Psychology, 87*, 763-778.

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social*

Psychology, 83, 5-25.

Bodenhausen, G. V., & Macrae, C. N. (2006). Putting a face on person perception. *Social Cognition, 24*, 511-515.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer (Eds.), *A Dual-Process Model of Impression Formation: Advances in Social Cognition* (Vol. 1, pp. 1-36). Hillsdale, NJ: Erlbaum.

Brinsmead-Stockham, K., Johnston, L., Miles, L., & Macrae, C. N. (2008). Female sexual orientation and menstrual influences on person perception. *Journal of Experimental Social Psychology, 44*, 729-734.

Brooks, J. A., Chikazoe, J., Sadato, N., & Freeman, J. B. (2019). The neural representation of facial-emotion categories reflects conceptual structure. *Proceedings of the National Academy of Sciences, 116*, 15861−15870.

Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational structure of facial emotion perception. *Nature Human Behaviour, 2*, 581-591.

Brooks, J. A., Shablack, H., Gendron, M., Satpute, A. B., Parrish, M. H., & Lindquist, K. A. (2017). The role of language in the experience and perception of emotion: A neuroimaging meta-analysis. *Social Cognitive and Affective Neuroscience, 12*, 169-183.

Brooks, J. A., Stolier, R. M., & Freeman, J. B. (2018). Stereotypes bias visual prototypes for sex and emotion categories. *Social Cognition, 36*, 481-493.

Brosch, T., Bar-David, E., & Phelps, E. A. (2013). Implicit race bias decreases the similarity of neural representations of black and white faces. *Psychological Science, 24*, 160-166.

Bruner, J. S., & Goodman, C. C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology, 42*, 33-44.

Carpinella, C. M., Chen, J. M., Hamilton, D. L., & Johnson, K. L. (2015). Gendered facial cues influence race categorizations. *Personality and Social Psychology Bulletin, 41,* 405-419.

Carpinella, C. M., Hehman, E., Freeman, J. B., & Johnson, K. L. (2016). The gendered face of partisan politics: Consequences of facial sex typicality for vote choice. *Political Communication, 33*, 21-38.

Carroll, J. M., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology, 70*, 205.

Carroll, N. C., & Young, A. W. (2005). Priming of emotion recognition. *The Quarterly Journal of Experimental Psychology Section A, 58*, 1173-1197.

Caruso, E. M., Mead, N. L., & Balcetis, E. (2009). Political partisanship influences perception of biracial candidates' skin tone. *Proceedings of the National Academy of Sciences, 106*, 20168-20173.

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, 88*, 885-894.

Collins, J. A., & Curby, K. M. (2013). Conceptual knowledge attenuates viewpoint dependency in visual object recognition. *Visual Cognition, 21*, 945-960.

Collins, J. A., & Olson, I. R. (2014). Knowledge is power: How conceptual knowledge transforms visual cognition. *Psychonomic Bulletin & Review, 21*, 843-860.

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.-P., . . . Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology, 48*, 1-33.

Curby, K. M., Hayward, W. G., & Gauthier, I. (2004). Laterality effects in the recognition of depth-rotated novel objects. *Cognitive, Affective, & Behavioral Neuroscience, 4*, 100-111.

Damaraju, E., Huang, Y.-M., Barrett, L. F., & Pessoa, L. (2009). Affective learning enhances activity and functional connectivity in early visual cortex. *Neuropsychologia, 47*, 2480-2487.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*: Harper Perennial.

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion, 14*, 289-311.

de Gelder, B., Vroomen, J., de Jong, S. J., Masthoff, E. D., Trompenaars, F. J., & Hodiamont, P. (2005). Multisensory integration of emotional faces and voices in schizophrenics. *Schizophrenia Research, 72*, 195-203.

Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18.

Dotsch, R., Wigboldus, D. H., Langner, O., & van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science, 19*, 978-980.

Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology, 22*, 22-37.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). The nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33*, 510-540.

Doyle, C. M., & Lindquist, K. A. (2017). Language and emotion: Hypotheses on the constructed nature of emotion perception. In J.M. Fernandez-Dols & J.A. Russell (Eds). *The Science of Facial Expression*. New York: Oxford University Press.

Doyle, C. M., & Lindquist, K. A. (2018). When a word is worth a thousand pictures: Language shapes perceptual memory for emotion. *Journal of Experimental Psychology: General, 147*, 62.

Durán, J. I., & Fernández-Dols, J.-M. (2018). Do Emotions Result in their Predicted Facial Expressions? A Meta-Analysis of Studies on the Link between Expression and Emotion. https://doi.org/10.31234/osf.io/65qp7

Durán, J. I., Reisenzein, R., & Fernández-Dols, J.-M. (2017). Coherence between emotions and facial expressions. In J.M. Fernandez-Dols & J.A. Russell (Eds). *The Science of Facial Expression*. New York: Oxford University Press.

Eberhardt, J. L., Goff, P. A., Purdie-Vaughns, V. J., & Davies, P. G. (2004). Seeing Black: Race, crime, and visual processing. *Journal of Personality and Social Psychology, 87*, 876-893.

Ekman, P. (1972). *Universal and cultural differences in facial expression of emotion.* Paper presented at the Nebraska symposium on motivation.

Ekman, P. (1993). Facial expression of emotion. *American Psychologist, 48*, 384-392.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Review, 3*, 364-370.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*, 124.

Ekman, P., Friesen, W. V., & Ellsworth, P. (2013). *Emotion in the Human Face: Guidelines for Research and an Integration of Findings* (Vol. 11). Elsevier.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science, 164*, 86-88.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience, 2*, 704-716.

Estes, S. G. (1938). Judging personality from expressive behavior. *The Journal of Abnormal and Social Psychology, 33*, 217.

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, 69*, 1013-1027.

Feleky, A. M. (1914). The expression of the emotions. *Psychological Review, 21*, 33.

Firestone, C., & Scholl, B. J. (2015). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, 1-72.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*, 77-83.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*, 878-902.

Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model: Ten years later. In *Dual-Process Theories in Social Psychology* (pp. 231-254). New York, NY: Guilford Press.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum model of impression formation from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology, 23*, 1–74.

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fraley, C. R., Niedenthal, P. M., Marks, M., Brumbaugh, C., & Vicary, A. (2006). Adult attachment and the perception of emotional expressions: Probing the hyperactivating strategies underlying anxious attachment. *Journal of Personality, 74*, 1163-1190.

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science*.

Freeman, J. B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science, 20*, 1183-1188.

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

Freeman, J. B., Ambady, N., Midgley, K. J., & Holcomb, P. J. (2011). The real-time link between person perception and action: Brain potential evidence for dynamic continuity. *Social Neuroscience, 6*, 139-155.

Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General, 137*, 673-690.

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology, 2*, 59.

Freeman, J. B., & Johnson, K. L. (2016). More than meets the eye: Split-second social perception. *Trends in Cognitive Sciences, 20*, 362-374.

Freeman, J. B., Ma, Y., Barth, M., Young, S. G., Han, S., & Ambady, N. (2015). The neural basis of contextual influences on face categorization. *Cerebral Cortex, 25*.

Freeman, J. B., Ma, Y., Han, S., & Ambady, N. (2013). Influences of culture and visual context on real-time social categorization. *Journal of Experimental Social Psychology, 49*, 206-210.

Freeman, J. B., Pauker, K., Apfelbaum, E. P., & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology, 46*, 179-185.

Freeman, J. B., Pauker, K., & Sanchez, D. T. (2016). A perceptual pathway to bias: Interracial exposure reduces abrupt shifts in real-time race perception that predict mixed-race bias. *Psychological Science, 27*, 502-517.

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

Freeman, J. B., Stolier, R. M., Brooks, J. A., & Stillerman, B. A. (2018). The neural representational geometry of social perception. *Current Opinion in Psychology*.

Freeman, J. B., Stolier, R. M., Ingbretsen, Z. A., & Hehman, E. A. (2014). Amygdala responsivity to high-level social information from unseen faces. *The Journal of Neuroscience, 34*, 10573-10581.

Friesen, E., & Ekman, P. (1978). Facial action coding system: a technique for the measurement of facial movement. *Palo Alto, 3*.

Fruhen, L. S., Watkins, C. D., & Jones, B. C. (2015). Perceptions of facial dominance, trustworthiness and attractiveness predict managerial pay awards in experimental tasks. *The Leadership Quarterly, 26*, 1005-1016.

Fugate, J., Gendron, M., Nakashima, S. F., & Barrett, L. F. (2018). Emotion words: Adding face value. *Emotion, 18*, 693.

Galinsky, A. D., Hall, E. V., & Cuddy, A. J. (2013). Gendered races: implications for interracial marriage, leadership selection, and athletic participation. *Psychological Science, 24*, 498-506.

Gaspelin, N., & Luck, S. J. (2018). Top-down″ Does Not Mean ″Voluntary". *Journal of Cognition, 1*.

Gauthier, I., James, T. W., Curby, K. M., & Tarr, M. J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology, 20*, 507-523.

Gendron, M., & Barrett, L. F. (2017). Facing the past: A history of the face in psychological research on emotion perception. In J.M. Fernandez-Dols & J.A. Russell (Eds). *The Science of Facial Expression*. New York: Oxford University Press.

Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion, 12*, 314.

Gendron, M., Mesquita, B., & Barrett, L. F. (2013). Emotion perception: Putting the face in context. In *The Oxford Handbook of Cognitive Psychology* (pp. 539-556). New York, NY: Oxford University Press; US.

Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron, 54*, 677-696.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*, 509-517.

Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*, 491-512.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition, 78*, 27-43.

Hamilton, D. L., Katz, L. B., & Leirer, V. O. (1980). Cognitive representation of personality impressions: Organizational processes in first impression formation. *Journal of Personality and Social Psychology, 39*, 1050-1063.

Hassin, R. R., Aviezer, H., & Bentin, S. (2013). Inherently ambiguous: Facial expressions of emotions, in context. *Emotion Review, 5*, 60-65.

Hehman, E., Carpinella, C. M., Johnson, K. L., Leitner, J. B., & Freeman, J. B. (2014). Early processing of gendered facial cues predicts the electoral success of female politicians. *Social Psychological and Personality Science, 5*, 815-824.

Hehman, E., Leitner, J. B., & Freeman, J. B. (2014). The Face−Time Continuum Lifespan Changes in Facial Width-to-Height Ratio Impact Aging-Associated Perceptions. *Personality and Social Psychology Bulletin*, *40*, 1624-1636.

Hehman, E., Stolier, R. M., Freeman, J. B., Flake, J. K., & Xie, S. Y. (2019). Toward a comprehensive model of face impressions: What we know, what we do not, and paths forward. *Social and Personality Psychology Compass, 13*, e12431.

Hehman, E., Sutherland, C. A., Flake, J. K., & Slepian, M. L. (2017). The Unique Contributions of Perceiver and Target Characteristics in Person Perception. *Journal of Personality and*

Social Psychology, *113*, 513-529.

Hess, U., Adams, R. B., Jr., & Kleck, R. E. (2004). Facial appearance, gender, and emotion expression. *Emotion, 4*, 378-388.

Hess, U., Senécal, S., Kirouac, G., Herrera, P., Philippot, P., & Kleck, R. E. (2000). Emotional expressivity in men and women: Stereotypes and self-perceptions. *Cognition & Emotion, 14,5*.

Huang, L. M., & Sherman, J. W. (2018). Attentional Processes in Social Perception. In *Advances in Experimental Social Psychology* (Vol. 58, pp. 199-241): Elsevier.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science, 14*, 640-643.

Hutchings, P. B., & Haddock, G. (2008). Look Black in anger: The role of implicit prejudice in the categorization and perceived emotional intensity of racially ambiguous faces. *Journal of Experimental Social Psychology, 44*, 1418-1420.

Izard, C. E. (1971). *The Face of Emotion*. East Norwalk, CT, US: Appleton-Century-Crofts.

Izard, C. E. (2011). Forms and functions of emotions: Matters of emotion–cognition interactions. *Emotion Review, 3*, 371-378.

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: how covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology, 102*, 116.

Johnson, K. L., Lick, D. J., & Carpinella, C. M. (2015). Emergent research in social vision: An integrated approach to the determinants and consequences of social categorization. *Social and Personality Psychology Compass, 9*, 15-30.

Johnson, S. L., Eberhardt, J. L., Davies, P. G., & Purdie-Vaughns, V. J. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science, 17*, 383-386.

Jozwik, K. M., Kriegeskorte, N., Storrs, K. R., & Mur, M. (2017). Deep convolutional neural networks outperform feature-based but not categorical models in explaining object similarity judgments. *Frontiers in Psychology, 8*, 1726.

Kaul, C., Ratner, K. G., & Van Bavel, J. J. (2014). Dynamic representations of race: processing goals shape race decoding in the fusiform gyri. *Social Cognitive and Affective Neuroscience, 9*, 326-332.

Kawakami, K., Amodio, D. M., & Hugenberg, K. (2017). Intergroup perception and cognition: An integrative framework for understanding the causes and consequences of social categorization. In *Advances in Experimental Social Psychology* (Vol. 55, pp. 1-80): Elsevier.

Kenny, D. A. (1994). *Interpersonal Perception: A Social Relations Analysis*. Guilford Press.

Kenny, D. A., & La Voie, L. (1984). The social relations model. In *Advances in Experimental Social Psychology* (Vol. 18, pp. 141-182): Elsevier.

Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology, 10*, e1003915.

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical−subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *Neuroimage, 42*, 998-1031.

Kraft-Todd, G. T., Reinero, D. A., Kelley, J. M., Heberlein, A. S., Baer, L., & Riess, H. (2017). Empathic nonverbal behavior increases ratings of both warmth and competence in a medical context. *PloS one, 12*, e0177758.

Krosch, A. R., & Amodio, D. M. (2014). Economic scarcity alters the perception of race. *Proceedings of the National Academy of Sciences, 111*, 9079-9084.

Kuhn, L. K., Wydell, T., Lavan, N., McGettigan, C., & Garrido, L. (2017). Similar representations of emotions across faces and voices. *Emotion, 17*, 912.

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

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience, 27*, 13232-13240.

Lay, C. H., & Jackson, D. N. (1969). Analysis of the generality of trait-inferential relationships. *Journal of Personality and Social Psychology, 12*, 12.

Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General, 135,* 501-512.

Li, W., Piëch, V., & Gilbert, C. D. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nature Neuroscience, 7*, 651-657.

Lindquist, K. A. (2013). Emotions emerge from more basic psychological ingredients: A modern psychological constructionist model. *Emotion Review, 5*, 356-368.

Lindquist, K. A. (2017). The role of language in emotion: existing evidence and future directions. *Current Opinion in Psychology, 17*, 135-139.

Lindquist, K. A., Barrett, L. F., Bliss-Moreau, E., & Russell, J. A. (2006). Language and the perception of emotion. *Emotion, 6*, 125.

Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion, 14*, 375.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behavioral and Brain Sciences, 35*, 121-143.

Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology, 82*, 5-18.

MacLin, O. H., & Malpass, R. S. (2001). Racial categorization of faces: The ambiguous-race face effect. *Psychology, Public Policy and Law, 7*, 98-118.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51*, 93-120.

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

Martin, D., & Macrae, C. N. (2007). A face with a cue: Exploring the inevitability of person categorization. *European Journal of Social Psychology, 37*, 806-816.

Mason, M. F., Cloutier, J., & Macrae, C. N. (2006). On construing others: Category and stereotype activation from facial cues. *Social Cognition, 24*, 540-562.

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the face in context: Cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology, 94*, 365-381.

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

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375-407.

Newell, A. (1980). Physical symbol systems. *Cognitive Science, 4*, 135-183.

Nook, E. C., Lindquist, K. A., & Zaki, J. (2015). A new look at emotion perception: Concepts speed and shape facial emotion recognition. *Emotion, 15*, 569.

Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2017). Increasing verbal knowledge mediates development of multidimensional emotion representations. *Nature Human Behaviour, 1*, 881.

Oh, D., Buck, E. A., & Todorov, A. (2019). Revealing Hidden Gender Biases in Competence Impressions of Faces. *Psychological Science, 30*, 65-79.

Olivola, C. Y., & Todorov, A. (2017). The biasing effects of appearances go beyond physical attractiveness and mating motives. *Behavioral and brain sciences, 40*.

Olson, I. R., McCoy, D., Klobusicky, E., & Ross, L. A. (2012). Social cognition and the anterior temporal lobes: a review and theoretical framework. *Social Cognitive and Affective Neuroscience*, *8*, 123-133.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*, 11087-11092.

2008Pinel, E. C. (1999). Stigma consciousness: The psychological legacy of social stereotypes. *Journal of Personality and Social Psychology, 76*, 114.

Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22*, 341-365.

Pylyshyn, Z. W. (1984). *Computation and Cognition*. Cambridge, MA: MIT Press.

Read, S. J., & Miller, L. C. (1998). On the dynamic construction of meaning: An interactive activation and competition model of social perception. In S. J. Read & L. C. Miller (Eds.), *Connectionist models of social reasoning and social behavior*. Mahwah, N. J.: Erlbaum.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019-1025.

Righart, R., & De Gelder, B. (2008). Rapid influence of emotional scences on encoding of facial expressions: An ERP study. *Social Cognitive and Affective Neuroscience, 3*, 270-278.

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

Rosenberg, S., Nelson, C., & Vivekananthan, P. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9*, 283.

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). *A General Framework for Parallel Distributed Processing*. Cambridge, MA: MIT Press.

Russell, J. A. (1997). Reading emotions from and into faces: Resurrecting a dimensional-contextual perspective. In J. A. Russell & J. M. Fernandez-Dols (Eds.), *The Psychology of Facial Expression*. Cambridge, UK: Cambridge University Press.

Russell, J. A., Bachorowski, J.-A., & Fernández-Dols, J.-M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54*, 329-349.

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion, 9*.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*, 227-256.

Schneider, D. J. (1973). Implicit personality theory: A review. *Psychological Bulletin, 79*, 294.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, 104*, 6424-6429.

Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., . . . Barrett, L. F. (2018). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychological Bulletin, 144*, 343.

Skerry, A. E., & Saxe, R. (2014). A common neural code for perceived and inferred emotion. *Journal of Neuroscience, 34*, 15997-16008.

Skerry, A. E., & Saxe, R. (2015). Neural representations of emotion are organized around abstract event features. *Current Biology, 25*, 1945-1954.

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology, 52*, 689.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin, 105*, 131.

Smith, E. R. (1984). Model of social inference processes. *Psychological Review, 91*(3), 392-413.

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

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science, 16*, 184-189.

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

Sofer, C., Dotsch, R., Oikawa, M., Oikawa, H., Wigboldus, D. H., & Todorov, A. (2017). For Your Local Eyes Only: Culture-Specific Face Typicality Influences Perceptions of Trustworthiness. *Perception*, *46*, 914-928.

Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. *Psychological Review, 96*, 58-83.

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Sciences, 22*, 531-543.

Stolier, R. M., & Freeman, J. B. (2015). The Neuroscience of Social Vision. In J. Cloutier & J. R. Absher (eds.), *Neuroimaging Personality, Social Cognition and Character: Traits and Mental States in the Brain* (pp. 139-157). Elsevier.

Stolier, R. M., & Freeman, J. B. (2016). Neural pattern similarity reveals the inherent intersection of social categories. *Nature Neuroscience, 19*, 795-797.

Stolier, R. M., Hehman, E., & Freeman, J. B. (2018). A dynamic structure of social trait space. *Trends in Cognitive Sciences*, *22*, 197-200.

Stolier, R. M., Hehman, E., & Freeman, J. B. (2018, June 11). Conceptual structure shapes a common trait space across social cognition. doi:https://doi.org/10.31234/osf.io/5na8m

Stolier, R. M., Hehman, E., Keller, M. D., Walker, M., & Freeman, J. B. (2018). The conceptual structure of face impressions. *Proceedings of the National Academy of Sciences*, *115*, 9210-9215.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*, 403-409.

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science, 314*, 1311-1314.

Sutherland, C. A., Liu, X., Zhang, L., Chu, Y., Oldmeadow, J. A., & Young, A. W. (2018). Facial first impressions across culture: Data-driven modeling of Chinese and British perceivers′ unconstrained facial impressions. *Personality and Social Psychology Bulletin, 44*, 521-537.

Sutherland, C. A., Young, A. W., Mootz, C. A., & Oldmeadow, J. A. (2015). Face gender and stereotypicality influence facial trait evaluation: Counter-stereotypical female faces are negatively evaluated. *British Journal of Psychology, 106*, 186-208.

Tajfel, H. (1981). *Human groups and social categories: Studies in Social Psychology*: CUP Archive.

Tamir, D. I., Thornton, M. A., Contreras, J. M., & Mitchell, J. P. (2016). Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. *Proceedings of the National Academy of Sciences, 113*, 194-199.

Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience, 3*, 764-769.

Thornton, M. A., & Mitchell, J. P. (2017). Theories of person perception predict patterns of neural activity during mentalizing. *Cerebral Cortex*, *28*, 3505-3520.

Todorov, A., Dotsch, R., Wigboldus, D. H. J., & Said, C. P. (2011). Data-driven methods for modeling social perception. *Social and Personality Psychology Compass, 5*, 775-791.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*, 1623-1626.

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66*.

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: a review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review, 3*, 397-405.

Tskhay, K. O., & Rule, N. O. (2013). Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Personality and Social Psychology Review, 17*, 72-86.

Uleman, J. S., & Kressel, L. M. (2013). A brief history of theory and research on impression formation. *Oxford Handbook of Social Cognition*, 53-73.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.), *Advances in*

Social Psychology (Vol. 28, pp. 211-279). San Diego: Academic Press.

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2008). The neural substrates of in-group bias a functional magnetic resonance imaging investigation. *Psychological Science, 19*, 1131-1139.

Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion, 7*, 487-494.

Vinson, D. W., Abney, D. H., Amso, D., Chemero, A., Cutting, J. E., Dale, R., . . . Gallagher, S. (2016). Perception, as you make it. *Behavioral and Brain Sciences, 39*.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*, 592-598.

Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., & Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia, 49*, 1105-1127.

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science, 26*, 1325-1331.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology, 47*, 237.

Wyer Jr, R. S., & Carlston, D. E. (2018). *Social Cognition, Inference, and Attribution*. Psychology Press.

Xiao, Y. J., Coppin, G., & Van Bavel, J. J. (2016a). Clarifying the role of perception in intergroup relations: Origins of bias, components of perception, and practical implications. *Psychological Inquiry, 27*, 358-366.

Xiao, Y. J., Coppin, G., & Van Bavel, J. J. (2016b). Perceiving the world through group-colored glasses: A Perceptual Model of Intergroup Relations. *Psychological Inquiry, 27*, 255-274.

Xie, S. Y., Flake, J. K., & Hehman, E. (2018). Perceiver and target characteristics contribute to impression formation differently across race and gender. *Journal of Personality and Social Psychology*.

Zebrowitz, L. A. (2006). Finally faces find favor. *Social Cognition, 24*, 657-701.

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

2008Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2*, 1497-1517.