Valence Biases and Emergence in the Stereotype Content of Intersecting Social Categories

Gandalf Nicolas[1] & Susan T. Fiske[2]

[1] Rutgers University – New Brunswick, NJ, United States

[2] Princeton University, NJ, United States

In press at the *Journal of Experimental Psychology: General*

Abstract

People belong to multiple social groups simultaneously. However, much remains to be learned about the rich semantic perceptions of multiply-categorized targets. Two pretests and three main studies (N = 1116) compare perceptions of single social categories to perceptions of two intersecting social categories. Unlike previous research focusing on specific social categories (e.g., race & age), our studies involve intersections from a large sample of salient societal groups. Study 1 provides evidence for biased information integration (vs. averaging), such that ratings of intersecting categories were more similar to the constituent with more negative and more extreme (either very positive or very negative) stereotypes. Study 2 indicates that negativity and extremity also bias spontaneous perceptions of intersectional targets, including dimensions beyond Warmth and Competence. Study 3 shows that the prevalence of emergent properties (i.e., traits attributed to intersecting categories but not the constituents) is greater for novel targets and targets with incongruent constituent stereotypes (e.g., one constituent is stereotyped as high Status and the other as low Status). Finally, Study 3 suggests that emergent (vs. present in constituents) perceptions are more negative and tend to be more about Morality and idiosyncratic content and less about Competence or Sociability. Our findings advance understanding about perceptions of multiply-categorized targets, information integration, and the connection between theories of process (e.g., individuation) and content.

Public significance statement

This paper shows how spontaneous measures reveal nuanced consequences of multiple categorization, with intersectional stereotypes showing unique content and properties.

Keywords: multiple categorization; intersectionality; text analysis; emergence; negativity bias

Valence Biases and Emergence in the Stereotype Content of Intersecting Social Categories

Categorization and stereotyping help perceivers make sense of a complex social world. However, much of this complexity is lost in dominant psychological models: Research in the field heavily relies on laboratory forced-choice numerical ratings of single social groups along researcher-predetermined categories. In reality, perceivers spontaneously use rich and diverse semantic dimensions to understand people who belong to multiple and often ambiguous social groups (Nicolas, Bai, & Fiske, 2022). This is ever more the case in an increasingly globalized and multicultural world (e.g., Phinney & Alipuria, 2006), where perceivers need to organize their mental societal structures around a larger number of identities. Here we examine perceptions of targets for whom membership in two group categories are activated, shedding light into how people perceive others as a function of intersecting social identities.

To examine perceptions of multiply-categorized targets, this paper explores whether (a) perceived characteristics of the constituent groups predict attributions for the intersection of the constituents, (b) beyond constituent stereotypes, emergent content readily appears in perceptions of the intersections, and (c) the content and associated properties of intersections' stereotypes yield systematic principles. Using a large sample of salient representative U.S. groups, the current studies of multiply-categorized targets combine models of information integration and emergence to test the role of incongruence along multiple dimensions of content, using spontaneous responses and natural language processing to provide replicable coding of emergence and content prevalence.

This report draws on and informs multiple theoretical perspectives. After a brief review of social categorization and associated content, the next section takes up perceptions of intersecting categories, including valence biases and emergence, and the consequences of

multiple categorization. Two pretests and three studies (plus replications in the Supplement) suggest meaningful principles for further research.

Single Social Categorization and Stereotype Content

Social categorization and stereotyping are fundamental social-cognitive processes (Fiske & Neuberg, 1990). Quickly upon encountering others, perceivers access information related to gender, race, and age categories (e.g., Ito & Urland, 2003), and subsequently activate associated stereotypes (Bodenhausen et al., 2012; Macrae & Bodenhausen, 2000). These functions can be adaptive, allowing perceivers to organize the social world and facilitate decision-making (Abele et al., 2021), but also bias intergroup relations and often result in discrimination (e.g., Dovidio et al., 2017).

Here, stereotypes are defined as beliefs that perceivers have about social groups (Katz & Braly, 1933; for a review, see Fiske et al., 2021). To understand the content of these beliefs, the Stereotype Content Model (SCM; Fiske et al., 2002) has documented the centrality of Warmth and Competence as stereotype dimensions. Warmth refers to perceptions about whether the target is an ally or a foe (trustworthy and friendly or not), while Competence refers to perceptions about whether the target can follow through on their intentions (capable and confident or not). These dimensions have other names (e.g., Agency and Communion; Abele et al., 2016). Several models agree on these two dimensions (for a synthesis, see Abele et al., 2021; Koch et al., 2021). More recent models (e.g., ABC; Koch et al., 2016; Koch et al., 2020) and theoretical refinements (e.g., Abele et al., 2016; Leach et al., 2007) have argued for alternative central dimensions (e.g., Progressive-Traditional Beliefs) or subdivision into facets of Warmth (into Sociability and Morality) and Competence (into Ability and Assertiveness[1]).

[1] The Assertiveness dimension includes related concepts of dominance and aggressiveness (see Rosette et al., 2016; Nicolas, Bai, et al., 2022).

Demonstrably, social group stereotypes vary relatively independently along the two central dimensions, such that some groups are stereotyped as being high Warmth and high Competence (e.g., people who are middle class), low Warmth and low Competence (e.g., people who are homeless), or as having ambivalent stereotypes, such as high Warmth but low Competence (e.g., people with disabilities), or low Warmth but high Competence (e.g., people who are wealthy). The different social groups thus are associated with distinct quadrants in the Warmth by Competence space. Furthermore, perceivers feel differentiated emotions toward groups in each of these quadrants: Admiration for high-high (Warmth, Competence), Envy for low-high, Pity for high-low, and Contempt for low-low (Fiske et al., 2002). Social groups' standing in the Warmth by Competence stereotype space further links to a myriad of consequences, including intergroup behavioral intentions (e.g., Cuddy et al., 2007), well-being (e.g., Townsend et al., 2009), biases in policing behavior (e.g., Correll et al., 2007), and impression management (Dupree & Fiske, 2019), among many others.

More recently, using open-ended perceptions of social groups coded through text analyses (an approach similar to the one used here), the Spontaneous Stereotype Content Model (SSCM; Nicolas, Bai, et al., 2022) supported not only the previously mentioned dimensions, but also formalized a larger variety of content that perceivers use to make sense of societal groups (from Emotion and Appearance to Deviance and Health). Moreover, the SSCM explicitly distinguishes between dimensions' valence (i.e., whether a group is evaluated positively or negatively on a dimension) and representativeness (i.e., whether the dimension is spontaneously prevalent in perceptions of the group). For example, although Americans may perceive doctors and nurses as similarly highly warm and competent, spontaneous stereotypes about doctors and nurses focus on either Competence or Warmth, respectively. Integrating representativeness and

direction improves predictions of social group attitudes (Nicolas, Bai, et al., 2022). The current paper makes use of these novel methods and the SSCM approach to provide a more nuanced view than previous research of the complex content that may accompany social perceptions.

Multiple Social Categorization

Perceiving social targets on the basis of a single activated social category is relatively well understood (Allport, 1954; Crisp & Hewstone, 2007). However, outside of the lab, perceivers belonging to multiple groups need to make sense of targets who themselves belong to multiple groups (e.g., Crisp et al., 2001), with differing associated (and often correlated, e.g., Johnson et al., 2012) stereotypes.

Next, we review relevant models of information integration that may explain patterns of stereotypes about intersecting categories. Research on these models has often focused on the stereotypes of intersections of particularly salient social categories, such as those based on race, age, and gender, oftentimes with contradictory results as to which model best explains the data, depending on the context of application (see Petsko & Bodenhausen, 2020; Petsko et al., 2022).

To preview, these models include simple averaging of constituent features and weighted averages where one constituent (e.g., the more negative or extreme) has more influence on perceptions of the intersection. In Studies 1 and 2 we test for the presence of negativity and extremity biases as evidence for a weighted (vs. simple) average model of information integration. Beyond averaging patterns, models of emergence and psychological intersectionality suggest that intersectional targets will exhibit distinct perceived attributes from the constituents (i.e., emergent attributes). Study 3 will examine emergence patterns, including whether incongruence in the stereotypes of the constituents predicts emergence, and how does the content and valence of emergent stereotypes differ from stereotypes shared with the constituents.

Simple Average Models

In the simplest model of information integration, perceptions of multiply-categorized targets can be described in terms of averaging of the perceptions of each constituent group in isolation. For example, perceptions of a target who belongs to two positively-evaluated groups will also be positive, as the average of the constituents' evaluations is positive. On the other hand, a target belonging to both a positive and a negative group will be evaluated as neutral, as the two constituents' scores average out to neutral (see Anderson, 1965).

Negativity and Extremity as Weights

Although simple average models may fit many patterns of multiple categorization perception, there is substantial evidence of robust biases in the information integration of valenced information. In particular, negativity and extremity biases have received considerable attention given their prevalence (e.g., Fiske, 1980). Negativity bias refers to the more negative traits having a greater impact on the evaluations of a target with multiple traits. Note that the opposite, a positivity bias is possible (e.g., Mende-Siedlecki et al., 2013), but less-often documented (we will review an important exception in the next paragraphs). Extremity bias refers to the more extreme (vs. more average or neutral) traits, either positive or negative, having a greater impact on the evaluations of a target with multiple traits. In other words, information integration of multiple social evaluations may be described as a weighted average of the multiple evaluations, where more negative or extreme evaluations receive more weight.

Multiple theories explain why negativity and extremity biases arise (see Skowronski & Carlston, 1989). For example, negative information may be more informative because positive traits and behaviors are often more similar to each other than negative ones (e.g., Alves et al., 2017), with distinctiveness of information relating to integration biases. Traits and behaviors that

occur less frequently are more informative (e.g., Fiske, 1980), which includes both negative and extreme traits, under the assumption that moderately positive traits and behaviors are more common (see also Gawronski & Brannon, 2019).

Of particular relevance here is how biases operate differently depending on the content of traits and behaviors (Reeder & Brewer, 1979; Skowronski & Carlston, 1987). Previous research has found support for the previously stated negativity bias for Warmth-related behaviors and traits. However, some research suggests that Competence-related information integration often exhibits a positivity bias, such that perceivers give more weight to positive Competence behaviors and traits in evaluating a target's overall Competence. Existing models explain this evidence by pointing to cultural expectations that moral people rarely behave in a blatantly immoral way, but that immoral people may frequently behave in moral ways without much expected contradiction (Reeder & Spores, 1983). On the other hand, the cultural expectation is that competent people sometimes do not-so-smart things, but incompetent people are less capable of smart behaviors (see also Rusconi et al., 2020; Wojciszke et al., 1993). This leads to individual pieces of information (e.g., a behavior at a point in time) to be weighted more heavily in person perception if they are about negative (vs. positive) Warmth, or if they are about positive (vs. negative) Competence.

Although much of the research on valence biases has been using combinations of traits or behaviors, we believe that these evaluation biases will also arise in perceptions of intersections of social groups. Specifically, we expect to find extremity biases for both Warmth and Competence (Hypothesis 1a), such that intersectional targets will be seen as more similar on Warmth and Competence to the constituent with the more extreme scores on the corresponding dimension. Similarly, we expect to find that intersectional stereotypes will show a negativity bias

for Warmth, such that intersectional target's Warmth will be more similar to the more negative constituent Warmth score (Hypothesis 1b). On the other hand, we expect either no significant valence asymmetry or a positivity bias for Competence (Hypothesis 1c). Furthermore, these patterns should extend beyond specific dimensions, such that the overall stereotype content associated with intersectional targets will more closely resemble one constituent's stereotypes over the other (Hypothesis 2a). Again, we expect resemblance at this abstract, across-dimension representations to be influenced by the more extreme (Hypothesis 2b) and negative (Hypothesis 2c) constituent.

However, we note some caveats when drawing connections to the previous literature. First, combinations of traits and behaviors provide more "bottom-up" information, while intersections of societal categories may provide more "top-down" information. Presumably, participants have pre-existing impressions and stereotypes about many of the group intersections, thus resulting in memory retrieval (c.f., Norris et al., 2019) rather than information integration. Given this, and the fact that some intersectional targets will be more novel, requiring impression formation, we expect negativity and extremity biases to arise in perceptions of multiply-categorizable targets (particularly for general or Warmth-related stereotypes), but the current studies are not able to robustly disambiguate these mechanisms. We additionally note that stereotypes of social groups, compared to traits and behaviors preselected by researchers for impression formation tasks, may be more ambivalent and show higher variability across participants (e.g., participants may think of different subgroups, Clausell & Fiske, 2005, or perceive the groups as more or less homogeneous, Quattrone & Jones, 1980). Thus, examining valence biases in the context of stereotypes (vs. traits or behaviors) is necessary for an understanding of their manifestation in this particular, societally relevant context.

Emergence

Beyond weighted average models, an alternative arises: the content of the stereotypes at the intersection of multiple social groups can differ from the contents of the constituents or their (weighted) average. That is, intersections may have emergent properties. A classic study of social-group intersections (Kunda et al., 1990) investigated whether targets who belong to two different groups would be associated with stereotypes that are not common stereotypes of either of the constituent groups. For example, participants encountered a target who was a Harvard-educated carpenter and described the target as possessing qualities such as *affluent* (part of the Harvard-educated stereotype) and *rugged* (part of the carpenter stereotype). However, participants also provided emergent trait attributions (not inherited from the constituents), such as a Harvard-educated carpenter being *nonmaterialistic* and *nonconformist*. Following these findings, we expect that emergent properties will also frequently arise when using a much larger sample of social category intersections (Hypothesis 3a).

Previous research suggests that emergence arises most often for category intersections that are surprising or incongruent (see Crisp & Hewstone, 2007). The main mechanism theorized behind emergence is complex reasoning (including causal reasoning) to resolve inconsistent cognitions (see Hastie et al., 1990). For example, when examining surprising intersectional targets (e.g., “a blind marathon runner”), participants engaged in more causal reasoning describing their expectations about the targets than for less surprising intersectional targets (e.g., “a feminist who is a bank teller”; Kunda et al., 1990). This surprisingness is partially due to incongruent stereotypes (see later for additional factors). Here, we also expect to find that, in general, stereotype incongruence (Hypothesis 3b) and novelty (Hypothesis 3c) predict higher emergence.

Most studies in this literature examine category intersections that are incongruent in terms of Competence-related dimensions (Ability, Assertiveness, Status), clearly distinct from Warmth-related dimensions (although some are incongruent on alternative dimensions, such as Beliefs, e.g., "communist ex-Marine"). Besides the Harvard-educated carpenter, other Competence-related-incongruent intersectional targets examined include "Oxford-educated bricklayer" (Hutter et al., 2009) and "blind lawyer" (Kunda et al., 1990). Whether the incongruence-emergence link exists for Warmth and other dimensions is less clear. For example, Competence-related dimensions are more structural and consensual, whereas Warmth is more personal and idiosyncratic (Koch et al., 2020; Nicolas, Fiske, et al., 2022). Perceivers may struggle to describe a Competence/Status incongruent "Harvard-educated carpenter" because the Status of these constituents is more structural and static (i.e., the constituent's Status is stable across many subgroups, contexts, or exemplars). On the other hand, it is possible that the Warmth of each constituent of the Warmth-incongruent "Middle class & Middle Eastern" person is more easily reinterpreted in light of the other constituent (because Warmth stereotypes are more variable across subgroups, contexts, or exemplars), resulting in less emergence. Thus, we hypothesize that higher incongruence in terms of structural-consensual Competence information will result in more causal reasoning and emergent attributions (Hypothesis 3d), but advance no hypothesis about the effect of incongruence on the more personal and potentially more malleable Warmth dimension.

As with the caveat previously raised about averaging models, pre-existing stereotypes about category intersections play an important role in theories of whether emergence occurs via complex processes such as causal reasoning. Specifically, for category intersections that are not novel to the perceiver, they may draw from exemplars or prototypes in memory to make trait

attributions. Primarily novel groups for which stereotypes do not exist should elicit emergent attributes online *through* causal reasoning. However, emergent properties, which we define here as perceptions representative of category intersections but not the constituents, may also be found in more familiar category intersections. We review some of this literature in the next section.

Intersectionality Perspectives

Drawing from Black feminist scholarship outside psychology (e.g., Crenshaw, 1989; see also Rosette et al., 2018), a number of psychological theories have touched on how people perceive targets at the intersection of identities in ways that differ from their constituents and alternative intersections. The intersectional invisibility hypothesis (e.g., Purdie-Vaughns & Eibach, 2008) suggests that Black women are more cognitively invisible to perceivers because they are not prototypical of the Black category (Black men are) nor the Women category (White women are). This translates, for example, to Black women's faces and speech (e.g., in a "who said what" task; Taylor et al., 1978) being remembered less often (vs. Black men; Sesko & Biernat, 2010; see also Biernat & Sesko, 2013). Intersectional invisibility occurs for other non-prototypical intersections of social categories (e.g., Asian men; Schug et al., 2015, 2017). There are multiple reasons that a category may become prototypical of another, including frequent co-occurrence of the categories in media and phenotypic similarity. Once categories are associated through prototypicality, exposure to an intersectional target may automatically spread activation to these associated categories and influence perceptions of the target (see Hall et al., 2019). For example, because of the prototypical association of the "Asian" and "woman" categories, perceptions of an Asian man may also be influenced by the "woman" category to the extent that the Asian category activates associated concepts in memory.

Beyond online information integration during evaluation of intersectional targets, intersectionality accounts suggest a different route for emergent attributes in group conjunctions when stereotypes are preexisting: The reason for emergent properties is that the constituents are, implicitly, already incorporating intersectional information about a different group than that included in the intersection. Thus, for example, Black women would be stereotyped differently from the constituents because stereotypes about "Black" as a single-category are mostly about Black men, and stereotypes about "Woman" as a single-category are mostly about White women.

Along these lines, some studies have explored how intersectional stereotypes for specific salient social groups differ from their constituents, using free response methods. For example, representative stereotypes of Black-White mixed-race individuals in the U.S. include traits such as "beautiful," "confused," and "not belonging," which are not as representative of the stereotypes of either Black or White people (Nicolas et al., 2019; Skinner et al., 2020)[2]. Other studies have looked at gender-by-ethnicity intersections and found evidence of emergent attributes, as well as patterns of bias and prototypicality (e.g., stereotypes of men and women having the White category as prototypical; Ghavami & Peplau, 2012). However, none of these studies have looked at generalized patterns across a large sample of societal groups (often by design, given a particular interest in identities most relevant to structural power dynamics). As a complement, here, we incorporate salient social categories distributed across status and valence of perception, for relevant comparison and analysis of information integration and

[2] Although people may perceive mixed-race individuals as belonging to a single "Multiracial" or alternative category (Nicolas et al., 2019) rather than as belonging to two racial groups simultaneously.

incongruence[3]. We expect to find significant diversity in the content of intersectional targets' stereotypes, as in single-group content models (e.g., Nicolas, Bai, & Fiske, 2022). However, we expect to find differences between constituent and emergent stereotypes. For example, emergent stereotypes may be more idiosyncratic or about Deviance, as the targets may be seen as non-prototypical (Hypothesis 3e; c.f., previous studies on specific intersections, e.g., Ghavami & Peplau, 2012; see also Kunda & Oleson, 1995). Emergent Warmth and Competence stereotypes may also be more negative, as perceivers need to resolve cognitive conflict from incongruent stereotypes (Hypothesis 3f; c.f., Proulx & Inzlicht, 2012; multiracial stereotypes, Skinner et al., 2020). However, in line with research on emergent attributions for targets categorized as Multiracial (e.g., "beautiful" vs. "confused" stereotypes), we expect to find variance in valence across different dimensions (Hypothesis 3g).

Consequences of Multiple Categorization

A better understanding of multiple categorization holds promise to not only improve the applicability, ecological validity, and generalizability of stereotype content research, but also to uncover previously unknown patterns. For example, previous research has found biases in the selection of Black (vs. Asian) candidates to positions more strongly associated with stereotypically masculine (vs. feminine) traits (Galinsky et al., 2013), or that perceptions of Black (vs. White) men as less fit for leadership positions are reversed when information about the targets being gay (vs. heterosexual) is included (Wilson et al., 2017).

The case of novel category intersections with incongruent stereotypes has also been linked to higher individuation (i.e., less reliance on categories and stereotypes; Fiske & Neuberg,

[3] As in other papers in this review, we do not use an intersectionality theory framework, which more fully considers social systems, power differentials, and activism; instead, we draw from intersectionality, multiple categorization, and other literatures and focus on the narrower concept of perceptions of intersecting identities.

1990) through the complex reasoning that leads to emergent attributes (Hutter & Wood, 2015). Individuation can allow perceivers to go beyond often-faulty stereotypes. However, this may not diminish negative evaluations across all dimensions and groups. For example, given that the emergence route to individuation is often elicited by incongruent groups, and incongruence is often aversive (e.g., see Proulx & Inzlicht, 2012), the individuated traits assigned to the targets may possibly incorporate more negativity than perhaps would be expected from the constituent stereotypes alone (c.f., Rudman & Fairchild, 2004).

Furthermore, the related literatures of subgrouping and subtyping (e.g., Maurer et al., 1995; Richards & Hewstone, 2001) suggest that multiple categorization may also have implications for stereotype change of the constituent categories. Subgrouping refers to the existence or formation of within-category groups, which are perceived as relatively prototypical of the superordinate category. On the other hand, subtyping refers to the separation of non-prototypical group exemplars from the superordinate category (Maurer et al., 1995). Subtyping involves causal reasoning to justify excluding the non-prototypical individual or subgroup from the superordinate category, and often results in stereotype maintenance because disconfirming information is not integrated with the superordinate stereotypes (Richards & Hewstone, 2001).

Thus, a perceiver encountering an intersectional target with incongruent stereotypes may subtype the target from both constituents, thus preventing change in the constituent stereotypes that would occur from incorporating disconfirming information. If the intersectional target is typical enough of one or both constituents, it may instead make salient to the perceiver that the constituents are not monoliths and have within-category variation (e.g., subgroups). The latter may be a more positive form of more-individuated thinking. See the Supplement for discussion of theoretical overlap with additional literatures and models.

Examining general patterns of bias and emergence can thus help us gain theoretical insights into the content, structure, and processes that underlie multiple categorization, as well as practical considerations to be potentially applied in interventions and policy.

Current Studies

The current paper presents the results of two pretest studies and three main studies (plus replications in the Supplement). The pretest studies report baseline open-ended and rating-scale stereotypes of salient societal groups. Studies 1 and 2 use two-group combinations of targets from Pretest 1 as stimuli. Pretest 2 expanded the list of groups from the first Pretest to increase generalizability for Study 3.

The main studies then explore and test several hypotheses related to bias, emergence, and stereotype content using random combinations of salient, representative social groups in the United States. Using scales (Studies 1 & 2) and spontaneous open-ended measures coded through word embeddings (Studies 2 & 3) and dictionaries (Study 3), we explore whether bias and emergent properties can be predicted from extremity, negativity/positivity asymmetries, and incongruence along multiple dimensions (including Warmth and Competence). In addition, we explore the content of emergent stereotypes, and compare it to the content of constituent stereotypes, in order to explore whether dimensions are systematically weighted differently depending on their emergence (Study 3).

In Study 1, we expect to find evidence of valence asymmetries and extremity biases. Specifically, we expect intersectional targets to be perceived as more similar to the more extreme constituent group, in terms of both Warmth and Competence (Hypothesis 1a). Additionally, we expect intersections' Warmth to be perceived as more similar to that of the constituent with the more negative Warmth (i.e., a negativity bias; Hypothesis 1b). Additionally, we expect either no

valence asymmetry or a positivity bias for Competence (such that the intersection's Competence is more similar to the more positive Competence score between the constituents; Hypothesis 1c).

In Study 2, we introduce a quantitative measure of the holistic semantic stereotypes of constituent and intersectional targets, allowing us to move beyond specific dimensions to explore biases across all the latent dimensions of stereotype content provided in open-ended responses. We expect to find that incongruence in the stereotype content of the constituents predicts biases in the holistic semantic stereotypes (across all latent dimensions) of the intersectional target (Hypothesis 2a). Specifically, the more incongruent the stereotypes of the constituents, the more the perceptions about the intersection will align with one constituent's stereotypes vs. the other (vs. equal similarity to both). Moreover, we expect that intersectional holistic stereotypes will be more similar to the holistic stereotypes of the constituent with the more extreme (Hypothesis 2b) and negative (2c) scores on general valence.

Finally, in Study 3, we expect that intersectional perceptions will have significant prevalence of emergent attributes (Hypothesis 3a)[4]. We hypothesize that emergence will be more prevalent when the holistic constituent stereotypes are more incongruent (Hypothesis 3b) and when the intersectional target is less familiar (Hypothesis 3c). Additionally, breaking down by dimension, we expect emergence to be higher when the constituents' Competence is more incongruent (Hypothesis 3d), but have no specific hypothesis about Warmth incongruence. In more data-driven analyses we expect that the prevalence and valence of the dimensions will differ across emergent and constituent stereotypes: emergent stereotypes will be more about alternative dimensions, beyond the big two of Warmth and Competence (in particular Deviance and more idiosyncratic content; Hypothesis 3e). We also expect that emergent stereotypes will

[4] Based on exploratory analyses of Study 2, we expected emergent properties to be approximately 38% to 58%, but our formal hypothesis was much more conservative, testing against 0%.

be, on average, more negatively valenced (Hypothesis 3f), although there should be differences in valence across dimensions (Hypothesis 3g).

Finally, in an exploratory analysis of Study 3 we test whether causal reasoning is a mediator between constituent stereotype incongruence and emergent properties. Previous research has suggested causal reasoning operates in the attribution of emergent properties for novel incongruent groups (Hutter et al., 2015). In this case, many of our category intersections are not completely novel. However, if causal reasoning is nonetheless a mediator, it should indicate some degree of individuation operating to explain stereotype incongruence, even when preexisting stereotypes may exist for the specific intersection. This compares, for example, to causal reasoning engagement by individuals in an effort to maintain coherent view of their identities, which are of course not novel to them (Gardner & Garr-Schultz, 2017). Indeed, we expect this to be the case, with causal reasoning, evidenced through coding of participants' stories about the targets, relating to incongruence and emergence as a partial mediator.

These findings will provide an initial view into the content and associated processes of stereotypes associated with social category intersections, under a generalist approach (i.e., attempting to find systematic patterns across many societal groups). This approach holds promise to advance our understanding of the degree to which generalizable patterns are found (or not) in perceptions of multiply-categorized targets. Moreover, our findings could help illuminate new avenues of integration for models of content (e.g., the Stereotype Content Models; Fiske at al., 2002; Nicolas, Bai, et al., 2022) and models of process (e.g., the continuum model of impression formation; Fiske & Neuberg, 1990), while highlighting the importance of frameworks that move the field toward an understanding of social targets along their multiple identities.

A series of variables are introduced throughout the studies, and we include a glossary of a subset in Table 1. Throughout, we attempt to replicate the findings from each previous study to establish robustness. We also make all data and code available in the online repository (Nicolas & Fiske, 2022): https://osf.io/8kad4/?view_only=669b91ffe24a43a9b8fefb9bb850e2e2. Studies were not preregistered.

Table 1.

Definitions for select variables, including the study where they are first introduced.

Study	Glossary of select variables	Definition
Study 1	Intersections' value bias	The degree to which the intersection's score on a variable (e.g., Warmth) is more similar to the score of its first (vs. second) constituent.
Study 1	Constituents' average value	The average score on a variable for the two constituents associated with an intersectional target.
Study 1	Constituents' relative value	The degree to which the first constituent associated with an intersectional target is seen more positively than the second constituent on a variable.
Study 1	Constituents' relative extremity	The degree to which the score on a variable for the first constituent associated with an intersectional target is more extreme (relative to the average score across all groups) than the second constituent's score.
Study 2	Holistic/Abstract stereotypes	A summary representation of the semantics of all the attributes associated with a target, encoded in a numerical space through word embeddings. Correlating the holistic stereotypes of two groups provides a measure of their semantic similarity.
Study 2	Intersections' spontaneous bias	The degree to which the intersection's holistic stereotypes are more semantically similar to the holistic stereotypes of its first (vs. second) constituent.
Study 2	Constituents' semantic incongruence	A measure of semantic dissimilarity between the constituents associated with an intersectional target, at the level of holistic stereotypes.
Study 3	Emergent properties	Attributions associated with an intersectional target that are not prevalent in the stereotypes of its constituents.

Pretest 1

In the first pretest we obtain scale ratings and spontaneous (open-ended) stereotypes about single groups. Unlike previous studies on similar topics, here we take a larger sample of societally salient social groups, with the purpose of arriving at systematic patterns shared across a variety of salient societal groups (see Fiske et al., 2002). The pretest results subsequently allow computing relevant variables and serve as baseline comparison, in Studies 1 and 2.

Methods and Results

Participants were 204 Amazon Mechanical Turk workers. The majority were women (55%; 44% men) and White (84%; 6% Black; 3% Asian, 3% Multiracial), with a mean age of 37.9[5]. For both pretests and main tests throughout the paper we powered conservatively, assuming small-medium effect sizes (r = .2) and a simple correlational analysis, resulting in > 80% power for samples of ~200 participants. In reality, our models include repeated measures, which would make our studies higher powered than suggested by this power analysis. Power analyses were conducted using G*Power (Faul et al., 2009).

The main stimuli consisted of 20 group labels that are salient representative U.S societal groups, selected (based on the literature, e.g., Fiske et al., 2002) to be distributed across the stereotype content model space (see Table 2). Labels chosen were based on the participants' more common responses for congruence with the literature and to avoid researcher input affecting the results (original question was: "what various types of people do you think today's society categorizes into groups?").

[5] Across all studies, we asked participants to report their gender identity choosing from "man", "woman" or "I identify as", followed by an open-ended box; and to choose one or more of the census options for racial/ethnic identities ("White", "Black or African American", "American Indian or Alaska Native", "Asian", "Native Hawaiian or Other Pacific Islander", and "Hispanic").

Table 2

List of groups used in: Pretest 1, Study 1, and Study 2.

Quadrant (a priori)	Group ("A person who is…")	Warmth	Competence
Admiration	American	3.69	3.91
Admiration	Christian	3.96	3.54
Admiration	Educated	3.55	4.41
Admiration	Middle-class	3.82	3.84
Admiration	White	3.72	3.87
Contempt	a Crossdresser	3.24	2.86
Contempt	a Drug addict	1.75	1.57
Contempt	a Welfare recipient	2.58	1.77
Contempt	an Undocumented immigrant	2.76	2.84
Contempt	Homeless	2.47	1.70
Envy	a Professional	3.46	4.05
Envy	an Entrepreneur	3.61	4.21
Envy	Asian	3.68	4.24
Envy	Wealthy	2.66	3.73
Envy	White-collar	3.26	3.92
Pity	Blind	3.82	2.69
Pity	Disabled	3.51	2.24
Pity	Elderly	3.99	2.55
Pity	Gay	3.81	3.68
Pity	Mentally disabled	3.52	1.95

Note. Groups were chosen so they were balanced across SCM stereotypical emotions quadrants, to maximize differences along the Warmth and Competence dimensions for the intersections. Warmth and Competence scores shown represent the pretest average score in a 5-point scale ranging from "Not at all" to "Extremely."

Each participant saw 5 group labels that we expected would be similar in terms of Warmth and Competence, and that would be closer to one of the SCM quadrants in relation to the other groups. The groups were presented one by one in random order and in two sequential blocks, one for the attributes list and one for the scaled ratings.

Participants provided 10 attributes that best described each group in isolation, as well as scaled ratings of Warmth, Competence, and familiarity. Specifically, based on the procedure in Hutter and colleagues (2009), and used in protocols of spontaneous stereotype content (e.g., Nicolas, Bai, et al., 2022), we first asked participants to list 10 "spontaneous thoughts about the characteristics that the type of person" who belongs to the group would possess. We asked

participants to preferably use 1-2 words to define each characteristic. Additionally, we included an open-ended question to allow participants to indicate if there was anything else they would like to know about the target (exploratory; not reported). Subsequently, participants rated each target group on how "sincere" and "friendly" (Warmth), and how "efficient" and "competent" (Competence) they were, using a 5-point scale ranging from 1 (*Not at all*) to 5 (*Extremely*). Drawing from previous use of these scales, we asked participants to rate the targets "as viewed by American society" and to note that "we are not interested in your personal beliefs, but in how you think these people are viewed by others." Additionally, we included three familiarity questions ("How surprised would you be to meet the type of person described above?" "How familiar did you find the person described above?" and "How frequently do you meet people like the one described above?"). Items were averaged, creating scores for Warmth ($\alpha = .73$), Competence ($\alpha = .91$), and familiarity ($\alpha = .86$). Finally, participants completed demographic questions, including an open-ended question to indicate to which of the presented groups they belonged (exploratory; not reported).

Results from the pretesting on average confirmed our a priori expectations of a well distributed SCM map, with groups varying in Warmth and Competence (although there was a higher than expected correlation between Warmth and Competence, $r = .56$).

Study 1

The first main study focuses on scale-measured evaluations of targets' Warmth and Competence. In particular, we examine negativity/positivity and extremity biases for category intersections, such that the ratings of intersectional targets are more in line with the more (vs. less) negative and extreme constituent. As in previous findings from traits and behavior (e.g., Skowronski & Carlston, 1989), we expect to find extremity biases for both Warmth and

Competence (Hypothesis 1a). We also expect to find a negativity bias for Warmth (Hypothesis 1b), and either no valence asymmetry or a positivity bias for Competence (Hypothesis 1c; Wojciszke et al., 1993; see also Rusconi et al., 2020).

Methods

Participants

Participants were 207 Amazon Mechanical Turk workers. The majority were women (52%; 48% men) and White (82%; 6% Black, 6% Asian), with a mean age of 36.2.

Materials & Procedure

Study 1 used the 20 group labels previously pretested as stimuli. The procedure was similar to the Pretest 1, with the following exceptions. First, only Pretest 1's second block, with the scale ratings, was included. Second, instead of evaluating single groups, participants rated Warmth ($\alpha = .88$), Competence ($\alpha = .94$), and familiarity ($\alpha = .85$) for targets who belonged to two of the pretested groups simultaneously. Thus, for example, some participants rated "a person who is Asian & a Welfare recipient" or "A person who is Elderly & an Entrepreneur." Participants saw 10 of these targets, random combinations of the 20 pretest labels, in random order, such that no group was seen twice by the same participant. The order of the social groups for each intersection was random: For each participant, the list of all groups was shuffled, and then each target was formed by sampling without replacement from the list. For the random effects of target, both orders for a given group were treated as the same target.

Data Analysis

Variable Computations. For data preparation, we created a series of new variables. All predictor variables are computed solely from the Pretest 1 ratings, becoming distinct for each

target in Study 1 based on the specific intersection of categories. All variables below were computed for both Warmth and Competence.

Constituents' Average Value. The first set of predictor variables consisted of the constituents' average value for Warmth and Competence of each target. To calculate these variables, we took the scores of each group to which the target belonged (these scores were themselves averages from all pretest participants ratings for each group) and averaged them to obtain the target's Warmth and Competence. Thus, for example, "a person who is White & American," two groups with high-Warmth stereotypes in the US (from Pretest 1, scores of 3.72 & 3.69, respectively: see Table 2) would have a high score on this variable (3.71), while the intersection of "educated" (3.55) and "homeless" (2.47) would have a more middling score (3.01). Average value was mostly used as a control variable, to examine results regardless of the intersection's average Warmth and Competence.

Constituents' Relative Value. The second set of predictor variables were measures of the constituents' relative value between the two groups on each dimension. Basically, for each of these variables, we subtracted the dimension score for the second group on the intersection from the dimension score for the first group on the intersection. Thus, "a person who is Christian & Asian" would receive a Warmth relative value score of 3.96 – 3.68 = 0.28, indicating the degree to which the first group is seen more positively (or negatively) than the second group on the dimension. The corresponding score for Competence for this target would be 3.54 – 4.24 = - 0.7. These variables range from more negative scores indicating the first group is lower on the dimension than the second, to more positive scores, indicating the first group is higher than the second.

Constituents' Relative Extremity. The final set of predictors were measures of the constituents' relative extremity between the two groups on each dimension. For each of these variables, we first calculate the distance from Group 1's Warmth score to the mean of all groups' Warmth (i.e., $|x - \bar{x}|$), and the distance from Group 2's Warmth to the mean of all groups' Warmth, and subtract these two values such that more negative scores indicate that Group 1 is less extreme (relative to the mean) than Group 2, and more positive scores indicate that Group 1 is more extreme than Group 2. Thus, for example, if the mean for Warmth was ~3.4 across all observations, then a target who is "educated" ($|3.55 - 3.4| = 0.15$) and "homeless" ($|2.47 - 3.4| = 0.93$) would get a score of – 0.78 (0.15 – 0.93), indicating that the first group is less extreme on the dimension than the second[6]. We calculate the same score for the Competence dimension.

Intersections' Value Bias. For outcome variables, we calculated the intersection's value-bias variables indicating the extent to which a participant's score for an intersection on a dimension was closer to Group 1's score vs. Group 2's score. Thus, these variables incorporate ratings from both Study 1 and Pretest 1. Specifically, we first calculate the distance from the intersection's Warmth to Group 1's Warmth (i.e., |S1 x – Pretest G1 x|) and the distance from the intersection's Warmth to Group 2's Warmth (i.e., |S1 x – Pretest G2 x|), and subtract these two values such that more negative scores indicate the intersection's score is more similar to Group 1's score and more positive scores indicate that the intersection's score is more similar to Group 2's score. Thus, for example, if a participant rated "a person who is Christian & Wealthy" with a Warmth score of 4, we first obtain the distance between this rating and the Pretest score for "Christian" ($|4 - 3.96| = .04$) and between this rating and the Pretest score for "wealthy" ($|4 - 2.66| = 1.34$). Then, we compute the difference between these scores (.04 – 1.34 = -1.3), which in

[6] We used the mean of each group at the participant-by-group level as the baseline, so it differs between studies. Analytic decisions on how to compute this variable made no difference on the results.

this example reflects that the intersection's Warmth rating was more similar to Group 1 than to Group 2. We calculate the same score for the Competence dimension.

Note that our outcome variable tells us how much the intersection deviates from the constituent's mean towards the first vs. the second group. On average, because the groups are randomized to first and second position, the outcome's mean should not significantly differ from zero if there are no order effects, which is what we find ($p > .05$). We can explain variability on the outcome variable from the relative valence of Group 1 vs. Group 2, and from the relative extremity of Group 1 vs. Group 2. Alternative analyses of bias are explored in the Supplement.

Model Specifications. Across all studies, we use mixed effects model with participants and category intersections as random intercepts and slopes, or a more minimal model if the previous did not converge. Predictors are uncentered (or standardized using the grand mean for Beta coefficients), but different centering strategies provided congruent results. All models include as predictors the constituents' average value, relative value, and relative extremity for the corresponding outcome dimension, and the intersection's value bias for Warmth and Competence as outcomes in two separate models.

Results

Table 3 summarizes the main results (additional results, such as average Warmth and Competence ratings of intersectional targets are included in the supplement and online repository). In line with Hypothesis 1a, we found significant extremity biases for both Warmth and Competence: An intersection's Warmth (or Competence) valence rating is more similar to the Warmth (or Competence) valence of the constituent with the most extreme Warmth (or Competence). Contrary to Hypothesis 1c, we found a negativity bias for Competence, such that an intersection's Competence valence rating is more similar to the Competence valence of the

constituent with the most negative Competence. Also, contrary to Hypothesis 1b, we found no valence bias for Warmth (but see the Supplement, where we find a negativity effect for Warmth in alternative models, and in replication analysis using Study 2 data)[7].

In exploratory analyses, an unexpected pattern appeared when incorporating information about the familiarity of the intersectional target. Specifically, Warmth's relative value and familiarity interacted (p = .004) such that the negativity bias was evident for novel targets (that is, those with the lowest score in the familiarity scale), b = 0.33, 95% CI [0.13, 0.52], *Beta* = .41, $t(1803) = 3.29$, p = .001, while bias was not significant for more familiar intersections, b = .046, 95% CI [-0.02, 0.11], *Beta* = .06, $t(207) = 1.33$, p = .185. These effects held when controlling for Competence-related predictors.

As with Warmth, we also find that Competence's relative value and familiarity interacted (p = .004) such that a negativity bias was strongest for the lowest familiarity targets, b = 0.47, 95% CI [0.30, 0.65], *Beta* = .59, $t(2053) = 5.41$, $p < .001$, while the effect was smaller for more familiar targets, b = 0.22, 95% CI [0.19, 0.26], *Beta* = .27, $t(2063) = 12.11$, $p < .001$. Results held when controlling for Warmth-related predictors.

[7] Note also that valence and extremity biases control for each other (e.g., a negativity bias is present for Warmth when relative value is the sole predictor, but the effect is explained away by the extremity bias effect, see online repository).

Table 3

Confirmatory Results Summary for Study 1.

Outcome (value bias)	**Predictor**	***b* [95% CI]**	***Beta***	***p***	**Interpretation**
Warmth	Warmth average value	-0.01 [-0.08, 0.05]	< 0.01	0.720	
	Warmth relative value	0.06 [-0.01, 0.12]	0.07	0.098	Intersections' Warmth is ***not*** significantly more similar to constituent with most negative Warmth
	Warmth relative extremity	-0.36 [-0.44, -0.27]	-0.26	< .001	Intersections' Warmth is more similar to constituent with most extreme Warmth
Competence	Competence average value	-0.03 [-0.09, 0.04]	0.02	0.388	
	Competence relative value	0.22 [0.16, 0.28]	0.27	< .001	Intersections' Competence is more similar to constituent with most negative Competence
	Competence relative extremity	-0.31 [-0.40, -0.22]	-0.16	< .001	Intersections' Competence is more similar to constituent with most extreme Competence

Note. Results from the two main analysis models. The first shows the intersection's Warmth value bias towards constituent 1 vs 2, with Warmth average value, relative value, and relative extremity as predictors. The second shows the intersection's Competence value bias towards constituent 1 vs 2, with Competence average value, relative value, and relative extremity as predictors. All predictors controlled for each other (and for familiarity of the intersection, *ns*.). Unstandardized coefficient, Standardized coefficient (Beta), *p*-value, and interpretation provided for each predictor.

Discussion

Study 1 provides evidence for valence asymmetries and extremity biases in perceptions of social category intersections across salient societal groups. In particular, multiply-categorized targets showed higher similarities in terms of Warmth and Competence stereotypes to the Warmth and Competence of the constituent group with the most extreme rating in the corresponding dimension. This finding agrees well with previous studies using bottom-up features such as behavior or trait information, which show extremity biases (e.g., Skowronski & Carlston, 1989; Hypothesis 1a).

We expected Warmth to show a negativity bias (Hypothesis 1b), which was only present for some targets. Additionally, Competence showed a negativity bias such that intersectional targets' Competence was more similar to the constituents with the lower Competence. This negativity effect was unexpected for Competence (Hypothesis 1c), based on previous studies which suggest Competence often shows a positivity bias when integrating traits/behaviors (e.g., Wojciszke et al., 1993). These unexpected findings may reflect the shift from studies using behaviors and traits to a study on top-down stereotype combinations, as well as the fact that perceivers may have preexisting stereotypes for the targets presented. In line with this latter point, familiarity (a potential indicator of prototype or exemplar availability, but also of other factors not controlled for here) moderated some of the results. For example, Warmth's negativity effect arose only for low familiarity targets (similar to judgments of unknown targets based on behaviors or traits, as in previous studies). However, Competence's positivity was still not evident when adding familiarity to the model.

Study 2

Study 2 replicates and extends the findings from Study 1. Continuing the focus on extremity and valence asymmetry biases, this study introduces new methods. Specifically, given the distinct information obtained from spontaneous measures of stereotypes (Nicolas, Bai, et al., 2022) here we use open-ended measures and code them using word embeddings, a machine learning model for the analysis of text. Word embeddings capture the semantics contained in the open-ended responses, which includes information not only about Warmth and Competence's valence, as in scales, but also about their prevalence, as well as information about multiple other dimensions of content. Thus, this approach allows us to move beyond specific researcher-determined dimensions of content, such as Warmth and Competence, to study bias at the level of the general stereotype structure across all latent dimensions. We expect to find that incongruence between constituents predicts bias across this semantic structure spanning multiple dimensions (Hypothesis 2a). Moreover, we expect extremity (Hypothesis 2b) and negativity (Hypothesis 2c) on general valence to predict bias at the level of this spontaneous, holistic semantic structure of all stereotype dimensions. In secondary analyses, we explore the role of familiarity.

Methods

Participants

Participants were 196 Amazon Mechanical Turk workers. The majority were women (61%; 39% men) and White (77%; 9% Black, 5% Hispanic, 4% Multiracial, 3% Asian), with a mean age of 37.2.

Materials & Procedure

The materials and procedure for Study 2 were similar to Study 1, with the following exceptions. First, block 1 from the Pretest 1 was added as an initial block, in addition to the

scaled ratings in a second block. Thus, participants first provided 10 open-ended responses for all targets, and then saw the targets again and rated them using scales for Warmth (α = .87), Competence (α = .92), and familiarity (α = .85). Second, participants saw only 4 targets, quasi-randomly assigned such that each participant saw intersections from across the SCM map, without repeating groups. Specifically, we had four between-subject baselines, based on our a priori clustering of groups in the pretest (see Table 2). Each target was a combination of a group from the assigned baseline quadrant with another group from each of the 4 quadrants. Thus, for example, a participant assigned to the admiration baseline saw all 5 groups from this baseline, each randomly paired with a group from each of the quadrants, including admiration. To further illustrate, this example participant could have seen the targets "White & American" (admiration & admiration), "Educated & Homeless" (admiration & contempt), "Middle-class & Asian" (admiration & envy), and "Christian & Blind." (admiration & pity). Note that the baseline group was always shown first, but the order of the trials was randomized, and the non-baseline groups were randomly sampled from each quadrant. We expected these procedures would increase variation along the Warmth and Competence dimensions, while keeping small the number of targets each participant saw. Controlling for baseline made no difference in results, and it is not further discussed.

Data Analysis

Variable Computations. First, for the scale-based variables, we computed the same variables as in Study 1: Constituents' average value, relative value, and relative extremity for Warmth and Competence (replication analyses in Supplement).

General Valence. For this study, we also coded an additional variable from Pretest 1, general valence, using a composite of sentiment dictionaries available through R (see Nicolas et

al., 2021). Dictionaries are lists of words that code for specific content, in this case, valence (negativity-positivity), such that participants' responses that appear in the dictionaries are coded accordingly. The valence scores ranged from -1 (negative) to 1 (positive). For example, using these dictionaries, words such as *attractive* (.96) and *righteous* (.94) score high, while words such as *unfortunate* (-.97) and *perverted* (-.96) score low. Since Pretest 1 included multiple responses per group, a group's general valence score consisted of the average valence across all its Pretest 1 stereotypes. We used these scores similarly to Warmth and Competence ratings and computed for each intersection the constituents' average value, relative value, and relative extremity in terms of valence.

Intersections' Spontaneous Bias. Next, we created additional variables based on the open-ended data. First, we obtained the word embeddings of all lower-cased responses. Word embeddings are numerical vector representations of words derived from machine learning models trained on vast amounts of text data. These embeddings are based on patterns of word co-occurrence in these text corpora (e.g., Google news archives, text data from millions of websites), as words that tend to co-occur with the same set of words tend to be more semantically related to each other. Thus, these embeddings provide information on the semantic similarity between words (by obtaining the cosine similarity between their embeddings[8]), allowing us to model open-ended responses based on semantic content. To illustrate, the word embedding similarity score for the words *friendly* and *amicable* will be higher than the embedding similarity score for the words *friendly* and *short*. Similarity scores can range from -1 (low similarity) to 1 (high similarity). For specifics on the word embeddings used here, see the Supplement and Nicolas and colleagues (2021).

[8] Cosine similarity is used to determine the angle between two vectors. If two vectors point in similar directions (which here encodes semantic meaning), then cosine similarity will be higher.

Once we had word embeddings for all the Pretest 1 responses, for each group, we selected only the embeddings for responses provided more than once and that contained only single words (to remove more idiosyncratic responses and responses from those who did not follow directions) and averaged them together. This provided us with an abstract semantic numerical representation of each group's spontaneous stereotypes, across any latent dimensions/structure captured by the embedding. We refer often to this *abstract or holistic stereotype*: the embeddings function as a black box that contains much more information than would be obtainable through human coding, including information about the semantics of the stereotypes, their prevalence, their valence, and other language properties the model encodes. Once we had these average embeddings (i.e., holistic spontaneous stereotypes), we could then perform operations on these vectors. In particular, we could subtract the embeddings for one group from the embeddings for another group, which provided us with a vector (which we will call the "embeddings bias vector") moving in one direction to the semantic representation of Group 1 and in the opposite dimension to the semantic representation of Group 2 (See Figure 1)[9].

We used these embeddings bias vectors (derived from Pretest 1) in conjunction with data from Study 2 in order to obtain a measure of bias toward one group or the other in the intersection's spontaneous stereotypes. Specifically, we obtained each intersection's (per participant) holistic stereotype representation by averaging the embeddings of all responses to that group. Then, we obtained the cosine similarity between the intersection's embeddings and the embedding bias vector, providing us with a score ranging from -1 (biased toward Group 1's direction) to 1 (biased toward Group 2's direction). Zero would indicate a lack of bias. We call

[9] More specifically, each direction points towards the content that makes each group distinct. Thus this measure may reflect more subtle differences between groups when the two groups' content is correlated.

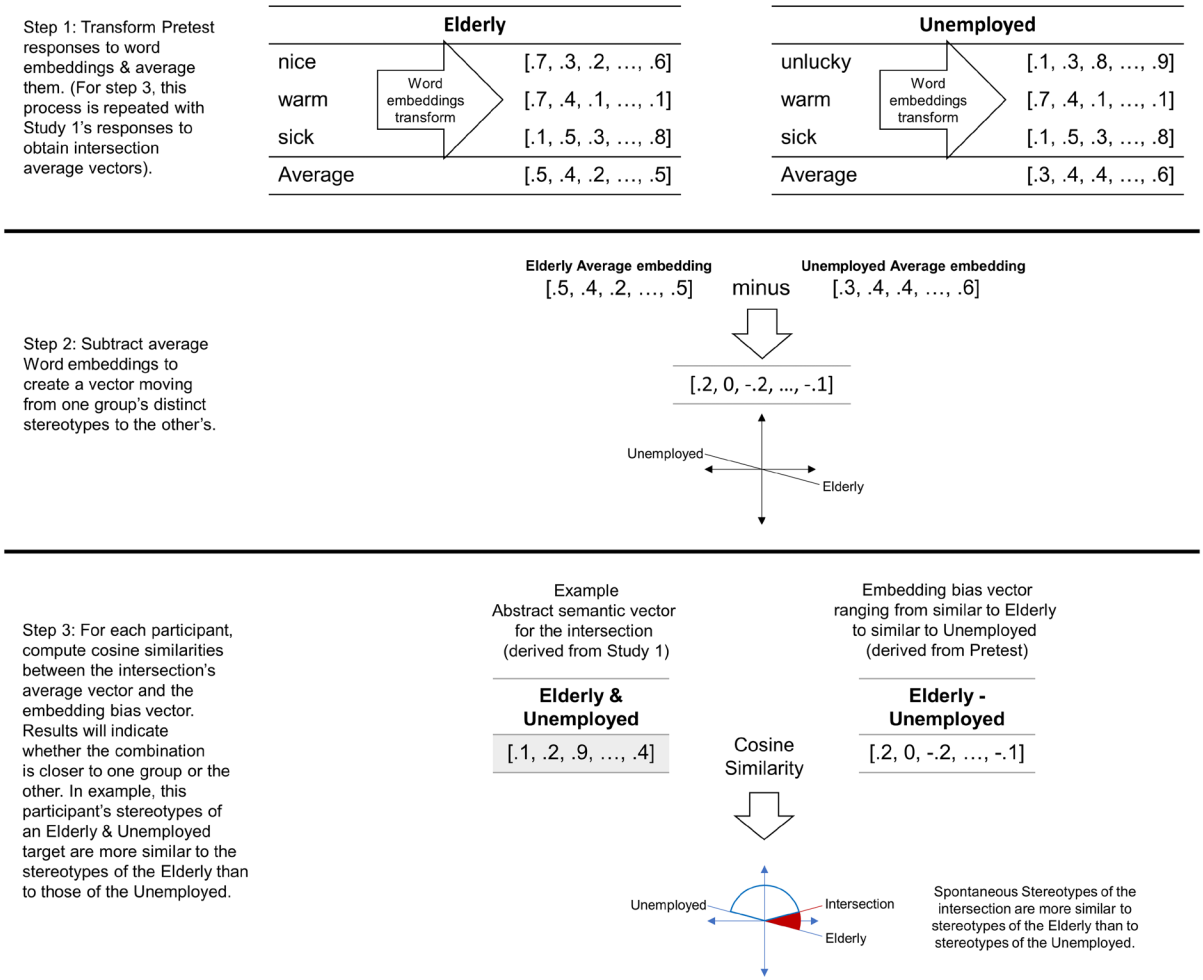
this score the "intersection's spontaneous bias" toward one or the other constituent. In other words, this metric indicates whether the participant's stereotypes for the intersection are more semantically similar to the Pretest 1 single-group stereotypes for Group 1 vs. Group 2. This is akin to the value-bias outcome from Study 1 we computed from each dimension's scales but incorporates information from all dimensions latent in the spontaneous responses. See Figure 1 for a visual representation of these steps.

Constituents' Semantic Incongruence. As a predictor variable, we also calculated the cosine (dis)similarity between the holistic semantic vectors of the two constituent groups (for each intersection), providing a measure of the constituents' semantic incongruence across all stereotypes (i.e., not dimension-specific).

Model Specifications. All models were mixed effects models as specified in Study 1. First, we introduce models with the intersection's spontaneous bias as the outcome. The first model for this outcome has no predictors and indicates the average absolute bias toward either constituent. Subsequent models for the outcome are predicted first by the constituent's spontaneous incongruence, and then by the average value, relative value, and relative extremity of Warmth and Competence (controlling for each other), and of general valence. Analyses replicating Study 1 included in the Supplement.

Figure 1

Computation of the intersection's spontaneous bias towards one constituent group vs. the other.



Note. All data in this figure are hypothetical. First, for each single group, we average across the word embeddings for each of its stereotypes (single-words provided by more than two pretest participants) This variable is an abstract, holistic semantic representation of the group's stereotypes. Second, we obtain embedding bias vectors by subtracting the holistic semantic vectors of each constituents pair. This embedding bias vector moves from the distinct stereotypes of one group to the other's. Third, we incorporate information from Study 1 to obtain cosine similarities to the embedding bias vector. This final step provides a measure of how similar the stereotypes of the intersection are to one group vs. the other (i.e., the intersection's spontaneous bias).

Results

General Evaluation of Bias

Before we presented evidence of bias for the specific Warmth and Competence dimensions. Here, we provide a broader perspective, using word embeddings to capture all the latent semantic dimensions in open-ended stereotypes. Using this approach, in a model predicting intersections' bias toward one vs. another group's direction in the semantic space, the more dissimilar the stereotypes of the constituents, the higher the bias toward one of the constituent's direction in perceptions of the intersection. In other words, semantic incongruence between the stereotypes of the constituents was a significant predictor of bias in the direction of one category over another's: b = .321, 95% CI [0.23, 0.41], *Beta* = .34, $t(148.8) = 6.95$, $p < .001$ (in line with Hypothesis 2a).

Negativity, Positivity, or Extremity bias?

Can we specify whether bias at the abstract semantic level is a function of constituent's relative valence or extremity? We may look at this by using information about the general valence of the spontaneous stereotypes as predictors. General valence is measured across dimensions, and thus better aligns with the general nature of the word embedding holistic stereotype measure.

Using this approach, the holistic semantic stereotypes of the intersection were more similar to the stereotypes of the more extreme constituent, based on general valence, b = .12, 95% CI [0.06, 0.18], *Beta* = .30, $t(183.6) = 4.25$, $p <.001$ (in line with Hypothesis 2b). The holistic semantic stereotypes of the intersection were more similar to the stereotypes of the more negative constituent, based on general valence, b = -0.14, 95% CI [-0.17, -0.11], *Beta* = -.35,

$t(185.75) = -10.1$, $p < .001$ (in line with Hypothesis 2c). None of these effects interacted with familiarity.

Discussion

Study 2 showed, using word embedding coding of spontaneous attributions, that incongruence at the level of abstract, holistic, stereotypes of the constituents led to higher bias in the conjunction, also at the level of holistic stereotypes. Breaking this down further revealed that both negativity and extremity, based on general valence, predicted biases in the structure of the spontaneous stereotype content of the intersections. This suggests that at the level of latent stereotypes, negativity and extremity biases are present in perceptions of multiply-categorized targets, providing further evidence for the cross-dimension generality of these effects at the level of stereotype integration. We note that familiarity did not moderate these effects, unlike in Study 1's results, suggesting differences in the role of familiarity for dimension-specific vs. holistic information integration. For example, it could be that spontaneous content prevalence, a component of holistic but not scale-based measures (see Nicolas et al., 2022), exhibits biases independently of familiarity.

In the last study we expand the number of social categories studied and go beyond valence bias patterns to explore emergent stereotypes. We first present the results of a second Pretest for evaluations of single groups used in Study 3 intersections.

Pretest 2

Pretest 2 follows the same format as Pretest 1 but expands the number of single groups. These data are then used for Study 3 instead of Pretest 1 data.

Methods and Results

Participants were 203 Amazon Mechanical Turk workers. The majority were women (58%; 42% men) and White (82%; 5% Black, 4% Asian, 4% Hispanic), with a mean age of 37.3.

For the main stimuli we returned to the literature and obtained a larger number of group labels. Our final list consisted of 43 group labels, with 29 groups being either completely new (e.g., a *Hacker*) or modifications from our previous list (e.g., *Rich* instead of *Wealthy*). We retained 14 groups from our previous list and obtained new pretest data for them (across both pretests, most of these repeated groups fall in the same relative positions). As in our previous pretest, participants provided attributes that best described each group in isolation, as well as scaled ratings of Warmth and Competence (as in previous studies) and familiarity (we retained only 1 item: "How familiar did you find the type of person described above?"). We lowered the number of open-ended attributes requested to 6 to shorten our survey. Items for the scales were averaged, creating scores for Warmth ($\alpha = .82$) and Competence ($\alpha = .91$). Finally, participants completed demographic questions. Each participant saw 10 groups randomly selected from the list. The groups were presented one by one in random order and in two blocks, one for the attributes list and one for the scaled ratings. For summary information on all the pretested groups for Pretest 2, see Table 4.

Table 4

List of Pretest 2 groups used in Study 3

Group ("A person who is…")	Warmth	Competence	In Study 3	Group ("A person who is…")	Warmth	Competence	In Study 3
a Banker	2.8	3.93	No	an Ivy-leaguer	2.83	3.77	Yes
a CEO	2.6	4.49	Yes	Asian	3.35	4.36	Yes
a Criminal	1.48	1.77	No	Black	3.01	2.96	Yes
a Crossdresser	2.97	2.58	Yes	Blind	3.86	2.91	Yes
a Drug addict	1.69	1.41	No	Christian	3.92	3.53	Yes
a Farmer	4.18	4.21	No	Disabled	3.72	2.76	Yes
a Hacker	2.04	4.01	Yes	Elderly	4.06	2.75	Yes
a Lawyer	2.41	3.93	Yes	Gay	3.63	3.34	Yes
a Nerd	3.53	4.14	Yes	German	3.12	4.02	Yes
a Nurse	4.41	4.58	No	Hispanic	3.41	3.23	Yes
a Politician	2.47	2.36	Yes	Homeless	2.51	1.87	No
a Sex worker	2.28	2.16	Yes	Home-schooled	3.38	3.6	Yes
a Religious extremist	1.92	1.86	Yes	Mentally Disabled	3.81	1.82	Yes
a Scientist	3.4	4.55	Yes	Middle-class	3.8	3.81	No
a Stutterer	3.45	2.79	Yes	Middle eastern	2.39	2.76	Yes
a Teacher	4.28	4.1	Yes	Obese	3.17	2.32	Yes
a Welfare recipient	2.52	2.04	Yes	Republican	2.76	3.14	Yes
American	3.63	3.57	No	Rich	2.68	3.98	Yes
an Accountant	3.17	4.39	Yes	Unemployed	2.94	2.21	Yes
an Engineer	3.39	4.55	No	Vegan	3.34	3.04	Yes
an Undocumented immigrant	2.53	2.38	No	White	3.59	3.71	Yes
an Investor	2.67	4.2	Yes				

The list of 43 groups resulted, as in our previous study, in a positive correlation between Warmth and Competence (r = .44). In order to obtain a more evenly distributed SCM map, prevent semantically impossible intersections ("a person who is Rich & Poor"), and avoid an excessive number of groups from of a particular type (e.g., professions), we dropped 10 of the groups. Also, again, our interest was in having good dispersion across salient societal groups in terms of stereotype content, and for studies looking at emergence, also having a large enough number of novel intersections. This resulted in categorical spaces, such as gender, as well as some race and age categories, not being included, which we understand is a significant tradeoff given the relevance of these categories and should be addressed in future research. The final list of 33 groups resulted in a smaller positive correlation between Warmth and Competence (r = .13). Furthermore, this list diversified the stimuli used, with 70% of the groups being new.

Study 3

Study 3 uses the Pretest 2 groups to expand the previous studies while testing the generalizability of our findings in a larger set of societally salient groups in the U.S. Study 3 also moves beyond questions of bias to study emergence and the distinct stereotypes associated with targets at the intersection of two categories. Specifically, we test whether emergent properties (i.e., perceptions associated with an intersection of social categories that is not present on either of the constituent groups' stereotypes) can be predicted from spontaneous stereotype incongruence (constituent's semantic incongruence**)**, using word embeddings on open-ended data as we did in Study 2. We expect emergent properties to be relatively prevalent (Hypothesis 3a), and that they will increase as incongruence at the level of the holistic semantic structure of the constituents' stereotypes increases (Hypothesis 3b). Moreover, we explore whether incongruence on specific dimensions predicts emergence differentially, using dictionaries (Nicolas et al., 2021)

on the same open-ended data. We hypothesize that, in line with previous research's focus on Competence/Status, emergence would increase with more Competence incongruence (in particular for the Ability facet and the related Status dimension; Hypothesis 3c). We had no hypotheses for the effects of Warmth, Beliefs, or general valence incongruence. Additionally, in line with the idea that emergent traits arrived at through causal reasoning are expected primarily for novel targets (Hutter et al., 2015), we expect that novel targets would elicit higher emergence than more familiar targets (Hypothesis 3d). In an exploratory analysis, we test whether causal reasoning is a mediator between constituent stereotype incongruence and emergent properties.

In further analyses of the spontaneous responses, we look at the content and valence of emergent attributions and compare these to the content and valence of constituent stereotypes. We expect stereotypes that are unique to group intersections to share some structure, such as lower frequency of more traditional content in favor of idiosyncratic content and less prevalent content dimensions, particularly Deviance (Hypothesis 3e; c.f., Kunda & Oleson, 1995).

In addition, although individuation may play a part in the use of emergent attributes (see Hutter & Wood, 2015), the cognitive effort of arriving at emergent properties and the need to deal with inconsistency in the constituent stereotypes may result in more negative emergent impressions of the intersectional targets (Hypothesis 3f; c.f., Proulx & Inzlicht, 2012; Rudman & Fairchild, 2004). Furthermore, we expected valence to vary by dimension (Hypothesis 3g; c.f., Nicolas et al., 2019; Skinner et al., 2019). Study 3 analyses were originally run in an exploratory manner for Study 2 (reported in the Supplement).

Methods

Participants

Participants were 306 Amazon Mechanical Turk workers (larger samples compared to previous studies reflects the higher number of possible group combinations; power remains over 80%). The majority were women (55%; 45% men) and White (75%; 8.5% Black, 6% Asian, 5.6% Hispanic, 3% Multiracial), with a mean age of 35.9.

Materials and Procedure

For Study 3 we used the 33 groups from the trimmed list of groups in Pretest 2.

The procedure was similar to Study 2, with a few exceptions. First, participants saw 6 targets instead of 4. The combination of constituents and their order was also completely random, as in Study 1. In addition, participants provided 6 responses per group, instead of 10. These changes allowed us to keep the survey a similar length, while increasing the number of targets each participant evaluated. We also added instructions before the task indicating that responses "will be completely anonymous and you do not need to personally believe they accurately define these groups of people… we are interested in any characteristics, traits, or descriptions of the groups that come up to your mind." Finally, in the first block, for each group, before providing the open-ended attributes, participants were asked to "write a short paragraph (2-4 sentences) telling a story about the following person's life, based on how you think most people would view this person." Instructions include common phrasing in stereotyping research aimed at incentivizing honest responding (e.g., Fiske et al., 2002).

Data Analysis

Variables Computations. For Study 3 analyses, we start with computations for emergent properties, followed by dictionary and causal reasoning codings.

Emergent Properties. Emergent responses were those provided for the intersections in Study 3 but not for the corresponding constituent groups in Pretest 2. To compute these

variables, we first preprocessed all text responses following the procedure by Nicolas and colleagues (2021; e.g., lower case, removing symbols; see Supplement for details). Then, for each single group, we compiled a list of all the single-word responses provided more than once in Pretest 2. These lists indicated the constituent attributes. Then, for each participant's responses to group intersections in Study 3 we computed the number of responses not included in the constituent's stereotypes, resulting in our "emergence" variable (reported in percentage of participant responses that are emergent). For better alignment with the attribute selection for Pretest 2, we used only single-word responses from Study 3 in emergence analyses.

Content. We used dictionary coding of Study 3 responses to characterize the content of emergent responses and compare them to constituent (non-emergent) responses. This dictionary coding included additional dimensions, such as whether responses where about Emotion, physical Appearance, Geographic origin, among others (see Figure 3 for full list; see Nicolas et al., 2021 for further details). We evaluate the content based both on prevalence (i.e., how many of the participant's responses are *about* the dimension, regardless of valence), as well as the valence variable described below (See Figure 2 for an illustration of the data format and sample of variables).

Valence. Using the approach developed by Nicolas and colleagues (2020), we coded both the Pretest 2 and Study 3 spontaneous responses based on their valence in terms of Sociability, Morality, Ability, Assertiveness, Status, Beliefs, and general valence (i.e., across all responses; same as in Study 2). These scores range from -1 (negative) to +1 (positive) and were obtained from a composite of sentiment dictionaries available through R and trained or developed based on the valence of words on their own, or in the context of product reviews (see Nicolas et al., 2021 for more information). As examples, "unfriendly" would be coded as - .91 for the

Sociability variable, while "amicable" would be coded as +.96; "unqualified" would be coded as -.87 for Ability, but "proficient" would be +.91; "fragile" would be -.79 for Assertiveness, while "diligent" would be +.75; and "fanatic" would be -.69 for (conservative) Beliefs, while "pious" would be +.81. Note that this per-dimension valence variable requires first that the response is categorized, based on the dictionaries, as being about the dimension. If it is, then the valence score is counted in the variable, and all other responses are treated as missing values. Thus, power is lower for less prevalent dimensions (e.g., fewer of the responses are about Beliefs than about Morality; c.f., Nicolas, Bai, et al., 2022) as compared to general valence, which simply uses the valence scores for all responses that have one available (including responses not coded into any of the dictionary dimensions).[10]

Constituents' Average Value & Incongruence. For spontaneous valence coded with the dictionaries, similarly to the scales, for each dimension (and across all responses, i.e., general valence), we calculated the constituents' average value as a predictor of Emergence. Additionally, we calculated the constituents' incongruence (in terms of valence of the dimension), which is simply the absolute value of the constituent's relative value (e.g., |G1's Valence – G2's Valence|). Thus, the constituent's incongruence is similar to the constituent's semantic incongruence variable introduced in Study 2, but specific to a dimension (or to general valence), rather than to the abstract stereotype structure encoded by word embeddings.

[10] Note that Nicolas, Bai, and Fiske (2021) present a different, related construct, "direction", which shares more similarities with traditional scales. For simplicity and consistency with other studies here, we present only valence results in the main text, and include direction information in the online repository data.

Figure 2

Example variables and computations based on Pretest and Main Study data used in Study 3 analyses.

Pretest Data

	Asian		Welfare recipient	
	Stereotype	**(General) Valence**	**Stereotype**	**(General) Valence**
	Intelligent	0.65	Poor	-0.69
	Small	-0.72	Lazy	-0.55
	Kind	0.63	Unlucky	-0.88
	Educated	0.73	Unfortunate	-0.97
	Foreigner	-0.5	Democrat	0
Aggregate		**.16**		**-.62**

Average Value	**Incongruence**
-.23	.78

Main Study Data

Participant	**Intersectional Target**	**Response**	**Emergent**	**Morality Prevalence**	**Ability Prevalence**	**Morality Valence**	**Ability Valence**	**(General) Valence**
122	Asian & Welfare recipient	Intelligent	0	0	1	NA	0.65	0.65
122	Asian & Welfare recipient	Poor	0	0	0	NA	NA	-0.69
122	Asian & Welfare recipient	Cheater	1	1	0	-0.63	NA	-0.63

Aggregate

Participant	**Intersectional Target**	**Response**	**Emergent**	**Morality Prevalence**	**Ability Prevalence**	**Morality Valence**	**Ability Valence**	**(General) Valence**
122	Asian & Welfare recipient	-	.33	.33	.33	-0.63	0.65	-0.22
			...					
157	Asian & Rich		0	0	1	NA	0.75	0.75
210	Nurse & Hacker		.33	.66	.33	-0.05	0.59	0.16

Note. All data in this figure are hypothetical. On top, average value and incongruence (absolute value of their difference) are calculated based on averaged Pretest 2 single-group data. Example shows general valence, but the same variables are computed for each dimension. Note that in actual data, values are weighted by how frequent responses are. The middle shows one participant's three responses to the intersection of previous constituents. Emergence is coded 0 ("Not emergent") if the response to the intersection is present in either of the constituent and 1 ("Emergent") if it is not. Prevalence (for all dimension) indicates whether the response is about the dimension. If the response is about the dimension, the valence variables indicate their negativity-positivity (else, coded as NA). General valence codes the negativity-positivity across all responses regardless of dimension. Finally, we show the averaged version for the example participant and others.

Causal Reasoning. Finally, for the causal reasoning variables, we had two research assistants blind to the hypotheses code all the stories that participants wrote about the targets. The coders were asked to answer the following question: "How much is the writer trying to explain *why* the person possesses both of these attributes?" (*1- Not at all* to *5 – Very much*), as an item theoretically relevant to causal reasoning (e.g., Kunda et al., 1990). Interrater reliability (for the average) was moderate (*ICC* = .55), and thus we use the average of both coders' ratings as our measure of causal reasoning. Other coding questions unrelated to causal reasoning were included for exploratory purposes (see Supplement).

Model Specifications. We examine prevalence and predictors of emergence using linear mixed models (Poisson or logistic models when using counts or binary responses instead of percentages did not seem to make a substantial difference). We present linear models for simplicity and convergence/computational purposes. Models control for the familiarity of the intersection. To compare the content of constituent and emergent properties, we use a linear model looking at the interaction between contrast-coded indicators for whether a response is emergent vs. constituent, and for the dimension into which the response is coded.

Then, we run linear mixed models predicting causal reasoning from incongruence, and emergence from causal reasoning. For this exploratory mediation analysis, we run linear models ignoring the multilevel structure as inputs to the mediation analysis (conducted in R using the *mediation* package; Tingley et al., 2019), given the higher complexity of running mediation analyses on crossed factor multilevel data. To obtain significance estimates of the indirect effect, the model relies on bootstrapping, with 5000 simulations. We want to note that ignoring the multilevel structure is not optimal, as well as note the limitations inherent in mediation analyses (e.g., see Fiedler et al., 2018): This exploratory analysis should be interpreted accordingly.

Results

Prevalence of Emergent Properties

In line with Hypothesis 3a, we find evidence for a substantial number of emergent properties for category intersections, not found in their constituent groups when evaluated independently. On average, 37.2%, 95% CI = [35.7%, 38.7%] of responses in Study 3 were found only in the intersections. Granted, this number may be an overestimate given that we do not account for synonymy, and some of the most idiosyncratic responses were removed from the constituent attributes (by requiring for the response to be provided at least twice).

Predictors of Emergent Properties

See Table 5 for results. As expected (Hypothesis 3b), semantic incongruence between the overarching spontaneous stereotypes of the constituents predicted higher emergence. This pattern can be further broken down by examining whether valence incongruence along specific dimensions predicts emergence. We found no incongruence effects for Warmth, nor its facets. However, in line with Hypothesis 3c, we found significant effects of Competence incongruence, as well as of its facet of Ability and related dimension of Status. Assertiveness, Beliefs, and general valence incongruence had no significant effects.

Finally, as expected (Hypothesis 3d), targets with the lowest familiarity score (vs. all others) elicited higher emergent traits (*M*s = .40 vs .36), $t(186) = 2.90$, $p = .004$, $d = .17$[11].

[11] These measures of novelty and incongruence did not significantly interact (as predicted by, e.g., Hutter et al., 2015), but this was not a planned analysis.

Table 5

Results for incongruence effects on emergence.

Incongruence	***b [95% CI]***	***Beta***	***p***
Spontaneous-Holistic	.31 [.08, .54]	0.07	0.008*
General Valence	.05 [-.01, .10]	0.06	0.081
Warmth	.03 [-.02, .07]	0.04	0.249
Morality	.04 [-.001, .08]	0.06	0.055
Sociability	-.001 [-.05, .05]	-0.002	0.953
Competence	.06 [.01, .10]	0.07	0.018*
Ability	.07 [.01, .09]	0.07	0.009*
Assertiveness	.02 [-.04, .08]	0.02	0.456
Status	.06 [.02, .08]	0.07	0.001*
Beliefs	-.07 [.13, .001]	-0.07	0.053

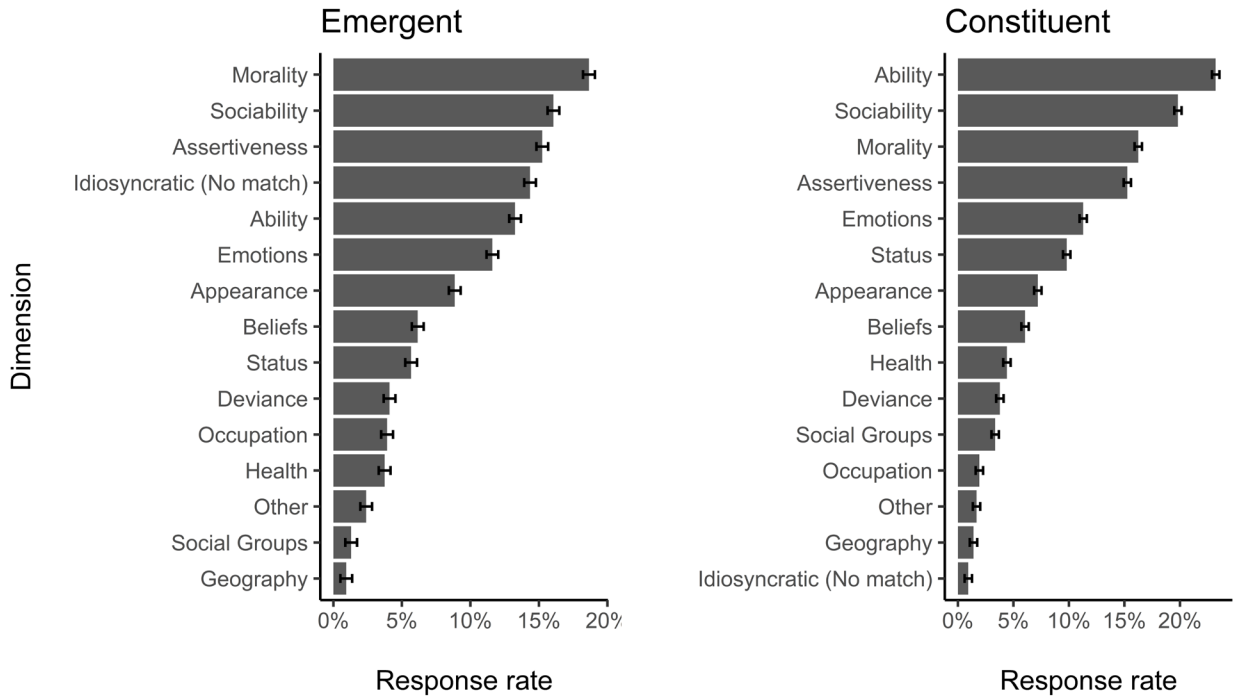
Note. Facets are italicized. * $p < .05$.

Content of Emergent Properties

We found significant differences in the prevalence of the various dimensions in constituent vs. emergent stereotypes, $F(14, 11417.8) = 79.34$, $p < .001$ (see Table 6 and Figure 3, with additional model details in the Supplement). Most notably, emergent (vs. constituent) properties were significantly more idiosyncratic (or at least, not captured by the major dimensions dictionaries) and more about Morality; constituent (vs. emergent) stereotypes were more about Ability, Sociability, and Status (supporting Hypothesis 3e, but Morality effect was not hypothesized).

Figure 3

Proportion of responses coded into each dimension, based on whether the responses is Emergent or not (Constituent).



Note. Error bars are Standard Errors.

Table 6

Emergent vs. Constituent Prevalence and Valence differences per dimension.

Dimension	Prevalence		Valence	
	d	***p***	***d***	***p***
Ability	0.37	< .001*	0.26	< .001*
Appearance	-0.06	0.002*	-0.23	< .001*
Assertiveness	0.00	0.975	0.15	0.001*
Beliefs	0.00	0.841	-0.10	0.140
Deviance	-0.01	0.561	0.27	0.002*
Emotions	-0.01	0.554	-0.02	0.674
Geography	0.02	0.392	-0.17	0.309
Health	0.03	0.214	0.08	0.363
Idiosyncratic	-0.50	< .001*	0.08	0.487
Morality	-0.09	< .001*	-0.04	0.292
Occupation	-0.07	< .001*	0.00	0.985
Other	-0.03	0.190	-0.25	0.042*
Sociability	0.14	< .001*	0.23	< .001*
Social Groups	0.08	< .001*	0.07	0.660
Status	0.15	< .001*	0.10	0.129

Note. Cohen's d values for are provided for the constituent - emergent mean difference. * $p <$ .05.

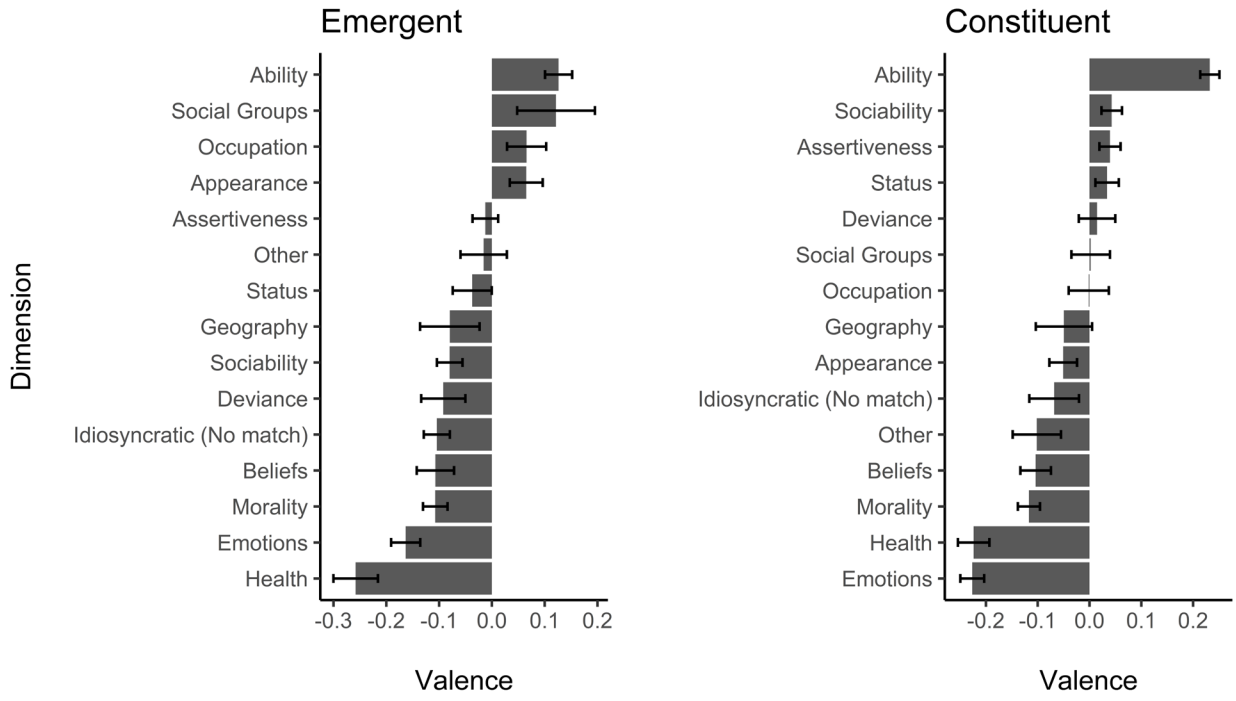
Valence of Emergent Properties

First, an evaluation of general valence reveals that emergent stereotypes are more negative (M = -.06) than constituent stereotypes (M = .007), $t(1012) = 6.19$, $p < .001$, $d = .11$ (Hypothesis 3f). Second, distinguishing between dimensions, we find significant differences in the valence of constituent vs. emergent stereotypes, $F(14, 12550) = 6.54$, $p < .001$ (see Table 6 and Figure 4; supplement for additional model details). Results suggest that Ability, Assertiveness, Sociability, and Deviance stereotypes were more positive in the constituent than the emergent responses. Appearance stereotypes were more positive in emergent perceptions (Hypothesis 3g).

We note that results so far mostly replicate exploratory analyses of Study 2 as well as Study 3S, with some variability across results (see Supplement).

Figure 4

Valence of responses (negative to positive) coded into each dimension, based on whether the responses is Emergent or not (Constituent).



Note. Error bars are Standard Errors.

Causal Reasoning and Emergence -Exploratory

Results indicate that causal reasoning was a predictor of emergence, $b = .011$, 95% CI = [.001, .02], *Beta* = .05, $t(1744) = 2.24$, $p = .026$. At the same time, dissimilarity on the holistic semantic content between the constituents was a significant predictor of causal reasoning in the stories describing the intersections: $b = 2.74$, 95% CI = [1.55, 3.92], *Beta* = .14, $t(475.6) = 4.53$, $p < .001$. Furthermore, a mediation analysis suggested that our data are consistent with the

model: Stereotype incongruence ➔ Causal reasoning ➔ Emergence (but see discussion of alternative models below for all mediation analyses), *Indirect effect* = .03, 95% *CI* = [.001, .06], $p = .038$. There was no effect of novelty on causal reasoning, $p = .281$. Additional exploratory analyses per dimension included in the Supplement.

Discussion

Study 3 examined emergent properties, which we define as attributes found in perceptions of the intersectional target but not representative of constituent groups' stereotypes. In this study, both the incongruence of the constituents' holistic stereotypes and the novelty of the intersection predicted emergence. This is in line with what may be expected from a causal reasoning account of emergence, in which people construct new impressions for more incongruent and more novel intersections (c.f. Hutter, et al., 2015). However, emergent properties, as defined here, may also result from simple exemplar or prototype retrieval, if the stereotypes of these intersections' exemplars or prototypes have distinct properties (e.g., due to non-prototypicality, subtyping, subgrouping).

When looking at incongruence along specific dimensions, we found an effect of Competence incongruence (as well as related facets of Ability and Status) predicting emergence, suggesting a primary role of Competence-related factors (vs. Warmth) in emergent properties. As noted earlier, most studies on this topic used category intersections that apparently differed along a Competence-related dimension, so we expected a pattern along these lines (perceptions of Status correlate highly with perceptions of Competence; Fiske et al., 2002). However, given that abstract semantic incongruence reliably predicts emergence in our data, it is likely that the holistic stereotypes are incorporating important information about other features (beyond valence) that perceivers find incongruent when integrating information about constituent groups.

For example, the holistic stereotypes derived from the word embeddings may also code for dimensional content representativeness (what the stereotype is *about*, regardless of valence), incongruence on the coherence between the stereotypes (are the stereotypes of one constituent more coherent than the other?), and various other linguistic features (e.g., are the stereotypes of one constituent more complex or more distinct compared to the other constituent's?). Given the black-box nature of the embeddings, we cannot determine which features are most important in this study, but the features suggested can be explored in future research.

Next, Study 3 presented an analysis of the prevalence and valence of different stereotype dimensions by whether the stereotype is emergent or not. Robust patterns (see Supplement for additional analyses) include a higher prevalence of emergent attributes related to Morality as well as more idiosyncratic content, but lower prevalence of emergent stereotypes related to major stereotype dimensions, such as Sociability, Ability, and Status. A valence examination reveals that in general, emergent attributes are more negative than constituent ones, suggesting that despite the possibility of increased individuation, the stereotype incongruence that leads to emergence may result in backlash and more negative evaluations (c.f., Rosette et al., 2016; Rudman & Fairchild, 2004). Further analyses of valence reveal that when emergent stereotypes do deal with Ability, Assertiveness, Sociability, or Deviance, they tend to be more negative, compared to constituent stereotypes (c.f., for example, "confused" and lack of belonging stereotypes of people categorized as Multiracial; Skinner et al., 2019). Appearance stereotypes were more positive when emergent (c.f., again as an example, perceptions of Multiracial people as "beautiful"; Nicolas et al., 2019).

Study 3 explored causal reasoning as one mediator of emergence in our sample of intersections of salient representative U.S. groups. We find that causal reasoning predicted

emergence, as expected. Causal reasoning was in turn predicted by constituents' stereotype incongruence at the abstract semantic level. Furthermore, conditional on the model assumption Incongruence ➔ Causal reasoning ➔ Emergence, our exploratory mediation analyses suggests that causal reasoning accounted for a significant portion of the model variance. These results are in line with previous research, suggesting that (particularly for novel targets), complex reasoning about incongruence results in emergent attributes that attempt to resolve the cognitive conflict and form a coherent impression of a multiply-categorized target. When the constituents have congruent stereotypes, it is easier to inherit those attributes or average them. Again, we expect other mediators (e.g., prototypicality) to operate here, given emergent properties in familiar conjunctions that likely activate preexistent exemplars and/or prototypes (c.f., Hutter et al., 2015). Because of the exploratory nature of the analysis and the limitations of mediation analyses, results should be interpreted as preliminary, and will not be discussed further. See the supplement for additional testing and discussion about the mediation analyses.

Finally, Study 3 varied the instructions to make them less likely to elicit social desirability biases. Given the robustness of the replication analyses presented in the Supplement (as well as results from other studies using a variety of instructions for open-ended responses; e.g., Nicolas, Bai, et al., 2022) these findings do not seem to depend on social desirability (at least as controlled for here; c.f. Fiske et al., 2002).

General Discussion

Social psychology has been slow to examine impressions and stereotypes of multiply-categorizable individuals. Our study joins a relatively small but growing number of studies exploring how multiple categorization affects social cognition. However, by taking a generalist perspective, our studies provide an initial framework to understand differential patterns between

single- and multiply-categorized targets across all major dimensions of content, and across a sample of representative and salient societal groups. This sample includes intersections studied relatively more frequently, such as race and age or race and sexual orientation, but also many others such as status- and beliefs- defined groups, occupations, and other groups that people often use to categorize others and themselves (both in the U.S. and around the world), to make sense of and interact with their society. As such, we examine this diverse sample of social groups through generalizable measures such as their valence or stereotypical content, and whether these are incongruent or extreme. Under this generalist framework, based on prevalent stereotype content models (Fiske et al, 2002; Koch et al., 2020; Nicolas, Bai, et al., 2022), we both confirm theories derived from more specific group intersections, and provide novel insights into the content and process of intersectional stereotypes. Several patterns appear consistently (see Table 7).

Table 7

Summary of confirmatory main findings.

Main findings	Specific Findings	Hypothesis
Intersectional stereotypes show a negativity bias	Intersectional targets' holistic stereotype structure is more similar to the holistic stereotype of the constituent with the more negative stereotypes.	2a, 2c
	Intersectional targets' Competence is more similar to the Competence of the constituent with the more negative Competence.	1c
	Intersectional targets' Warmth is more similar to the Warmth of the constituent with the more negative Warmth.	1b
Intersectional stereotypes show an extremity bias	Intersectional targets' holistic stereotype structure is more similar to the holistic stereotype of the constituent with the more extreme stereotypes.	2a, 2c
	Intersectional targets' Competence is more similar to the Competence of the constituent with the more extreme Competence.	1a
	Intersectional targets' Warmth is more similar to the Warmth of the constituent with the more extreme Warmth.	1a
Intersectional perceptions have significant prevalence of emergent attributes	Over 35% of perceptions of intersectional targets are emergent (i.e., not found in the constituent groups).	3a
	Emergent properties are *more* about idiosyncratic content and Morality (vs. attributes also found in the constituents).	3e
	Emergent properties are *less* about traditional dimensions of Ability, Sociability, and Status (vs. attributes also found in the constituents).	3e
Emergent (vs. constituent) attributes are more negatively valenced	Emergent attributes are in general more negative than attributes also found in the constituents' stereotypes.	3f
	Emergent attributes about Ability, Assertiveness, Sociability, and Deviance are more negative (vs. attributes also found in the constituents).	3g
	Emergent attributes about Appearance are more positive (vs. attributes also found in the constituents).	3g
Constituents' stereotype incongruence and Intersection novelty predict emergence	Higher dissimilarity (i.e., incongruence) of the constituents' holistic stereotype structure predicts more emergent attributes for the intersectional target.	3b
	Higher incongruence of the constituents' Competence stereotypes predicts more emergent attributes for the intersectional target.	3d
	Novel (vs. familiar) intersectional targets elicit more emergence	3c

Note. Most analyses show robust patterns across replications included in the Supplement. The hypothesis codes include the study number and hypothesis letter.

We find that across groups, the big two dimensions of Warmth and Competence exhibit extremity effects, such that a multiply-categorized target's Warmth/Competence is more similar to the Warmth/Competence of the constituent with the more extreme score on the dimension. This finding is in line with a breadth of research on information integration, and several mechanisms have been posited to explain its existence, including that extreme traits occur less frequently and are more distinctive, making them more informative and diagnostic (e.g., Fiske, 1980; Skowronski & Carlston, 1989). These same mechanisms are thought to underlie valence asymmetries, such as negativity or positivity biases, albeit with potential variation depending on the dimension of content. For example, while a negativity bias has been established when moral traits and behaviors are combined (and more generally in the literature), a positivity bias is sometimes found for competence-related traits and behaviors (see e.g., Reeder & Brewer, 1979; Skowronski & Carlston, 1987; Mende-Siedlecki et al., 2013). Here, we found a negativity bias for both Competence and Warmth (particularly for novel targets). Multiple possibilities arise for this discrepancy, including that we are dealing with the combination of two top-down pieces of information (social categories) rather than relatively more bottom-up information (traits and behaviors). In other words, social categories are informative depending on the prototypes and exemplars retrieved from memory, and are thus more variable across participants, can be more selectively applied, and are not as strongly associated with the outcome of interest, in this case attributions of Warmth or Competence (as opposed to behaviors or traits that are more consensually and strongly associated with the outcome of interest). Extremity and negativity biases were also found when using a holistic spontaneous measure of stereotype structure.

Follow-up questions may be asked about this pattern. For example, to what extent are these biases the result of online information integration vs. being already present in retrieved

prototypes? Given that familiarity significantly moderated some of the valence asymmetry effects, it is possible that both are operating here, differently. To illustrate, extremity biases occurred regardless of familiarity, suggesting they may operate similarly in preexisting stereotypes (e.g., in long-term stereotype learning) and at impression formation about a relatively novel target. But valence asymmetries operated differently, with negativity biases for Warmth and Competence often being stronger for novel group intersections, suggesting it may be occurring more often at an information integration step than being ingrained in preexistent stereotypes about the intersection. Future research should further disentangle these processes, as well as answer related questions such as what mechanisms may lead to preexisting intersecting categories stereotypes showing these biases.

Beyond bias, we also examined patterns of emergence in the stereotypes of intersections of salient social categories. Previous research exploring specific group conjunctions has found multiple ways in which their stereotypes and perceptions are distinct from those of their constituent (e.g., Hutter et al., 2015; Purdie-Vaughns & Eibach, 2008). Here, we also found evidence of substantial emergent perceptions across many intersections and explored predictors and content patterns. Although a strict definition of emergence may involve perceptions that arise from complex interactions of information about the constituents, we use a more inclusive definition to mean any perceptions found in a category conjunction not found in the constituents, regardless of whether they arise from an online process based on such complex interactions of information, or preexistent prototypes/exemplars that may or may not have involved such complex interactions at encoding (c.f. Hutter et al., 2015).

Among patterns of emergence we find, in line with existing theorizing, that both the novelty of the conjunction and the stereotypical incongruence (at the level of abstract semantic

structure) of the constituents are relevant in predicting emergence (Hutter et al., 2015). The role of novelty highlights a path consistent with a strict definition of emergence, where intersections with no preexistent stereotypes are ascribed emergent properties that attempt to reconcile incongruence between the constituents. However, we also find emergent properties (and predicted by incongruence) in non-novel intersectional targets, suggesting that preexistent stereotypes about these groups already incorporate emergent stereotypes. This may again be due to similar conflict resolution (e.g., causal reasoning) processes that occur at encoding of these pre-existing stereotypes, as well as the result of subgrouping when the intersection includes non-prototypical groups (e.g., Black women, where when thinking about Black people perceivers tend to think about Black men, and when thinking about women tend to think of White women), among other potential mechanisms.

When looking at incongruence at the level of specific dimensions, we find the most robust pattern for Competence incongruence. Most previous studies on emergence of novel targets seemed to combine groups based on incongruence along a dimension resembling Competence (i.e., Ability, Assertiveness, or Status). In addition, Status and power are particularly relevant in intersectionality accounts, and incongruence along this dimension has been previously explored from the target's perspective (e.g., King et al., 2019). Our results suggest that, compared to other specific dimensions, targets combining high and low Competence constituents (e.g., a rich welfare recipient) trigger more emergent attributions. Status, as a dimension that is more structural (vs. psychological) than others studied here (see Nicolas, Fiske, et al., 2022), may constitute a particularly conflicting category that requires perceivers to concoct emergent properties that explain away the incongruence. On the other hand, incongruence along Warmth dimensions might be resolved primarily through biased

information integration, discarding information about one of the constituents and focusing on the other (more extreme, more negative; c.f., behavior-trait literature, Reeder & Brewer, 1979).

In addition to predictors, we also used insights and methods from the Spontaneous Stereotype Content Model (SSCM; Nicolas, Bai, et al., 2022) to examine the content of emergent attributes, and how they compare to stereotypes present in the constituents. We found some expected results, such as more idiosyncratic responses for emergent properties, but also found some unexpected patterns, such as higher number of Morality-related responses in emergent (vs. constituent) attributes, while constituent attributes instead focused on the rest of the large content domains (Ability, Sociability, Status). In particular, Status (and to a smaller extent, Ability) being more common in the constituent responses is a notable pattern, given that Status incongruence was a reliable predictor of emergence. In other words, status incongruence may be more likely solved by moving away from the conflicting dimension and focusing emergent properties on alternative content. These findings suggest that a more spontaneous approach, accompanied by an understanding of a taxonomy more comprehensive than traditional two-dimensional models (such as Warmth and Competence) is needed to capture the nuances associated with perceptions of multiply-categorizable targets.

Finally, emergent attributes tended to be more negative than constituent ones, both in a general valence metric, and independently for several dimensions. In particular, greater emergent negativity was present for Ability, Assertiveness, and Sociability. Potentially related patterns have been previously seen in studies of specific multiply-categorized targets, such as Black-White Multiracials being stereotyped as "confused" and "not belonging" (Skinner et al., 2019), or even novel intermediate/conjunction groups being judged as less "socially real" (Burke, 2016). Stereotypes about Deviance were also more negative when emergent, suggesting that

some of these intersectional targets may seem norm-violating or strange. As one exception, Appearance stereotypes appeared more positive when emergent, potentially paralleling perceptions of Multiracial people as "beautiful" (Nicolas et al., 2019), and perhaps related to a compensation strategy (c.f., Durante et al., 2017), where higher negativity along major dimensions is partially offset by higher positivity in a non-psychological dimension.

Theoretical Implications

Theoretically, our findings are a generalist examination of multiple theories that have been tested either at different levels (e.g., traits and behaviors) or with specific social category intersections (e.g., age and race). Here, by using conjunctions of a large sample of salient groups in a society (the United States), we were able to study overarching properties of the groups to test against existing theories and uncover novel patterns. For example, we find that assumptions of information integration from traits and behaviors, such as a positivity bias for Competence (Reeder & Brewer, 1979; Skowronski & Carlston, 1987), do not necessarily apply to multiply-categorized targets' stereotypes. However, some biases did arise more consistently, including extremity and negativity (Fiske, 1980; Skowronski & Carlston, 1987). This should highlight the need for future theorizing on multiple categorization stereotypes to distinguish between preexisting prototypes, which may incorporate bias at encoding, and online impression formation which more directly aligns with the traits and behaviors information integration literature.

Our findings are consistent with emergence theories (Hutter et al., 2015) that posit the roles of intersection novelty and stereotype incongruence as predictors of emergence. This again aligns with the distinction of preexisting exemplars and prototypes versus online impression formation solely from the information given. Furthermore, it has implications for the integration of content and process models (e.g., the SCM and the continuum model; Fiske et al., 2002; Fiske

& Neuberg, 1990). Specifically, Status, a structural dimension strongly related to Competence, was one of the more robust predictors of emergence. This may suggest that incongruence along a relatively consensual, non-psychological (yet very socially relevant) dimension may be harder to disregard and requires more cognitive engagement to resolve, including emergent inferences and individuation. Future research could explore the connection between specific content dimensions, emergence, and individuation more directly, including understanding when individuation of targets with multiple activated categories may heighten negativity (Rudman & Fairchild, 2004).

Another theoretical contribution to highlight: Our exploration on the content of multiple categorization stereotypes speaks to psychological intersectionality theories about the unique ways in which multiply-categorized targets are perceived. A particularly robust finding is a high level of relatively idiosyncratic emergent content, which highlights the need to understand that specific category intersections will have unique properties requiring a non-generalist understanding, along the lines of much current theorizing (see Nicolas et al., 2017; Petsko & Bodenhausen, 2019; Remedios & Sanchez, 2018). However, we also find patterns related to well-known dimensions of content reliably differing, such as higher Morality but lower Sociability and Competence-related emergent (vs. constituent) attributes (the latter of which also tended to be more negative than their constituent counterparts). This suggests that further explorations of mechanisms may uncover shared pathways through which these dimensions become accessible in perceptions of intersecting categories.

Finally, our use of both scaled and open-ended metrics analyzed through text analysis methods adds to evidence from the SSCM (Nicolas, Bai, et al., 2022). The SSCM, developed using single-group stereotypes, found significant prevalence of not only the big two dimensions

of Warmth and Competence (Fiske et al., 2002), but also their facets (i.e., Sociability, Morality, Ability, and Assertiveness; Abele et al., 2016), dimensions from alternative models (e.g., Beliefs and Status; Koch et al., 2020), as well as understudied dimensions such as Health and cross-sectional group stereotypes. The stereotypes of multiply-categorized targets largely reflect the general structure of content of single-group stereotypes, from the prevalence of the various dimensions to their valence (albeit with the previously discussed differences based on the constituent vs. emergent breakdown). By revealing not only the valence of the different dimensions used to stereotype intersectional targets, but also how representative (i.e., prevalent) a dimension is of a target's stereotypes, the spontaneous stereotypes examined here reveal how these variables may diverge. For example, we find that the valence (which is the variable that most traditional scale-based studies explore) of Morality is not significantly different between emergent and constituent perceptions (Morality attributes tend to be negative for both). However, because Morality attributes are more *prevalent* in emergent perceptions, it implies that the multiply-categorized target is moralized more often, and because Morality attributions tend to be negative, so may be the effective evaluations of these targets along this dimension (this is also supported by additional analyses of our data; see online repository). Thus, the SSCM applied to multiple categorization provides new insights into stereotyping, spanning multiple existing models, and improving the predictive power of our models (see Nicolas, Bai, et al., 2022).

Practical Implications

At the practical level, our studies have multiple takeaways. First, by acknowledging and exploring perceptions of targets at the intersection of multiple categories, our findings add to a growing literature calling for psychologically informed interventions and policies to take into deeper consideration the complex dynamics of stereotypes and impression formation associated

with real-world intergroup relations (see e.g., Hall et al., 2019; Remedios & Sanchez, 2018). In fact, categorization along multiple dimensions has been previously linked to reduced prejudice (e.g., Crisp et al., 2001), but our results add more nuance to evaluate when strategies to incorporate multiple categorization for prejudice reduction may work.

Beyond consequences for targets perceived at the intersections of categories, our results may also provide insights about constituent stereotype change. Drawing from the subtyping literature (see Richards & Hewstone, 2001), we may think of intersectional targets with high emergence arrived at through causal reasoning as subtyped from both of the constituents. Encoding of these intersectional targets' perceived attributes may not help change stereotypes of their constituents. Instead, these intersectional targets are either individuated deviants or represented into a separate categorical space from both constituents. Furthermore, intersectionality models suggest that these targets may become invisible, resulting for example in lack of organizational influence (see Hall et al., 2019). Alternatively, parallels with the subgrouping literature may suggests that intersectional targets that do not engage as much causal reasoning may be typical enough of one or both constituents. This typicality may lead the perceiver to view the constituents' stereotypes as more variable and help reduce overgeneralization (see Richards & Hewstone, 2001).

Our findings also have parallels to the literature on multiple identities from the perspective of the target (e.g., conflict resolution and emergence may relate to the process of achieving identity coherence; c.f., Gardner & Garr-Schultz, 2017), potentially informing strategies for improving the well-being of stigmatized targets with complex social identities.

Finally, through the use of spontaneous measures, the framework employed here, and our findings, may be used to further our understanding of how biases related to intersecting identities

are replicated and reinforced in social media and through widely deployed Artificial Intelligence algorithms (c.f., Guo & Caliskan, 2020), in order to minimize their contribution to inequality.

Limitations and Future Directions

Our findings in general confirm many predicted patterns from the bias, emergence, and psychological intersectionality literature, and at the same time provide new insights that are hypothesis-generating and rife with potential future directions. In addition, there are certainly also limitations that should be addressed in future research (some of these have been addressed in specific studies' discussions). First, we focused on examining relatively high-level characteristics of social groups, such as valence incongruence. However, certainly many more characteristics of the constituent groups could have been examined as predictors of bias and emergence. In fact, previous literature has identified several variables that could have an impact, including differential weight given to identities along societally salient dimensions such as race and gender (e.g., Levin, et al., 2002) or perceived conflict and symbolic threat (e.g., Grigoryan et al., 2020). We make our data available for further exploration, and future studies could explore additional factors using our approach.

Our findings are also context-agnostic. As much research on stereotype content, we explore stereotypes about social groups under whatever default or activated context the particular participant has during the task. However, we know that specific contexts can alter perceptions of single groups (e.g., see Nicolas & Skinner, 2017). And in fact, context may be of particular importance when attempting to understand the different weights assigned to different social groups (e.g., Crisp & Hewstone, 2007; Petsko et al., 2022). For example, in a context where there is salient racial conflict, perceivers may give higher weight to stereotypes of the racial category constituent in an intersection (c.f., Petsko et al., 2022). Future studies would benefit

from incorporating context, behavioral, and trait information (as well as individual differences) into the framework used here, to further disentangle these effects. Moreover, future research would benefit from further expanding the framework by exploring perceptions of intersections of more than two social categories.

As another limitation to highlight, several of our finding are inconclusive as to the specific mediators underlying them. For example, as we have discussed previously, perceptions of multiply-categorized targets can be the end product of multiple dynamic cognitive processes (c.f., Freeman & Ambady, 2011); these range from retrieving particular exemplars from memory, retrieving group prototypes, online algebraic information integration (e.g., averaging), and engaging in causal complex reasoning, among potential others. Disentangling these processes is relevant to understand where biases are occurring, and to better fit models of continuum group-to-individuated perceptions, along which these different mechanisms vary. We do provide some evidence for some of these processes, but it is unclear how big a role each plays, and if they are involved differently in the various descriptive patterns we found.

Constraints on Generality

A final limitation to discuss relates to generalizability constraints based on sample characteristics. Our studies relied of mostly online White participants from the United States, as a convenience sample, limiting our understanding of how cultural factors may impact our findings. For example, cross-cultural and cross-lingual results from the SSCM highlight that different societies may have differing levels of prevalence for various stereotype dimensions studied here (Nicolas, Bai, & Fiske, 2022), either because of the types of groups that are salient in the society, or because of the stereotypes that get attached to them. Added complexity is introduced when considering that the prototypicality of different categories, the familiarity with

different intersectional targets, and other intergroup and structural dynamics may also vary across cultures. Thus, it will be very important for future research to explore cultural moderators of our findings, for example by exploring nation-level structural factors' interplay with stereotype content (e.g., Bai et al., 2020).

Conclusion

As an initial attempt at exploring potentially generalizable patterns in perceptions of multiply-categorized targets that we know face distinct biases, and who are evaluated through highly dynamic and expansive information processing, our framework and findings may provide one path toward a deeper understanding of the intricacies of social cognition.

Context of the research

For both authors, these studies provided an opportunity to move closer to an understanding of the nuances of social cognition and the challenges that perceivers face in an increasingly diverse world. For the first author, the research was also an opportunity to expand spontaneous approaches to person perception and for the second author to tie models of content and process to which she has contributed (including her dissertation on negativity and extremity biases). Along the way from motivation to realization, we have done interdisciplinary work to develop the methods, engaged in adversarial collaborations that have strengthened the underlying theories, and collaborated in multiple other projects that shaped how we approached the current research. We expect to continue developing the ideas in this paper in future research, including examining the role of structural factors through cross-cultural research, examining intersectional stereotypes embedded in Artificial Intelligence models, and further strengthening the link between models of stereotype content and processes.

References

Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*, *128*(2), 290–314. https://doi.org/10.1037/rev0000262

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness—Communion with warmth and morality. *Frontiers in Psychology, 7*, 1810. https://doi.org/10.3389/fpsyg.2016.01810

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

Alves, H., Koch, A., & Unkelbach, C. (2017). The "common good" phenomenon: Why similarities are positive and differences are negative. *Journal of Experimental Psychology: General, 146*(4), 512–528. https://doi.org/10.1037/xge0000276

Anderson, N. H. (1965). Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology, 70*(4), 394-00. https://doi.org/10.1037/h0022280

Blinded for review (2022). Valence Biases and Emergence in the Stereotype Content of Intersecting Social Categories: Supplement. https://osf.io/8kad4/?view_only=669b91ffe24a43a9b8fefb9bb850e2e2

Bai, X., Ramos, M. R., & Fiske, S. T. (2020). As diversity increases, people paradoxically perceive social groups as more similar. *Proceedings of the National Academy of Sciences*, *117*(23), 12741-12749. https://doi.org/10.1073/pnas.2000333117

Bodenhausen, G. V., Kang, S. K., & Peery, D. (2012). Social Categorization and the Perception of Social Groups. In S. Fiske, & C. N. Macrae (Eds.), *The SAGE handbook of social cognition* (pp. 311-329). SAGE Publications Ltd.

Burke, S. E. (2016). *The excluded middle: Attitudes and beliefs about bisexual people, biracial people, and novel intermediate social groups* (Doctoral dissertation, Yale University).

Clausell, E., & Fiske, S. T. (2005). When do subgroup parts add up to the stereotypic whole? Mixed stereotype content for gay male subgroups explains overall ratings. *Social Cognition*, *23*(2), 161-181. https://doi.org/10.1521/soco.23.2.161.65626

Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology, 92*(6), 1006–1023. https://doi.org/10.1037/0022-3514.92.6.1006

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum, 1989*(1), 139-167. http://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8

Crisp, R. J., & Hewstone, M. (2007). Multiple social categorization. *Advances in Experimental Social Psychology, 39*, 163–254. https://doi.org/10.1016/S0065–2601(06)39004–1

Crisp, R. J., Hewstone, M., & Rubin, M. (2001). Does multiple categorization reduce intergroup bias? *Personality and Social Psychology Bulletin, 27*(1), 76–89. https://doi.org/10.1177/0146167201271007

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, *92*(4), 631–648. https://doi.org/10.1037/0022-3514.92.4.631

Dovidio, J. F., Love, A., Schellhaas, F. M., & Hewstone, M. (2017). Reducing intergroup bias through intergroup contact: Twenty years of progress and future directions. *Group Processes & Intergroup Relations*, *20*(5), 606-620. https://doi.org/10.1177/1368430217712052

Dupree, C. H., & Fiske, S. T. (2019). Self-presentation in interracial settings: The competence downshift by White liberals. *Journal of Personality and Social Psychology*, *117*(3), 579-604. https://doi.org/10.1037/pspi0000166

Durante, F., Tablante, C. B., & Fiske, S. T. (2017). Poor but warm, rich but cold (and competent): Social classes in the stereotype content model. *Journal of Social Issues, 73*(1), 138–157. https://doi.org/10.1111/josi.12208

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Fiedler, K., Harris, C., & Schott, M. (2018). Unwarranted inferences from statistical mediation tests—An analysis of articles published in 2015. *Journal of Experimental Social Psychology, 75,* 95–102. https://doi.org/10.1016/j.jesp.2017.11.008

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology, 38*(6), 889–906. https://doi.org/10.1037/0022-3514.38.6.889

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, *23*, 1-74. https://doi.org/10.1016/S0065-2601(08)60317-2

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*(6), 878-902. https://doi.org/10.1037//0022-3514.82.6.878

Fiske, S. T., Nicolas, G., & Bai, X. (2021). Stereotype content model: How we make sense of individuals and groups. In P. A. M. Van Lange, E. T. Higgins, & A. W. Kruglanski (eds.). *Social psychology: Handbook of basic principles* (3rd ed). New York: Guilford.

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

Galinsky, A. D., Hall, E. V., & Cuddy, A. J. C. (2013). Gendered races: Implications for interracial marriage, leadership selection, and athletic participation. *Psychological Science, 24*(4), 498–506. https://doi.org/10.1177/0956797612457783

Gardner, W. L., & Garr-Schultz, A. (2017). Understanding our groups, understanding ourselves: The importance of collective identity clarity and collective coherence to the self. In *Self-concept clarity* (pp. 125-143). Springer, Cham.

Gawronski, B., & Brannon, S. M. (2019). *What is cognitive consistency, and why does it matter?* In E. Harmon-Jones (Ed.), *Cognitive dissonance: Reexamining a pivotal theory in psychology* (p. 91–116). American Psychological Association. https://doi.org/10.1037/0000135-005

Ghavami, N., & Peplau, L. A. (2013). An intersectional analysis of gender and ethnic stereotypes: Testing three hypotheses. *Psychology of Women Quarterly, 37*(1), 113–127. https://doi.org/10.1177/0361684312464203

Grigoryan, L., Cohrs, J. C., Boehnke, K., van de Vijver, F. (A. J. R.), & Easterbrook, M. J. (2020). Multiple categorization and intergroup bias: Examining the generalizability of three theories of intergroup relations. *Journal of Personality and Social Psychology.* Advance online publication. https://doi.org/10.1037/pspi0000342

Guo, W., & Caliskan, A. (2020). Detecting Emergent Intersectional Biases: Contextualized Word Embeddings Contain a Distribution of Human-like Biases. arXiv. https://arxiv.org/abs/2006.03955

Hall, E. V., Hall, A. V., Galinsky, A. D., & Phillips, K. W. (2019). MOSAIC: A model of stereotyping through associated and intersectional categories. *The Academy of Management Review, 44*(3), 643–672. https://doi.org/10.5465/amr.2017.0109

Hastie, R., Schroeder, C., & Weber, R. (1990). Creating complex social conjunction categories from simple categories. *Bulletin of the Psychonomic Society*, *28*(3), 242-247. https://doi.org/10.3758/BF03334016

Hutter, R. R., & Wood, C. (2015). Applying individuation to conflicting social categories. *Group Processes & Intergroup Relations*, *18*(4), 523-539. https://doi.org/10.1177/1368430214550424

Hutter, R. R. C., Allen, R. J., & Wood, C. (2015). The formation of novel social category conjunctions in working memory: A possible role for the episodic buffer? *Memory, 24*(4), 496-512. https://doi.org/10.1080/09658211.2015.1020814

Hutter, R. R., Crisp, R. J., Humphreys, G. W., Waters, G. M., & Moffitt, G. (2009). The dynamics of category conjunctions. *Group Processes & Intergroup Relations*, *12*(5), 673-686. https://doi.org/10.1177/1368430209337471

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology, 85*(4), 616–626. https://doi.org/10.1037/0022-3514.85.4.616

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*, 116–131. https://doi.org/10.1037/a0025335

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *The Journal of Abnormal and Social Psychology, 28*(3), 280. https://doi.org/10.1037/h0074049

King, T. L., Shields, M., Shakespeare, T., Milner, A., & Kavanagh, A. (2019). An intersectional approach to understandings of mental health inequalities among men with disability. *SSM-population health, 9*, 100464. https://doi.org/10.1016/j.ssmph.2019.100464

Kunda, Z., & Oleson, K. C. (1995). Maintaining stereotypes in the face of disconfirmation: Constructing grounds for subtyping deviants. *Journal of Personality and Social Psychology, 68*(4), 565–579. https://doi.org/10.1037/0022-3514.68.4.565

Kunda, Z., Miller, D. T., & Claire, T. (1990). Combining social concepts: The role of causal reasoning. *Cognitive Science*, *14*, 551–577. https://doi.org/10.1207/s15516709cog1404_3

Koch, A., Yzerbyt, V., Abele, A., Ellemers, N., & Fiske, S. T. (2021). Social evaluation: Comparing models across interpersonal, intragroup, intergroup, several-group, and many-

group contexts. *Advances in Experimental Social Psychology*, *63*, 1-68. https://doi.org/10.1016/bs.aesp.2020.11.001

Koch, A., Imhoff, R., Unkelbach, C., Nicolas, G., Fiske, S. T., Terache, J., Carrier, A., & Yzerbyt, V. (2020). Groups' warmth is a personal matter: Understanding consensus on stereotype dimensions reconciles adversarial models of social evaluation. *Journal of Experimental Social Psychology, 89*, 103995. https://doi.org/10.1016/j.jesp.2020.103995

Leach, C., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality vs. competence and sociability in the evaluation of in-groups. *Journal of Personality and Social Psychology, 93*, 234-249. https://doi.org/10.1037/0022-3514.93.2.234

Levin, S., Sinclair, S., Veniegas, R. C., & Taylor, P. L. (2002). Perceived discrimination in the context of multiple group memberships. *Psychological Science*, *13*(6), 557-560. https://doi.org/10.1111/1467-9280.00498

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*(1), 93-120. https://doi.org/10.1146/annurev.psych.51.1.93

Maurer, K. L., Park, B., & Rothbart, M. (1995). Subtyping versus subgrouping processes in stereotype representation. *Journal of Personality and Social Psychology, 69*(5), 812–824. https://doi.org/10.1037/0022-3514.69.5.812

Mende-Siedlecki, P., Baron, S. G., & Todorov, A. (2013). Diagnostic value underlies asymmetric updating of impressions in the morality and ability domains. *The Journal of Neuroscience, 33*(50), 19406–19415. https://doi.org/10.1523/JNEUROSCI.2334-13.2013

Nicolas, G., & Skinner, A. L. (2017). Constructing race: How people categorize others and themselves in racial terms. In H. Cohen, & C. Lefebvre (Eds.), Handbook of

categorization in cognitive science (2nd ed., pp. 607–635). California: Elsevier Science. https://10.1016/B978-0-08-101107-2.00025-7

Nicolas, G., Bai, X., & Fiske, S. T. (2022). A spontaneous stereotype content model: Taxonomy, properties, and prediction. Journal of Personality and Social Psychology. Advance online publication. https://doi.org/10.1037/pspa0000312

Nicolas, G., Bai, X., & Fiske, S. T. (2021). Comprehensive stereotype content dictionaries Using a semi-automated method. *European Journal of Social Psychology, 51*(1), 178-196. https://doi.org/10.1002/ejsp.2724

Nicolas, G., de la Fuente, M., Fiske, S. T. (2017). Mind the overlap in multiple categorization: A review of crossed categorization, intersectionality, and multiracial perception. *Group Processes & Intergroup Relations, 20*(5): 621–631. https://doi.org/10.1177/1368430217708862

Nicolas, G., Skinner, A. L., & Dickter, C. L. (2019). Other than the sum: Hispanic and Middle Eastern categorizations of Black–White mixed-race faces. *Social Psychological and Personality Science, 10*(4), 532-541. https://doi.org/10.1177/1948550618769591

Nicolas, G., Fiske, S. T., Koch, A., Imhoff, R., Unkelbach, C., Terache, J., Carrier, A., & Yzerbyt, V. (2022). Relational versus structural goals prioritize different social information. *Journal of Personality and Social Psychology, 122*(4), 659–682. https://doi.org/10.1037/pspi0000366

Norris, C. J., Leaf, P. T., & Fenn, K. M. (2019). Negativity bias in false memory: moderation by neuroticism after a delay. *Cognition and Emotion*, *33*(4), 737-753. https://doi.org/10.1080/02699931.2018.1496068

Petsko, C. D., & Bodenhausen, G. V. (2019). Racial stereotyping of gay men: Can a minority sexual orientation erase race? *Journal of Experimental Social Psychology, 83,* 37–54. https://doi.org/10.1016/j.jesp.2019.03.002

Petsko, C. D., & Bodenhausen, G. V. (2020). Multifarious person perception: How social perceivers manage the complexity of intersectional targets. *Social and Personality Psychology Compass*, *14*(2), e12518. https://doi.org/10.1111/spc3.12518

Petsko, C. D., Rosette, A. S., & Bodenhausen, G. V. (2022). Through the looking glass: A lens-based account of intersectional stereotyping. *Journal of Personality and Social Psychology*. Advance online publication. https://doi.org/10.1037/pspi0000382

Phinney, J. S., & Alipuria, L. L. (2006). Multiple social categorization and identity among multiracial, multiethnic, and multicultural individuals: Processes and implications. In R. J. Crisp & M. Hewstone (Eds.), *Multiple social categorization: Processes, models and applications* (p. 211–238). Psychology Press.

Proulx, T., & Inzlicht, M. (2012). The five "A"s of meaning maintenance: Finding meaning in the theories of sense-making. *Psychological Inquiry, 23*(4), 317–335. https://doi.org/10.1080/1047840X.2012.702372

Purdie-Vaughns, V., & Eibach, R. P. (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles*, *59*, 377–391. https://doi.org/10.1007/ s11199–008–9424–4

Quattrone, G. A., & Jones, E. E. (1980). The perception of variability within in-groups and out-groups: Implications for the law of small numbers. *Journal of Personality and Social Psychology, 38*(1), 141–152. https://doi.org/10.1037/0022-3514.38.1.141

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review, 86*(1), 61–79. https://doi.org/10.1037/0033-295X.86.1.61

Reeder, G. D., & Spores, J. M. (1983). The attribution of morality. *Journal of Personality and Social Psychology, 44*(4), 736–745. https://doi.org/10.1037/0022-3514.44.4.736

Remedios, J. D., & Sanchez, D. T. (2018). Intersectional and dynamic social categories in social cognition. *Social Cognition*, *36*(5), 453-460. https://doi.org/10.1521/soco.2018.36.5.453

Richards, Z., & Hewstone, M. (2001). Subtyping and subgrouping: Processes for the prevention and promotion of stereotype change. *Personality and Social Psychology Review, 5*(1), 52–73. https://doi.org/10.1207/S15327957PSPR0501_4

Rosette, A. S., de Leon, R. P., Koval, C. Z., & Harrison, D. A. (2018). Intersectionality: Connecting experiences of gender with race at work. *Research in Organizational Behavior, 38*, 1-22. https://doi.org/10.1016/j.riob.2018.12.002

Rudman, L. A., & Fairchild, K. (2004). Reactions to counterstereotypic behavior: The role of backlash in cultural stereotype maintenance. *Journal of Personality and Social Psychology, 87*(2), 157–176. https://doi.org/10.1037/0022-3514.87.2.157

Rusconi, P., Sacchi, S., Brambilla, M., Capellini, R., & Cherubini, P. (2020). Being honest and acting consistently: Boundary conditions of the negativity effect in the attribution of morality. *Social Cognition*, *38*(2), 146-178. https://doi.org/10.1521/soco.2020.38.2.146

Schug, J., Alt, N. P., & Klauer, K. C. (2015). Gendered race prototypes: Evidence for the non-prototypicality of Asian men and Black women. *Journal of Experimental Social Psychology, 56,* 121–125. https://doi.org/10.1016/j.jesp.2014.09.012

Schug, J., Alt, N. P., Lu, P. S., Gosin, M., & Fay, J. L. (2017). Gendered race in mass media: Invisibility of Asian men and Black women in popular magazines. *Psychology of Popular Media Culture, 6*(3), 222–236. https://doi.org/10.1037/ppm0000096

Sesko, A. K., & Biernat, M. (2010). Prototypes of race and gender: The invisibility of Black women. *Journal of Experimental Social Psychology, 46*(2), 356–360. https://doi.org/10.1016/j.jesp.2009.10.016

Skinner, A. L., Perry, S. P., & Gaither, S. (2020). Not quite monoracial: Biracial stereotypes explored. *Personality and Social Psychology Bulletin, 46*(3), 377–392. https://doi.org/10.1177/0146167219858344

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*(4), 689–699. https://doi.org/10.1037/0022-3514.52.4.689

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin, 105*(1), 131–142. https://doi.org/10.1037/0033-2909.105.1.131

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

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis.

Townsend, S. S. M., Markus, H. R., & Bergsieker, H. B. (2009). My choice, your categories: The denial of multiracial identities. *Journal of Social Issues, 65*(1), 185–204. https://doi.org/10.1111/j.1540-4560.2008.01594.x

Wilson, J. P., Remedios, J. D., & Rule, N. O. (2017). Interactive effects of obvious and ambiguous social categories on perceptions of leadership: When double-minority status may be beneficial. *Personality and Social Psychology Bulletin, 43*(6), 888–900. https://doi.org/10.1177/0146167217702373

Wojciszke, B., Brycz, H., & Borkenau, P. (1993). Effects of information content and evaluative extremity on positivity and negativity biases. *Journal of Personality and Social Psychology, 64*(3), 327. https://doi.org/10.1037/0022-3514.64.3.327