Good is generally more alike than bad in the information ecology: Investigating the case of (the ABC model of) group stereotypes

Inauguraldissertation

zur

Erlangung des Doktorgrades

der Humanwissenschaftlichen Fakultät

der Universität zu Köln

nach der Prüfungsordnung vom 10.05.2010

vorgelegt von

Alex Koch

aus

Klausenburg

Abgabedatum: 07.03.2016

Diese Dissertation wurde von der
Humanwissenschaftlichen Fakultät
der Universität zu Köln
im Juni 2016 angenommen.

Erklärung

Chapter 1 enthält Teile des folgenden Manuskripts:

Alves, H., Koch, A., & Unkelbach, C. (in press) Why Good Is More Alike Than Bad: Processing Implications. *Trends in Cognitive Sciences.*

Hans Alves und ich haben die Idee entwickelt. Hans Alves und ich haben das Manuskript geschrieben, und Christian Unkelbach hat zu jedem Schritt wertvolle Vorschläge beigetragen.

Chapter 2 beruht auf folgendem Manuskript:

Koch, A., Alves, H., Krüger, T., & Unkelbach, C. (2016). A general valence asymmetry in similarity: Good is more alike than bad. *Journal of Experimental Psychology: Language, Memory, and Cognition*, *42*, 1171-1192.

Ich habe die Idee entwickelt, die Datenerhebung überwacht, die Analyse der Daten durchgeführt, und das Manuskript geschrieben. Hans Alves, Tobias Krüger, und Christian Unkelbach haben zu jedem Schritt wertvolle Vorschläge beigetragen.

Chapter 3 beruht auf folgendem Manuskript:

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency / socio-economic success, conservative-progressive beliefs, and communion. *Journal of Personality and Social Psychology: Attitudes and Social Cognition*, *110*, 675-709.

Roland Imhoff und ich haben die Idee entwickelt. Ich habe die Datenerhebung überwacht und die Analyse der Daten durchgeführt. Roland Imhoff hat den Einleitungsteil des Manuskripts geschrieben. Ich habe den Methoden-, Ergebnis-, und Diskussionsteil des Manuskripts geschrieben. Roland Imhoff, Ron Dotsch, Christian Unkelbach, und Hans Alves haben zu jedem Schritt wertvolle Vorschläge beigetragen.

Chapter 4 enthält Teile des folgenden Manuskripts:

Imhoff, R., & Koch, A. (in press). How orthogonal are the Big Two of social perception? On the curvilinear relation between agency and communion. *Perspectives on Psychological Science.*

Roland Imhoff und ich haben die Idee entwickelt. Roland Imhoff und ich haben die Datenerhebung überwacht und die Analyse der Daten durchgeführt. Roland Imhoff hat das Manuskript geschrieben. Ich habe das Manuskript überarbeitet.

Alex Koch, Köln, März 2016

Auszug

Für viele Bewertungsdimensionen (z.B. die Temperatur, der Geschmack, und der Nährwert einer Mahlzeit) gilt: Ein Spektrum positive Zustände ist von zwei Spektren negativer Zustände umgeben: zu viel und zu wenig. Diese Verteilung von Zuständen in der Informationsökologie resultiert in einer höheren Ähnlichkeit positiver Objekte, Personen, und Ereignisse im Vergleich zu negativen Stimuli. Das heißt, es gibt weniger Möglichkeiten positiv zu sein als negativ zu sein. Stimuli können oftmals nur auf eine einzige Art und Weise positiv sein (z.B. bei einer guten Mahlzeit müssen sowohl die Temperatur als auch der Geschmack und der Nährwert angemessen sein). Stimuli können aber in vielerlei Hinsicht negativ sein (z.B. eine schlechte Mahlzeit kann zu heiß oder zu kalt, zu scharf oder zu fad, oder zu fett oder zu karg sein). Diese höhere Ähnlichkeit von positiver verglichen zu negativer Information ist relevant, weil Ähnlichkeit auf nahezu allen Ebenen der Informationsverarbeitung einen großen Einfluss auf Schnelligkeit und Genauigkeit hat. Wenn die höhere Ähnlichkeit von positiver verglichen zu negativer Information also ein allgemeingültiges Phänomen ist, dann könnte dieses Phänomen eine Reihe von Valenzasymmetrien in der Informationsverarbeitung vorhersagen / erklären. Ich zeige, dass die höhere Ähnlichkeit von positiver verglichen zu negativer Information in der Tat ein allgemeingültiges Phänomen ist, welches auf tausende Wörter und Bilder zutrifft. Außerdem zeige ich, dass die höhere Ähnlichkeit von positiver verglichen zu negativer Information auch auf soziale Gruppen zutrifft. Gruppen, welche hinsichtlich Durchsetzungskraft / sozioökonomischer Status (A für agency) und konservative-progressive Überzeugungen (B für beliefs) als durchschnittlich beurteilt werden, werden hinsichtlich Gemeinschaftssinn (C für communion) als hoch beurteilt. Gruppen, welche hinsichtlich A und B als extrem beurteilt werden, werden hinsichtlich C als niedrig beurteilt. Da durchschnittliche Gruppen einander ähnlicher sind als extreme Gruppen, sind positive verglichen zu negativen Gruppen ähnlicher. Zum Abschluss diskutiere ich Implikationen dieses ABC-Modells von Gruppenstereotypen, wobei ich auf weiterführende Forschungsmöglichkeiten zum Einfluss von Stereotypen auf soziale Wahrnehmung, Kognition, und Verhalten hinweise.

Schlagwörter: Valenzasymmetrien, mentale Repräsentation, Ähnlichkeit, natürliche Stimulusauswahl, räumliches Anordnen, Allgemeingültigkeit, Inhalt von Stereotypen, Gruppen, Durchsetzungskraft / sozioökonomischer Erfolg, konservative-progressive Überzeugungen, Gemeinschaftssinn, ABC-Model

Abstract

On most if not all evaluatively relevant dimensions such as the temperature level, taste intensity, and nutritional value of a meal, one range of adequate, positive states is framed by two ranges of inadequate, negative states, namely too much and too little. This distribution of positive and negative states in the information ecology results in a higher similarity of positive objects, people, and events to other positive stimuli as compared to the similarity of negative stimuli to other negative stimuli. In other words, there are fewer ways in which an object, a person, or an event can be positive as compared to negative. Oftentimes, there is only one way in which a stimulus can be positive (e.g., a good meal has to have an adequate temperature level, taste intensity, and nutritional value). In contrast, there are many different ways in which a stimulus can be negative (e.g., a bad meal can be too hot or too cold, too spicy or too bland, or too fat or too lean). This higher similarity of positive as compared to negative stimuli is important, as similarity greatly impacts speed and accuracy on virtually all levels of information processing, including attention, classification, categorization, judgment and decision making, and recognition and recall memory. Thus, if the difference in similarity between positive and negative stimuli is a general phenomenon, it predicts and may explain a variety of valence asymmetries in cognitive processing (e.g., positive as compared to negative stimuli are processed faster but less accurately). In my dissertation, I show that the similarity asymmetry is indeed a general phenomenon that is observed in thousands of words and pictures. Further, I show that the similarity asymmetry applies to social groups. Groups stereotyped as average on the two dimensions agency / socio-economic success (A) and conservative-progressive beliefs (B) are stereotyped as positive or high on communion (C), while groups stereotyped as extreme on A and B (e.g., managers, homeless people, punks, and religious people) are stereotyped as negative or low on C. As average groups are more similar to one another than extreme groups, according to this ABC model of group stereotypes, positive groups are mentally represented as more similar to one another than negative groups. Finally, I discuss implications of the ABC model of group stereotypes, pointing to avenues for future research on how stereotype content shapes social perception, cognition, and behavior.

Keywords: valence asymmetries, mental representation, similarity, natural sampling, spatial arrangement, generality, stereotype content, groups, agency / socio-economic success, conservative-progressive beliefs, communion, ABC model

Structure

Chapter 1 – Introduction

Valence or the continuum from positive to negative is surely one of the most important dimensions of meaning (Lazarus, 1966; Lewin, 1963; Osgood, Suci, & Tannenbaum, 1957), as distinguishing between “good for me” and “bad for me” is key to survive and thrive. Accordingly, the valence literature provides a plethora of insights about how positive and negative information is learned, stored, retrieved, and processed. A fundamental insight is that there is a big difference in how people learn, store, retrieve, and process positive compared to negative information, a phenomenon known as valence asymmetry in cognitive processing (Anderson, 1965; Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001; Taylor, 1991). To give a few examples, negative information binds people’s attention for a longer time (Pratto & John, 1991). Negative information is recognized with greater accuracy (Alves et al., 2015) and recalled in greater detail (Kensinger, Garoff-Eaton, & Schacter, 2006). Negative information is also learned better (Fazio, Eiser, & Shook, 2004) and weighs more in object and person perception (Knobe, 2003; Peeters & Czapinski, 1990). However, people detect, classify, and process positive information faster (Becker, Anderson, Mortensen, Neufeld, & Neel, 2012; Unkelbach, Fiedler, Bayer, Stegmüller, & Danner, 2008), lump positive information into fewer categories (Koch, Alves, Krüger, & Unkelbach, 2016), and generalize more between positive entities such as traits and behaviors (Gräf & Unkelbach, 2015). In sum, processing valence is a vital part of human cognition, and valence asymmetry occurs on virtually all levels of cognitive processing. Thus, to understand human cognition, it is important to understand how valence asymmetry in cognitive processing comes about.

Traditionally, valence asymmetry in cognitive processing is explained by the higher affective and motivational potential of negative compared to positive information (“bad is stronger than good”; Baumeister et al., 2001, p. 323). Specifically, to initiate efforts that end or avoid a problematic situation, negative stimuli have a greater affective and motivational impact on the organism (see also Rozin & Royzman, 2001). A closely related account is that negative stimuli elicit a stronger affective and motivational reaction with the purpose of mobilizing resources to remove a present threat and to minimize its damage (Taylor, 1991). The resources mobilized / efforts initiated in the face of negative stimuli are increased vigilance, improved encoding and retrieval etc., resulting in different forms of valence asymmetry in cognitive processing to be explained. No such resources / efforts are required in the face of positive objects, people, and events. In contrast, to prolong and maximize the pleasure from hedonic situations, individuals minimize the resources used / efforts expended for processing positive stimuli (Baumeister et al., 2001), resulting in higher speed as well as increased superficiality (e.g., fewer categories, more generalizations) – that is, other forms of valence asymmetry in cognitive processing to be explained. Recently, Unkelbach and colleagues (2008) proposed an alternative explanation, namely that positive information is more similar to other positive information compared to negative information's similarity to other negative information – that is, good is more alike than bad.

Stimulus similarity greatly impacts all stages of cognitive processing (Goldstone & Son, 2005; "sameness is […] the backbone of our thinking", James, 1980/1950; p.459), including classification, categorization, generalization, judgment, recognition, and recall (Nosofsky, 1986; Shepard, 1987). For example, facilitated processing effects are stronger for similar prime-target pairs (Perea, Duñabeitia, & Carreiras, 2008). Exemplars are classified faster and more accurately if they are similar to the category's prototype (Nosoksky, 1986; Smith & Sloman, 1994). People are more likely to apply previous decisions in similar situations (Shepard; 1987; Tenenbaum & Griffiths, 2001). Similar stimuli are more likely to be recalled (Roediger & McDermott, 1995) but also more often confused with one another (DeSoto & Roediger, 2014). Given that stimulus similarity influences all stages of cognitive processing, the proposed higher similarity of positive compared to negative information might explain valence asymmetry in cognitive processing (Unkelbach, 2012).

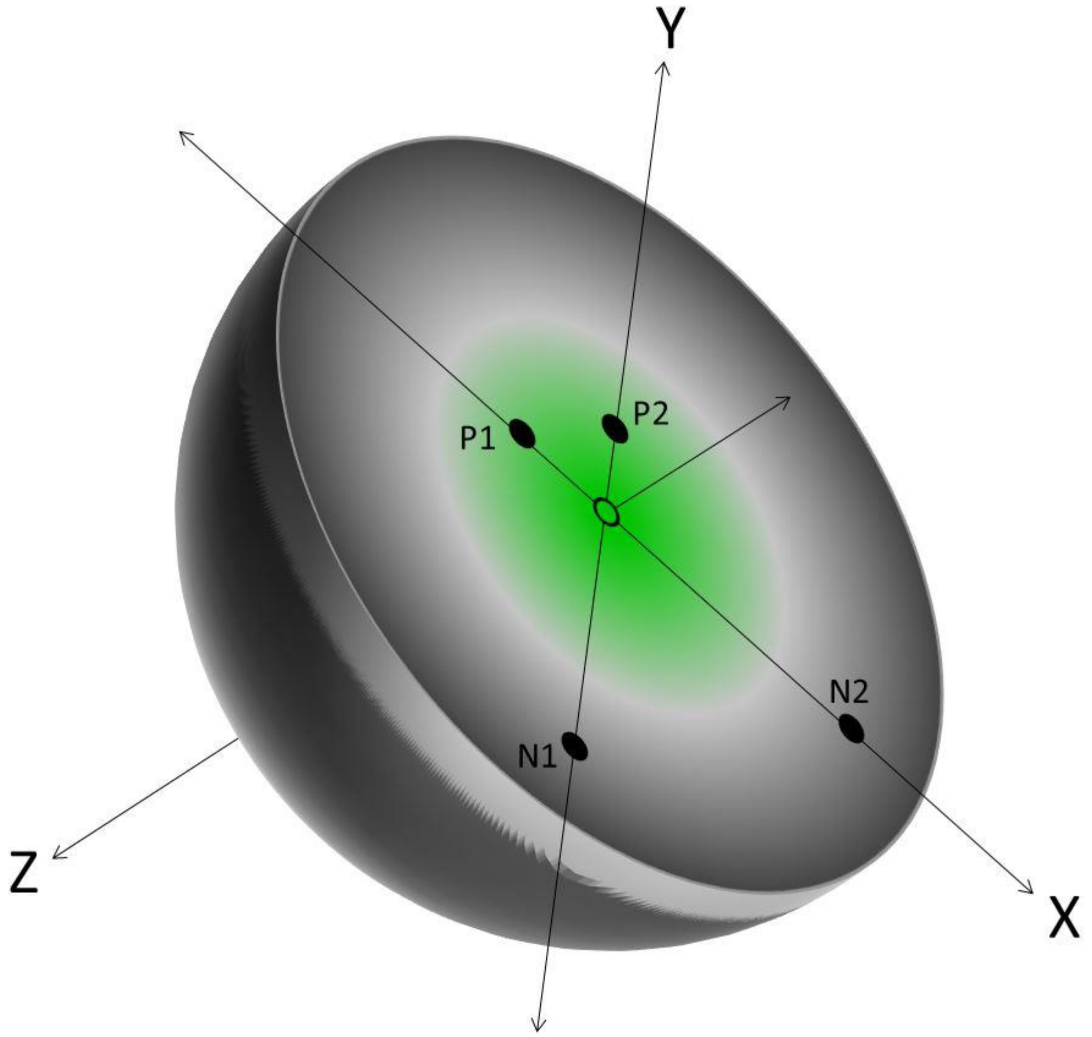
There is evidence in support of this. Unkelbach and colleagues (2008) showed that the higher classification speed of positive targets is due to their higher similarity (see also Chapter 2, Study 1). The lower sensitivity and higher response bias for positive stimuli has been found to be due to their higher similarity, too (Alves et al., 2015, see also Chapter 2, Study 1). Also, halo effects are more likely to occur for positive traits, because positive information is more similar (Gräf & Unkelbach, 2016). Further, the effect of additional positive information during impression formation might be less strong (Knobe, 2003; Peeters & Czapinski, 1990) because positive information is more similar and thus more repetitive. The lower precision for details during recall of positive information (Kensinger, Garoff-Eaton, & Schacter, 2006) might also due to that positive information is more similar and thus more easily confused. Taken together, valence asymmetry in stimulus similarity already explains several forms of valence asymmetry in cognitive processing (e.g., processing speed and memory accuracy), and there are other forms that valence asymmetry in stimulus similarity might explain.

To count as an alternative explanation of valence asymmetry in cognitive processing, valence asymmetry in similarity should at least in principle be independent of valence asymmetry in affective-motivational potential. The first aim of this work is to propose a homeostatic model of valence in which valence asymmetry in stimulus similarity can occur independently of valence asymmetry affective-motivational potential. If this model is correct, valence asymmetry in stimulus similarity indeed qualifies as an explanation of valence asymmetry in cognitive processing that is independent of, and thus complementary to, the traditional explanation in terms of valence asymmetry in affective-motivational potential.

Valence asymmetry in stimulus similarity as a novel and independent explanation of valence asymmetry in cognitive processing

To model that valence asymmetry in similarity is in principle independent of valence asymmetry in affective-motivational potential, I extend a classical, geometric model of similarity. In this model, stimuli (i.e., objects, people, events etc.) are points in a metric space with n dimensions. Each dimension scales one stimulus feature that is evaluatively relevant within a given set and context of stimuli. The space equates proximity to similarity – therefore, stimuli that lie closer to one another are more similar to one another. I propose that on all dimensions, there is an ideal, maximally positive quantity (i.e., a subgoal) that is framed by two spectra of increasingly less positive / more negative quantities, namely quantities that are increasingly “too little” on one side of the subgoal and “too much” on the other side of the subgoal. This assumption is based on the notion of homeostasis, the idea that humans will survive and thrive only if they remain within a certain range of quantities on certain physical dimensions (e.g., light, sound, touch, smell, taste, all kinds of organ functions and blood values, sleep and waking, motion and rest, height, weight; Bernard, trans. 1974; Cannon, 1926). Stimuli that meet the subgoal on all dimensions meet the overall goal or, in other words, cannot get better. As illustrated in Figure 1, I model the overall goal at the sphere’s center. Thus, stimuli that cannot get better all lie exactly at the sphere’s center, which means that due to their maximal proximity their similarity is maximal, too. Importantly, stimuli can always get worse by deviating further from the sphere’s center, becoming less positive (green layer) to neutral (thin white layer) to more negative (gray layer) on one or more or even all dimensions. The further from the sphere’s center stimuli are located (i.e., the more negative stimuli are), the more dissimilar they will, on average, be, as the maximum possible distance between stimuli increases as a function of their distance from the sphere's center. In other words, higher similarity of positive compared to negative stimuli (Unkelbach et al., 2008; Unkelbach, 2012) is built into this homeostatic valence model.

Figure 1



Note. The homeostatic model of valence.

Figure 1 illustrates a case with three goal-relevant dimensions (X, Y, and Z) and two positive and two negative stimuli (P1 and P2, N1 and N2, respectively). The stimuli's affective-motivational potential is given by their beeline distance to the thin white neutral layer; the further away stimuli are from the neutral layer, the more positive or negative they are. As negative compared to positive stimuli can in principle be further away from the neutral layer, negative stimuli will, on average, have greater affective-motivational potential (Baumeister et al., 2001; Rozin & Royzman, 2001; Taylor, 1991). That is, valence asymmetry in affective-motivational potential is built into the homeostatic valence model, too.

Can valence asymmetry in stimulus similarity occur independently of valence asymmetry in affective-motivational potential in this homeostatic valence model? As depicted in Figure 1, P1 and P2 have the same affective potential as N1 and N2, because the beeline distance between these four stimuli and the neutral layer is equal. Nevertheless, the beeline distance between P1 and P2 is shorter than the beeline distance between N1 and N2. As interstimulus similarity is given by interstimulus proximity, the positive stimuli P1 and P2 are more similar to one another than the negative stimuli N1 and N2, even though all stimuli's affective-motivational potential is equal. That is, in the homeostatic valence model higher similarity of positive compared to negative stimuli (Unkelbach et al., 2008) does not necessitate higher affective-motivational potential of negative compared to positive stimuli (Baumeister et al., 2001). Thus, if the homeostatic valence model is correct, valence asymmetry in stimulus similarity might be an explanation of valence asymmetry in cognitive processing that is independent of, and thus complementary to, the traditional explanation in terms of valence asymmetry in affective-motivational potential.

Is there evidence that the homeostatic valence model is correct? The core idea of the homeostatic valence model is that on most if not all evaluatively relevant content dimensions positive, adequate quantities are framed by two ranges of negative quantities, namely increasing deficiency or "too little" and increasing excess or "too much". In other words, to fail at falsifying the homeostatic valence model, it is crucial to show that the relation between evaluatively relevant content quantities and (negative to positive) evaluation is curvilinear in the sense that extreme and intermediate quantities are (perceived as) negative and positive, respectively. Facial beauty is a prominent example for the homeostatic distribution (i.e., too little-adequate-too much) of evaluatively relevant content quantities: unattractive faces have too small or too big eyes or lips, or the distance between the ears or chin and hairline is too small or too big, or the skin is too dry or too oily or too light or too dark. Attractive faces, however, have none of these: they are similarly average in a physical sense. Thus, they all look alike (Langlois & Roggman, 1990; Potter et al., 2007; Rhodes, 2006), representing one of many instances of the proposed higher similarity of positive compared to negative stimuli (Alves, Koch, & Unkelbach, 2016; Koch et al., 2016a; Unkelbach et al., 2008). Likewise, people can be too skinny or too heavy. Some people suffer from being too small. Others have problems being too tall, whereas average heighted people rarely ever think about their height. Too low amounts of stress and adversity have a negative impact on health and well-being, but the same is true for too high amounts of stress and adversity (Seery, 2011). Motivation and performance collapse if demand and arousal are too low, but the same is true for too high demand and arousal (Csikszentmihályi, 1990; Yerkes & Dodson, 1908). Potential partners, friends, and colleagues are negative if they are too conservative or too progressive, or too stupid or to smart, or too introverted or too extraverted (Barry & Stewart, 1997). All these and many other evaluatively relevant content dimensions, including openness, experience, complexity (Janssen, 2001; Sturman, 2003), practice, persistence, optimism, self-efficacy, self-esteem (Baumeister, Campbell, Krueger, & Vohs, 2003; Berman, Down, & Hill, 2002; Brown & Marshall, 2001; Wrosch, Scheier, Carver, & Schulz, 2007), generosity, empathy (Flynn, 2003; Galinsky, Maddux, Gilin, & White, 2008; Gino & Pierce, 2009; Windsor, Anstey, & Rodgers, 2008), commitment, and assertiveness (Ames & Flynn, 2007; Somers & Casal, 1994) feature the homeostatic distribution of quantities, namely a positive middle ground between two negative extremes (Grant & Schwartz, 2011). Thus, there is evidence suggesting that the homeostatic valence model is correct, supporting the claims that valence asymmetry in stimulus similarity can occur independently of valence asymmetry in affective-motivational potential, and that valence asymmetry in stimulus similarity might thus explain valence asymmetry in cognitive processing independently of valence asymmetry in affective-motivational potential.

Chapter 2 provides empirical evidence in support of these claims. First, Studies 1, 2, 4, and 6 in Chapter 2 showed valence asymmetry in stimulus similarity in the absence of valence asymmetry in affective-motivational potential (see Footnote 10 in the General Discussion in Chapter 2; we did not measure affective-motivational potential in Studies 3 and 5). And second, in Study 1 in Chapter 2 stimulus similarity predicted five different stages of cognitive processing (classification and evaluation speed, probability of being categorized, and recognition sensitivity and response bias; see Table 3 in Study 1 in Chapter 2) for which valence asymmetry has been shown before (Alves et al., 2015; Unkelbach et al., 2008).

Valence asymmetry in stimulus similarity as a powerful and appealing explanation of valence asymmetry in cognitive processing

According to Fiedler (2014; see also Fiedler & Wänke, 2009), intrapsychic causes such as stimuli’s affective-motivational impact (i.e., positive or negative affect paired with an intention to approach or avoid a positive or negative stimulus) are often not suitable as an explanation of intrapsychic effects such as different stages of cognitive processing, because intrapsychic causes tend to be theoretically too proximal to intrapsychic effects. The theoretical proximity between a cause and an effect of interest is maximal if cause and effect are in fact the same construct. Obviously, in such a case the explanation is tautological or, in other words, has no power at all. The power and appeal of an explanation thus increases as a function of the theoretical distance between cause and effect. Fiedler (2014) argues that research should explain intrapsychic effects by means of causes that lie outside the mind. For example, explanations of affect, motivation, cognition, and behavior in terms of structural or functional aspects of the brain and body (e.g., neurotransmitter's and hormone's production, circulation, and resorption, gain and loss of neural networks, lesions, injuries, all sorts of blood values, organ conditions, genes etc.) are powerful and appealing, because they are theoretically distal. Even more theoretically distal and thus more powerful and appealing are explanations that lie outside the organism – that is, ecological explanations.

Ecological explanations of intrapsychic effects are explanations that relate to the way in which information is distributed in, and sampled from, people's ecology. For example, it is safe to assume that people encounter more information about their ingroup compared to outgroups, because they spend more time with ingroup compared to outgroup members. This ingroup-outgroups difference in information sample size provides a powerful and appealing ecological explanation of intergroup bias (i.e., people's tendency to favor the ingroup, and to disfavor outgroups). The greater the information sample size, the better ratios of positive to negative are learned (Fiedler, 2014). If a person encounters a ratio of hundred positive to twenty negative ingroup behaviors, and a ratio of only ten positive to two negative outgroup behaviors, he or she will learn his or her ingroup's behavioral positivity prevalence better than his her outgroups' behavioral positivity prevalence, resulting in a more favorable evaluation and treatment of the ingroup compared to outgroups (i.e., intergroup bias). Importantly, this intrapsychic effect occurs in the absence of real differences in behavioral positivity prevalence between the ingroup and outgroups. Even more importantly, the intrapsychic effect's ecological explanation in terms of a ingroup-outgroups difference in information sample size is theoretically more distal and thus more powerful and appealing than the widely accepted intrapsychic explanation in terms of a positive distinctiveness or self-enhancement motive (Tajfel & Turner, 1979). Critically speaking, this intrapsychic explanation traces people's perception of the ingroup as superior to outgroups to their motive to perceive the ingroup and thus themselves as superior to outgroups and thus other people. That is, critically speaking this intrapsychic explanation is tautological / has no power.

Likewise, the intrapsychic explanation of valence asymmetry in cognitive processing in terms of valence asymmetry in affective-motivational potential is theoretically proximal and thus not very powerful / appealing. Specifically, valence asymmetry in affective-motivational potential can be paraphrased as higher organismic importance of negative compared to positive stimuli. Thus, this intrapsychic explanation traces higher processing accuracy (e.g., Alves et al., 2015; Fazio, Eiser, & Shook, 2004) and lower processing speed (e.g., Unkelbach et al., 2008) of negative compared to positive stimuli to higher organismic importance of negative compared to positive stimuli – that is, a construct that is theoretically proximal and thus hardly insightful / interesting as a cause. In contrast, the ecological explanation in terms of higher similarity of positive compared to negative stimuli is theoretically more distal and thus more insightful / interesting.

But is valence asymmetry in stimulus similarity not a cognitive and thus intrapsychic, theoretically proximal cause, too? I propose that valence asymmetry in stimulus similarity is a theoretically distal cause that is rooted in people’s information ecology. Ecological similarity is commonly measured in terms of frequency of co-occurrence in time and space (Griffiths, Steyvers, & Tenenbaum, 2007; Jones & Mewhort, 2007). For example, two stimuli are ecologically the more similar the more often they co-occur on webpages accessible through Google Search (henceforth: Google similarity; Cilibrasi & Vitanyi, 2007), or in passages of a large collection of books that is representative of the general knowledge of undergraduates (LSA similarity; Landauer & Dumais, 1997). If valence asymmetry in stimulus similarity is indeed rooted in people’s information ecology, the Google and LSA similarity of positive compared to negative stimuli should be higher. Study 1 in Chapter 2 showed that this is the case for a widely investigated set of valenced words (e.g., Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Unkelbach et al., 2008). Further, if valence asymmetry in similarity is indeed an ecological, theoretically distal cause of valence asymmetry in cognitive processing, Google and LSA similarity should predict different stages of cognitive processing for which valence asymmetry has been shown before. Study 1 in Chapter 2 showed that this is the case for classification and evaluation speed as well as recognition sensitivity and response bias (see Table 3 in Study 1 in Chapter 2), establishing that valence asymmetry in stimulus similarity is an ecological, theoretically distal and thus powerful and appealing explanation of valence asymmetry in cognitive processing.

On a side note, Study 1 in Chapter 2 also showed that Google and LSA similarity (i.e., ecological similarity) substantially correspond to two cognitive measures of similarity (SpAM and Pairwise similarity), that valence asymmetry is also found for these cognitive measures, and that these measures also predict the aforementioned stages of cognitive processing (see Tables 2 and 3 in Study 1 in Chapter 2).

Valence asymmetry in stimulus similarity as a robust and general explanation of valence asymmetry in cognitive processing

Up to this point, I have argued that valence asymmetry in stimulus similarity is a novel, independent, powerful, and appealing explanation of valence asymmetry in cognitive processing. However, these theoretical advantages are not of any practical use if valence asymmetry in stimulus similarity is not a robust and general phenomenon that thus can be shown reliably for large and representative sets of positive and negative stimuli. Is there evidence that the proposed higher similarity of positive compared to negative stimuli (Unkelbach et al., 2008; Unkelbach, 2012) is a robust and general phenomenon? I propose that this is the case. Specifically, there are several broad and relevant stimulus domains in which valence asymmetry in stimulus similarity has been shown: emotions, faces, persons, social groups (see Chapter 3), and, more generally, verbal and visual stimuli (see Chapter 2).

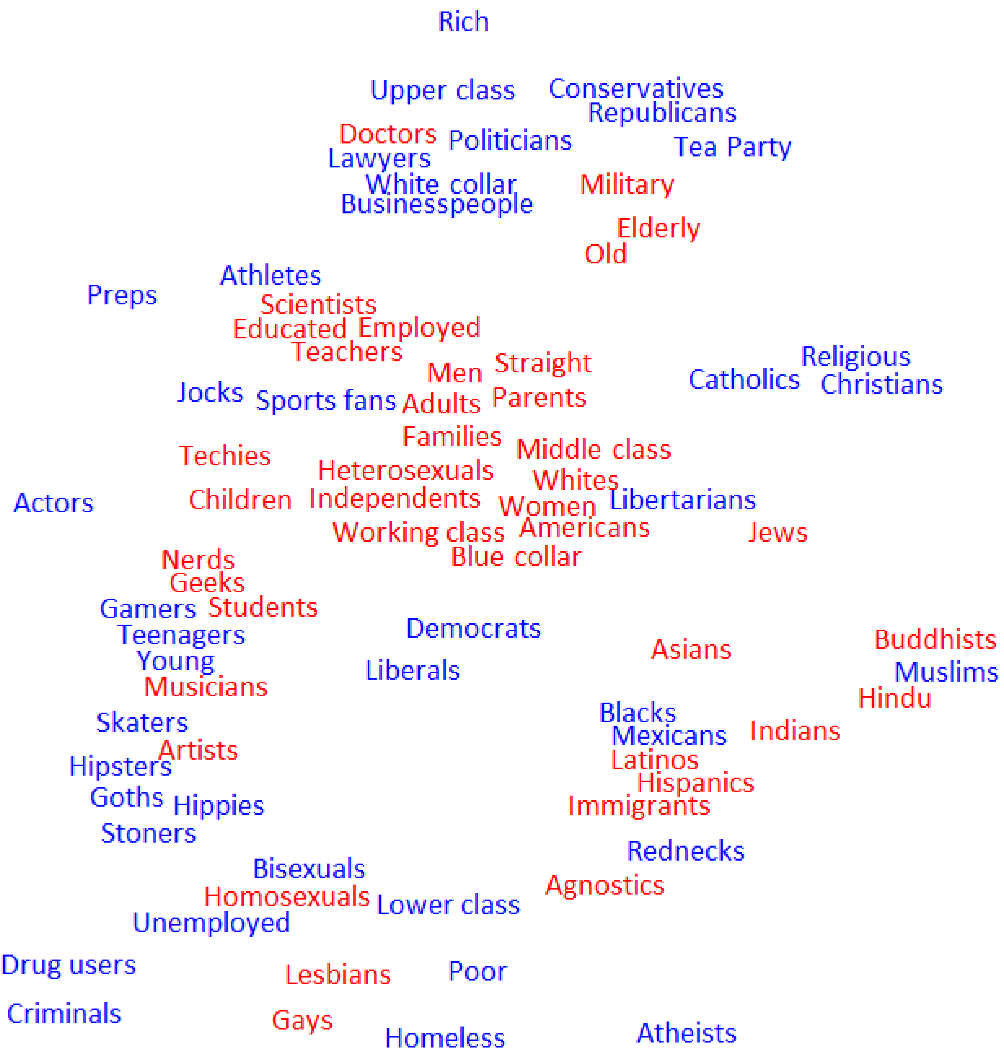
Within the dominant emotion classification systems, there are less positive than negative emotions. In English, German, and Spanish the spectrum of words for positive emotional states is less diverse than the vocabulary for negative emotional states (Schrauf & Sanchez, 2004; Semin & Fiedler, 1992). Within the six basic emotions proposed by Ekman and Friesen, (1971), happiness is the only positive emotion, while on the negative side, there are anger, disgust, fear, and sadness. Surprise has no clear valence, but according to Noordewier and Breugelmans (2013), surprise is more often negative than positive. Plutchik's (2001) classification contains trust and joy versus anger, disgust, fear, and sadness. And Izard's (1971) list contains interest and joy versus anger, contempt, disgust, distress, fear, guilt, and shame. While there are appeals to better differentiate positive emotions (Sauter, 2010), we could not find published evidence for more diversity on the positive side of emotions compared to the negative side. Greater diversity can be taken as greater dissimilarity, as the number of categories (i.e., the diversity) required to organize a stimulus domain increases as a function of the dissimilarity between the domain's stimuli.

Concerning faces, averaging more and more faces results in a face that is more and more attractive, and individual faces with average features are also more attractive than faces with features that deviate from the prototype of the population. Consequently, attractive faces all look alike, while there are many ways in which faces can be unattractive (Langlois & Roggman, 1990; Rhodes, 2006). Faces with symmetrical halves are also more attractive than faces whose left side does not resemble the right side (Rhodes, Proffitt, Grady, & Sumich, 1998; Grammer & Thornhill, 1994). And indeed, if cognitive similarity is directly assessed (i.e., not in terms of averageness or symmetry), participants judge attractive compared to unattractive faces to be more similar to one another (Potter et al., 2007). Also, average / prototypical faces are also more trustworthy (Sofer, Dotsch, Wigboldus, & Todorov, 2015). In sum, if one accepts facial attractiveness and trustworthiness as proxies for facial positivity, positive faces are more similar to one another than negative faces.

Concerning persons, Leising, Ostrovski, and Borkenau (2012) reported that people describe liked persons much less differentiated compared to disliked persons. This conclusion was based on the higher frequency of unique terms used to describe disliked persons. The other way round, Bruckmüller and Abele (2013) kept the number of terms describing persons constant and directly assessed the cognitive similarity of these terms. Participants judged 20 character traits related to the basic dimensions of person perception (agency / competence and communion / warmth), and they judged the traits' positive versions (e.g., clever, confident, warm, and friendly) to be more similar to one another than the traits' negative versions (e.g., cold, mean, stupid, and insecure). Also, Alves and colleagues (2016) showed in seven studies that people perceived people they like (e.g., friends) as more similar to one another than people they dislike.

The studies reported in Chapter 3 attested that the same is true for groups (Koch, Imhoff, Dotsch, Unkelbach, & Alves, 2016). Based on either pairwise similarity ratings – the classical method to measure cognitive similarity – or spatially arranged similarity – a new, more efficient method in which people simultaneously rate the similarity of all stimuli of interest by arranging more similar stimuli closer to one another on the screen (for detailed discussions of this spatial arrangement method, see Goldstone, 2014; Hout & Goldinger, 2016; Hout, Goldinger, & Ferguson, 2013; Koch et al., 2016a; Verheyen, Voorspoels, Vanpaemel, & Storms, 2016) – we scaled well-fitting two-dimensional similarity spaces (for a review of multidimensional scaling, see Borg & Groenen, 2005) for three different samples of U.S. and German groups based on either frequency of naming (e.g., "Off the top of your head, what various types of people do you think today's society categorizes into groups?"; Fiske, Cuddy, Glick, & Xu, 2002) or frequency of occurrence in contemporary mass media (Davies, 2011). Just like in the homeostatic valence model, in these spaces groups are points at certain locations, and intergroup similarity is given by intergroup proximity. Thus, if valence asymmetry in stimulus similarity also applies for groups, more warm, trustworthy, and liked (i.e., more communal) groups should be located closer to other groups.

Figure 2



Note. One of the group spaces that we scaled.

Figure 2 shows the group space that we scaled in Study 2 in Chapter 3. The 40 most and 40 least communal groups are marked in red and blue, respectively. Evidently, the proximity of the 40 most communal groups to other groups was higher than the proximity of the 40 least communal groups to other groups. Accordingly, more communal groups were more proximal and thus more similar to other groups, confirming that valence asymmetry in stimulus similarity also applies for groups. In fact, this was the case for every single group space that we scaled in Chapter 3 (see the General Discussion in Chapter 3).

Finally, concerning verbal stimuli, a domain that is much broader than words that denote only just emotions, traits, persons, or groups, Unkelbach and colleagues (2008) did a similarity scaling of 20 extremely positive and 20 extremely negative words of all kinds (i.e., objects, persons, events etc.) sampled from a widely investigated set (Fazio et al., 1986). The resulting space showed that the positive words were more densely clustered – that is, more similar to one another – than the negative words. However, 40 words are hardly enough to generalize valence asymmetry in stimulus similarity to all positive and negative words (Wells & Windschitl, 1999; Westfall, Judd, & Kenny, 2015; Westfall, Kenny, & Judd, 2014). The stimulus sample examined by Unkelbach and colleagues (2008) presumably did not exceed $N = 40$ because the people in their study judged the words' similarity in pairs, a procedure that takes a lot of time and effort. To illustrate, for to obtain similarity estimates for each pair that can be formed with a set of 40 words, 780 judgments have to be made. Is there a more time- and effort-efficient measure that allows testing the robustness and generality of valence asymmetry in stimulus similarity with a large variety of positive and negative words?

Recently, Hout and colleagues (2013) validated such a similarity measure: the spatial arrangement method (SpAM; see also Goldstone, 1994; Kriegeskorte & Mur, 2012). Goldstone (1994) was the first to measure perceptual similarity based on how close to one another stimuli were arranged on the screen. The advantage of SpAM is that arranging a stimulus simultaneously readjusts the proximity and thus rated similarity between that stimulus and all other stimuli on the screen, decreasing measurement speed and effort (another advantage is that all stimuli can be rearranged at all times during the task). In Goldstone's study (1994), the similarity proximities between the spatially arranged stimuli (i.e., the capital letter A in different fonts) correlated almost perfectly their similarity judged in pairs, suggesting that SpAM is an effective way to measure perceptual similarity. Hout and colleagues (2013) generalized this from perceptual similarity within a stimulus domain (i.e., schematic wheels and rudimentary bugs) to conceptual similarity within a stimulus domain (i.e., animal names). To validate SpAM (Hout et al., 2013) as an effective method to measure the similarity of conceptual stimuli from different domains, Study 1 in Chapter 2 correlated the SpAM similarity of Unkelbach and colleagues' (2008) 40 conceptually diverse words with their similarity judged in pairs (Pairwise similarity), and with their frequency of co-occurrence on webpages (Google similarity) and in the print media (LSA similarity; see Table 2 in Chapter 2). Additionally, Study 1 in Chapter 2 correlated the 40 words' SpAM, Pairwise, Google, and

LSA similarity with their evaluation and classification speed, recognition response bias and sensitivity, and probability of being categorized (see Table 3 in Chapter 2). Results showed that the construct and predictive validity of SpAM similarity is high, which encouraged us to use SpAM to test the robustness and generality of valence asymmetry in stimulus similarity with a large variety of positive and negative words.

In Study 2a in Chapter 2, people first generated and then spatially arranged 20 positive and 20 negative words. This procedure delivered a large and representative sample of positive and negative stimuli (N = 1,044). To avoid retrieval biases, Study 2b had participants spatially arrange 20 positive and 20 negative stimuli generated by other participants in Study 2a. Study 3 then examined whether the proposed asymmetry in similarity holds true for stimuli of both consensual and idiosyncratic valence; participants generated and spatially arranged 40 words that are positive / negative either generally (i.e., for everybody; N = 995) or personally (i.e., for themselves; N = 1,139). In Study 4, to take greater care of avoiding retrieval biases, participants spatially arranged 20 positive and 20 negative words randomly drawn from a pool (N = 64) to which other participants had added only one positive and only one negative word each. Study 5 shifted from investigating memory-based information to investigating experience-based information. Participants named one positive and one negative event of their day on seven consecutive days. Thereafter, they spatially arranged these unique everyday life events from their last week (N = 1,518). In all these studies, people spatially arranged the positive words closer to another than the negative ones, establishing that valence asymmetry in similarity holds true for all kinds of verbal stimuli.

Finally, Study 6 in Chapter 2 switched from examining strongly to examining strongly, moderately, and mildly positive / negative words, and from examining verbal to examining visual stimuli. Specifically, we compared the similarity of all positive and negative words in the database by Warriner, Kuperman, & Brysbaert (2013; N = 13,915), and all positive and negative pictures in the international affective picture system (Lang, Bradley, & Cuthbert, 2005; N = 956). Results confirmed that valence asymmetry in similarity holds true for all kinds of visual stimuli, too. An open research question is whether the phenomenon can be shown for auditive, olfactory, and haptic stimuli, too.

I summarize Chapter 1 as follows: valence asymmetry in stimulus similarity can be taken as a novel, independent, powerful, appealing, robust, and general explanation of valence asymmetry in cognitive processing. Chapter 2 and 3 elaborate on the empirical robustness and generality of the theoretically predicted higher similarity of positive compared to negative stimuli, showing the effect for thousands of words and pictures (Chapter 2), and for hundreds of groups (Chapter 3). Thus, redundancy is inevitable.

Chapter 2 – A general valence asymmetry in similarity: Good is more alike than bad

Abstract

The density hypothesis (Unkelbach et al., 2008) claims a general higher similarity of positive information to other positive information compared to the similarity of negative information to other negative information. This similarity asymmetry might explain valence asymmetries on all levels of cognitive processing. The available empirical evidence for this general valence asymmetry in similarity suffers from a lack of direct tests, low representativeness, and possible confounding variables (e.g., differential valence intensity, frequency, familiarity, or concreteness of positive and negative stimuli). To address these problems, Study 1 first validated the spatial arrangement method (SpAM) as a similarity measure. Using SpAM, studies 2-6 found the proposed valence asymmetry in large, representative samples of self- and other-generated words (Studies 2a/2b), for words of consensual and idiosyncratic valence (Study 3), for words from one and many independent information sources (Study 4), for real-life experiences (Study 5), and for large data sets of verbal (i.e., ~14,000 words reported by Warriner, Kuperman, & Brysbaert, 2013) and visual information (i.e., ~1,000 pictures reported in the IAPS; Lang, Bradley, & Cuthbert, 2005) (Study 6). Together, these data support a general valence asymmetry in similarity, namely that good is more alike than bad.

Positively and negatively evaluated information differentially influences all stages of information processing; from attention (e.g., Pratto & John, 1991) to categorization (e.g., Billig & Tajfel, 1973) to memory (e.g., Alves et al., 2015). Traditionally, these influences are explained by negative information's higher emotional and motivational significance due to basic survival needs (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001; Taylor, 1991). While emotional and motivational effects are uncontested, Unkelbach and colleagues (2008) formulated the *density hypothesis* as an informational explanation of valence asymmetries in information processing. The density hypothesis states that positive information is generally more similar to other positive information compared to negative information's similarity to other negative information; in visualizations of mental representations, positive information is thus more densely clustered. In other words, they hypothesized a *general valence asymmetry in stimulus similarity*. And as inter-stimulus similarity influences numerous cognitive processes, such as classification, categorization, generalization, judgment, recognition, and recall (e.g., Hintzman, 1988; Nosofsky, 1986; 1988; Shepard, 1987), the hypothesized asymmetry in stimulus similarity might explain many valence asymmetries in cognitive processing independently of information's emotional and motivational potential. But does positive and negative information really differ in similarity?

The present article tests the hypothesis of a general valence asymmetry in stimulus similarity. To do so, we first review existing valence asymmetries in cognitive processing that the proposed asymmetry might explain. Next, we provide a theoretical explanation for the density hypothesis (Unkelbach, 2012; Unkelbach et al., 2008). Then, we review existing evidence for the proposed generality of higher similarity of positive compared to negative information, concluding that the available evidence suffers from a lack of direct tests, low representativeness, and possible confounding variables, namely valence intensity, frequency, familiarity, and concreteness. Addressing these problems, six studies provide converging evidence that, as predicted by the density hypothesis and its underlying theoretical rationale, across all kinds of positive and negative information, positive information is more similar compared to negative information and therefore clusters more densely in in visualizations of mental representations.

Why we should care that positive stimuli are more similar than negative stimuli

Despite the debate over the use of similarity as a construct (Goodman, 1972) and different models of similarity (Goldstone & Son, 2005; Hahn, Chater, & Richardson, 2003; Krumhansl, 1978; Medin, Goldstone, & Gentner, 1993; Tversky, 1977), inter-stimulus similarity undoubtedly impacts learning, memory, and cognition in profound ways ("sameness is […] the backbone of our thinking", James, 1980/1950; p.459). Targets are classified faster/easier when following similar primes (e.g., McRae & Boisvert, 1998; Perea, Duñabeitia, & Carreiras, 2008). Categorization is more accurate when stimuli are similar to the representative prototype and/or available exemplars of the target category (Nosoksky, 1986; 1988; Smith & Sloman, 1994). Generalizations of processing strategies, judgments, and decisions to similar stimuli are more likely (Ames, 2004; Shepard; 1987; Tenenbaum & Griffiths, 2001). Dissimilar targets are recognized with greater accuracy (DeSoto & Roediger, 2014; Hintzman, 1988). And, similar items are more likely to be recalled (Roediger & McDermott, 1995; Schwartz & Humphreys, 1973). In sum, inter-stimulus similarity plays a major role in how people make sense of the world (Quine, 1969).

If, as predicted by the density hypothesis (Unkelbach, 2012; Unkelbach et al., 2008), positive stimuli are generally more similar than negative stimuli, there should be similarity-based differences between positive and negative stimuli on all levels of information processing (Unkelbach, 2012). There is evidence to suggest that this is indeed the case. For example, positive words are classified faster (Fazio, Sabonmatsu, Powell, & Kardes, 1986; Klauer & Musch, 1999), because they are more similar to one another than negative words (Unkelbach et al., 2008; extremity of positive/negative word meaning and word frequency and length were ruled out as confounding variables). Similarly, people recognize negative information more accurately, because it is more dissimilar (Alves et al., 2015; word frequency and length were ruled out as confounding variables). Furthermore, generalizations of positive evaluations (i.e., halo effects; Dion, Berscheid, & Walster, 1972; Langlois et al., 2000) are more likely than generalizations of negative evaluations (horn effects), and this has been argued to be due to that positive aspects of objects, people, and events are more similar and thus more relatable to one another than negative stimuli (Gräf & Unkelbach, 2015).

Furthermore, negative behaviors might impact impression formation more strongly (Knobe, 2003; Peeters & Czapinski, 1990), because such behaviors are more dissimilar to the behavioral norm (Skowronski & Carlston, 1989). And, people might recall negative events with greater precision for details (Kensinger, Garoff-Eaton, & Schacter, 2006), because they are more dissimilar to one another than positive events. In sum, there is evidence for similarity-based valence asymmetries in cognitive processing (e.g., processing speed, likelihood of generalization, and memory accuracy), and there are many more possible candidates to be explained by the assumed differential similarity of positive and negative information.

Why positive information should be more similar than negative information

There are two arguments for the proposed similarity asymmetry: stimulus range and frequency of stimulus co-occurrence. Both arguments are based on the structure of people's information ecology, and not the biased processing of evaluative information (for reviews on why and how to distinguish between ecological and psychological analyses, see De Houwer, Gawronski, & Barnes-Holmes, 2014; Fiedler & Wänke, 2009).

The *stimulus range* argument is that on any given content dimensions, positive states are framed by "too much" and "too little", resulting in a narrower multidimensional range of positive compared to negative states; in fact, we were not able to think of a single content dimension on which both "more" and "less" are better than the range in between those two boundaries – that is, a dimension on which negative states are framed by positive states. Thus, as long as there are two ways to be negative on a given dimension, the higher similarity of positive information follows as a necessity. For example, a dinner meal can be evaluated on several content dimensions (e.g., temperature level, taste intensity, and nutritional value). For each dimension, a deviation of "too much" (e.g., too hot, too spicy, and too fat) or "too little" (e.g., too cold, too bland, and too thin) will make the dinner negative. In contrast, positive meals will all have a similarly adequate temperature level, taste intensity, and nutritional value, making them more similar within the space constituted by the relevant content dimensions.

The stimulus range argument thereby reflects the notion of homeostasis ("staying similar" in Greek), the idea that organisms are in a positive state only if they remain within certain boundaries on certain physical dimensions (e.g., light, sound, touch, smell, taste, all kinds of organ functions and blood values, sleep and waking, motion and rest, height, weight; Bernard, trans. 1974; Cannon, 1926).

The *frequency of stimulus co-occurrence* argument is that positive compared to negative information is more frequent both objectively (i.e., in large text corpora; Augustine, Mehl, & Larsen, 2011; Rozin, Berman, & Royzman, 2010) and subjectively (Unkelbach et al., 2010). In their Pollyanna hypothesis Matlin and Stang (1974) argued that this positivity bias might reflect a basic human tendency, and it follows from many other psychological principles, such as the need to maintain favorable evaluations, relatedness, belonging to others (Langston, 1994; Reis et al., 2010), good mood, and life satisfaction (Lyubomirsky, Sousa, & Dickerhoof, 2006). In addition, most people most often comply with social norms to act and interact in a positive way to avert reputation damage (Feinberg, Willer, & Schultz, 2014), resource deprivation (Balliet, Mulder, & van Lange, 2011), and ostracism (Williams, 2007). This valence asymmetry in frequency is further amplified by the fact that people keep positive and negative objects, people, and events near and at distance, respectively (Denrell, 2005; Fazio, Eiser, & Shook, 2004).

Positive information's higher frequency must lead to higher frequency of co-occurrence, which translates to higher perceived similarity, because stimuli that co-occur more often are more strongly associated in memory (Fiedler, Kutzner, & Vogel, 2013; Verhaeghen, Aikman, & van Gulick, 2011); frequency of co-occurrence per se is a widely accepted proxy for inter-stimulus similarity (e.g., Griffiths, Steyvers, & Tenenbaum, 2007; Jones & Mewhort, 2007; Landauer & Dumais, 1997).

Taken together, these two arguments predict a general higher similarity of positive compared to negative information in the ecology, because the range of positive information is more restricted and positive information co-occurs more frequently; a more detailed discussion of these two explanations is provided by Koch, Alves, and Unkelbach (2016). However, the main focus of the present research is the generality of the predicted valence asymmetry in similarity, and not the underlying theoretical explanations. Thus, we now turn to the available evidence.

Existing evidence that positive information is more similar than negative information

Indirect evidence. Three research areas indirectly support a general valence asymmetry in stimulus similarity: facial beauty, basic emotions, and vocabulary for states and traits.

First, morphed (i.e., averaged) and naturally average faces are more attractive than less prototypical faces. Because average/prototypical faces are highly similar, "attractive faces all look alike", while faces are unattractive in many different ways (too big or too small eyes or lips, *or* the distance between the ears or between chin and hairline is too big or too small, *or* the skin is too dry or too oily or too light or too dark; Langlois & Roggman, 1990, p. 115; Rhodes, 2006). Potter, Corneille, Ruys, and Rhodes (2007) directly tested this prediction and found that in a multidimensional scaling of attractive and unattractive faces people judged attractive faces as more similar to one another compared to unattractive faces.

Second, the diversity of positive "basic" emotions is lower than the diversity of their negative counterparts (for a critique of basic emotions, see Ortony & Turner, 1990). In the taxonomy of basic emotions by Ekman and Friesen (1971), there is only happiness on the positive side, while anger, disgust, fear, and sadness form the more diverse negative side; according to Noordewier and Breugelmans (2013), the valence of surprise is ambivalent but more often negative rather than positive. Furthermore, this higher diversity of negative compared to positive basic emotions is also apparent in Plutchik's (2001; anger, disgust, fear, and sadness vs. trust and joy) and Izard's (1971; anger, contempt, disgust, distress, fear, guilt, and shame vs. joy) taxonomies. There are appeals to better differentiate positive emotions

(Sauter, 2010); however, we could not find published evidence for a greater diversity of positive compared to negative emotions.

And third, in English, German, and Spanish the spectrum of words for positive emotional states and character traits is less diverse than the vocabulary for negative emotional states and character traits (Leising, Ostrovski, & Borkenau, 2012; Schrauf & Sanchez, 2004; Semin & Fiedler, 1992).

However, diversity is not a direct measure of inter-stimulus similarity, and thus studies that compare the diversity of positive and negative terms do not provide direct evidence that positive information is mentally represented as more similar to one another than negative information (e.g., competent and warm are less diverse but less similar than untrustworthy, dishonest, and insincere).

Direct evidence. Two studies provide direct evidence: Bruckmüller and Abele (2013) showed that 20 character traits related to agency and communion were judged to be more similar to one another in their positive formulations (e.g., warm, friendly, clever, and confident) than their negative formulations (e.g., cold, mean, stupid, and insecure). In Unkelbach and colleagues' (2008) study, participants used a 'dissimilar-similar' scale to judge all 780 pairs of words that can be formed with their set of 20 extremely positive words and 20 extremely negative words that refer to not just people, but also objects and events. These similarity judgments were averaged across participants and subjected to multidimensional scaling (MDS; Borg & Groenen, 2005). The MDS algorithm estimates coordinates for each word in a geometric space in which proximity equates to similarity. Finally, in this geometric space the authors compared the average proximity of the positive words to the average proximity of the negative ones. Consistent with their *density hypothesis*, the positive words were more densely clustered – that is, more similar to one another – compared to the negative words.

However, these direct tests are restricted in scope. Bruckmüller and Abele (2013) only used 20 traits specifically describing communion or agency, and Unkelbach and colleagues (2008) used only the most extremely positive and negative words from a list of 92 words (Klauer & Musch, 1999), which was originally compiled in an arbitrary fashion (Fazio et al., 1986). Thus, there is a chance that due to a sampling bias, this list consists of similar positive words and dissimilar negative words. Following the arguments by Westfall, Judd, and Kenny (2015), this small sample of stimuli does not provide the necessary power to generalize to the population of positive and negative information. Small samples of participants do not allow generalizing, and small samples of stimuli do not allow this either. The following empirical investigation aims to fill this gap.

Testing a general valence asymmetry in stimulus similarity

The solution for the discussed limitations is to collect similarity data for large samples of freely selected positive and negative stimuli. However, the standard procedure, pairwise similarity judgments to feed an MDS algorithm, prohibits large stimulus samples due to high numbers of repetitive trials. For example, scaling 40 stimuli requires 780 similarity comparisons – if one, for example, wants pairwise similarity judgments for 20 samples of 40 words, 15,600 pairwise similarity judgments must be made. Thus, testing a general valence asymmetry in stimulus similarity and its possible predictive power for cognitive processes necessitates another method of measuring inter-stimulus similarity.

An early alternative to avoid the efforts of pairwise judgment was that participants sort similar and different stimuli into same and different piles, respectively (e.g., Forgas, 1976; Rosenberg, Nelson, & Vivekananthan, 1968, Shaver, Schwartz, Kirson, & O'Connor, 1987). Sorting is more efficient than pairwise judgment, because each stimulus is sorted into only one pile, whereas in the pairwise method each stimulus is judged in conjunction with each other stimulus. However, sorting is disadvantaged in terms of precision of measurement, because responses between similar (same pile) and different (different piles) are not admitted.

Recently, Hout, Goldinger, and Ferguson (2013) validated a new similarity measurement method. This spatial arrangement method (SpAM; Hout, Goldinger, & Ferguson, 2013; Goldstone, 1994; Kriegeskorte & Mur, 2012) provides a psychometrically effective and highly efficient method to measure the similarity of large samples of stimuli. Goldstone (1994) was the first to measure perceptual stimulus similarity based on how close to one another stimuli were arranged on a computer screen. The averaged proximities between the spatially arranged stimuli (i.e., the capital letter "A" in different fonts) correlated highly with averaged pairwise similarity judgments, $r(62) = .93$, suggesting that SpAM might be an effective way to measure perceptual similarity. Hout and colleagues (2013) generalized this from perceptual similarity within a stimulus domain (i.e., schematic wheels and rudimentary bugs) to conceptual similarity within a stimulus domain (i.e., animal names, $r(23) = .81$ for the animals examined by

Hornberger, Bell, Graham, & Rogers, 2009; *r*(23) = .61 for the animals examined by Henley, 1969).

Overview of the studies

To validate SpAM (Hout et al., 2013) as an effective method to measure the similarity of conceptual stimuli from different domains, Study 1 compared the SpAM similarity of 20 positive and 20 negative conceptually diverse stimulus words (see Unkelbach et al., 2008) with their similarity judged in pairs (Pairwise similarity), and with their frequency of co-occurrence in the internet (Google similarity) and the print media (LSA similarity). Additionally, Study 1 compared the predictive power of SpAM, Pairwise, Google, and LSA similarity by correlating the obtained similarities with stimuli's evaluation speed, classification speed, recognition response bias and sensitivity, and probability of being subsumed under a category.

Having validated SpAM, Study 2a instructed participants to *generate* and spatially arrange 20 positive and 20 negative words. This procedure should deliver a large and representative sample of positive and negative stimuli. To avoid retrieval biases, Study 2b had participants spatially arrange 20 positive and 20 negative stimuli generated by other participants in Study 2a. Study 3 then examined whether the similarity asymmetry holds true for stimuli of both *consensual* and *idiosyncratic* valence; participants named and spatially arranged 40 words that are positive/negative either generally (i.e., for everybody) or personally (i.e., for themselves). In Study 4, to avoid processing and retrieval biases, participants spatially arranged 20 positive and 20 negative words randomly drawn from a pool to which other participants had added only one positive and only one negative word each. Study 5 shifted from investigating *memory-based* information to investigating *experience-based* information. Participants named one positive and one negative event of their day on seven consecutive days. Thereafter, they spatially arranged these unique everyday life events from their last week. Finally, Study 6 switched both from *strongly* to *strongly, moderately, and mildly* positive/negative stimuli and from *verbal* to *visual* stimuli by comparing the similarity of all positive and negative words in the database by Warriner, Kuperman, & Brysbaert (WKB; 2013; ~14,000 words), and all positive and negative pictures in the international affective picture system (IAPS; Lang, Bradley, & Cuthbert, 2005; ~1000 pictures).

In addition to corroborating that the proposed valence asymmetry in similarity is a general phenomenon, Studies 1 and 4 corroborated that the asymmetry is actually due to valence, as the positive stimuli were seen as more similar to one another than the negative stimuli even when controlling for their valence intensity, frequency, familiarity, and concreteness.

Throughout these studies, we report all manipulations, measures, and data exclusions. The reported studies represent the full set we conducted for the present research question. We based our sample sizes on the effect sizes reported by Unkelbach and colleagues (2008).

Study 1

Participants spatially arranged the 20 positive and 20 negative words investigated by Unkelbach and colleagues (2008) and then divided these words into between 2 and 7 unlabeled categories. With these 40 stimulus words, we validated how well SpAM and classical pairwise judgment measure the same aspects of conceptual similarity. Up to this point, the validity of SpAM similarity has only been confirmed for perceptual/conceptually simple stimulus sets such as color patches, letters, letter-like forms, schematic wheels, rudimentary bugs, and animal names (see Goldstone, 1994; Hout et al., 2013; Kriegeskorte & Mur, 2012). To further test and compare the validity of SpAM similarity, we correlated SpAM similarity and similarity judged in pairs with two ecological indicators of the 40 words' inter-stimulus similarity, namely their frequency of pairwise co-occurrence on webpages (as indicated by the most widely used search engine: Google Search; Cilibrasi & Vitanyi, 2007) and in a large collection of book passages that is representative of the literature read by US college students (Latent Semantic Analysis; see Landauer & Dumais, 1999). Finally, to test and compare the predictive strength of SpAM similarity and Pairwise similarity, we correlated these measures with basic aspects of cognitive processing, including the 40 words' evaluation speed (based on data from Klauer & Musch, 1999), their classification speed (based on data from Unkelbach et al., 2008), their probability of being falsely recognized (based on data from Alves et al., 2015), and their probability of being subsumed under a category (present study). As participants spatially arranged the 40 words right before they sorted them into the same or different categories, in contrast to prior research we did not operationalize this sorting into categories as a separate similarity measure, but rather as a possible effect of similarity measured with SpAM.

Method

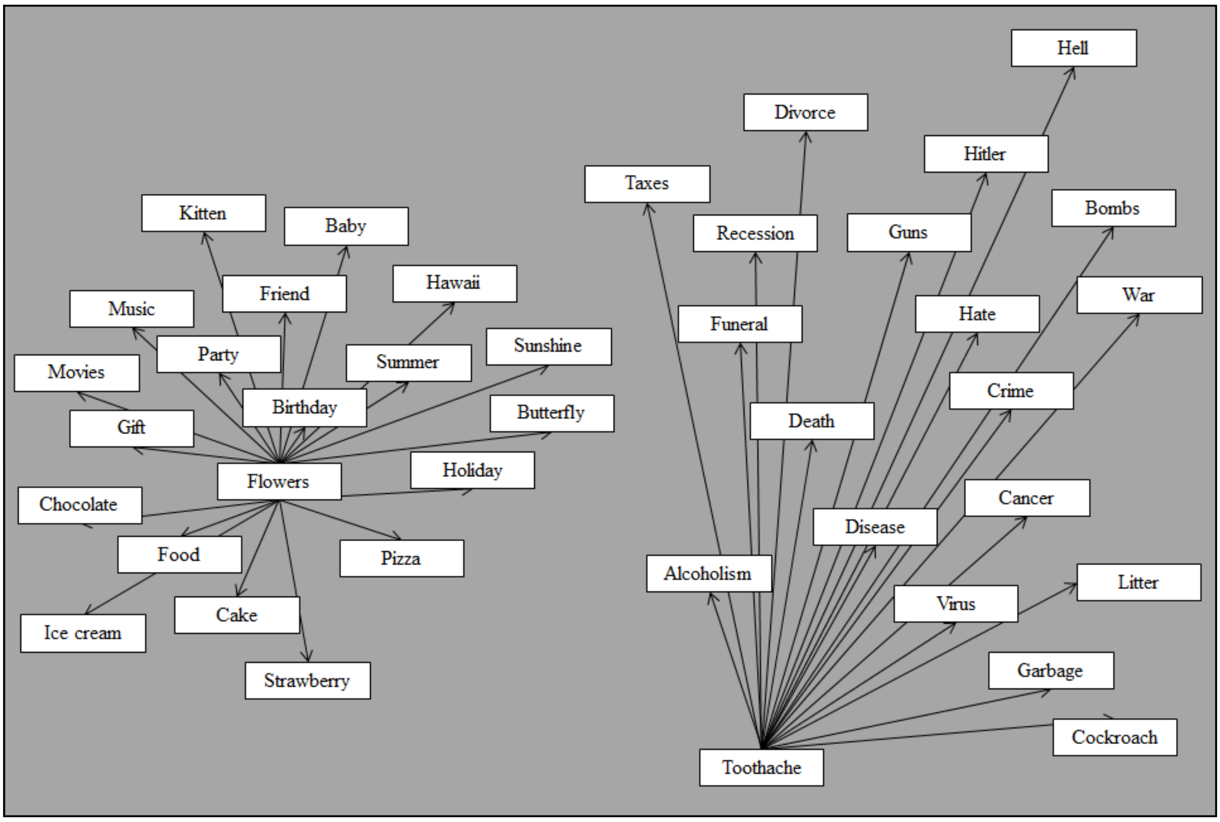
Participants, design, and stimuli. 55 students (40 women, 15 men; 52 native German speakers) participated for course credit. We used the 20 positive and 20 negative words investigated by Unkelbach and colleagues (2008; see Appendix A). These 40 words were first used by Fazio and colleagues (1986), and were translated into German by Klauer & Musch (1999). Stimulus valence varied within-participants.

Procedure. Upon arriving, participants read an informed consent form. If they agreed to participate, experimenters lead them to computer-equipped cubicles and started a Visual Basic program that presented German instructions (translated into English here) and stimuli, and recorded dependent variables. The first screen informed participants that "Your task is to sort 40 words based on how similar/dissimilar they are. The words will appear in the middle of the screen one at a time, and you can drag-and-drop them at any time to change their position on the screen. Please sort the words in such a way that more similar words are more close to one another, while more dissimilar words are further away from one another. That is, your task is to use the 40 words to draw a map in which greater proximity indicates greater similarity, and in which greater distance indicates greater dissimilarity".

The instruction did not mention the evaluative connotation of the stimuli. After clicking on an "I understand" button, the background color of the screen (1920 x 1080 pixels) changed to gray, and a word randomly drawn from the set of 20 positive and 20 negative words appeared in the middle of screen in black font in a white label (100 x 22 pixels) with a black margin. Once participants dragged this word to another location on the screen, a "Next word" button appeared at the bottom of the screen. A click on the button presented the next randomly drawn word in the middle of the screen. At the same time, the button disappeared. Participants repeated this procedure for all 40 words. All words already on the screen could be dragged to another location at all times during the spatial arrangement task. After participants arranged the fortieth word on the stimulus map, an "I have finished" button appeared. With a click, participants ended the spatial arrangement. Figure 1 presents an example for such a stimulus map. The arrows in Figure 1 show the pixel proximities of the stimuli "flowers" and "toothache" to all other stimuli of the same valence. For each of the 40 words, the program computed the average pixel proximity to all same-valence words in relation to the length of the screen diagonal (i.e.

the lowest possible proximity). We termed this indicator *SpAM similarity*[1] (lower values indicate higher similarity). This indicator is identical to the density computation used by Unkelbach and colleagues (2008). The screen diagonal serves as a fixed calibration divisor.

Figure 1



Note. Example of distances in a possible SpAM solution. The 20 positive target words are clustered on the left side and the 20 negative target words are clustered on the right side. Here, the proximity, and thus similarity, between "flowers" and the other 19 positive words is greater than the proximity/similarity between "toothache" and the other 19 negative concepts.

The final stimulus map was compressed to fit into the upper two thirds of the screen, making space for seven equal and unlabeled boxes that appeared side by side in the lower third of the screen. Participants read "Your next task is to divide the 40 words that you have sorted into between two and seven categories. To assign a word to a category, please drag-and-drop it

[1] We calculated the target concepts' within-valence similarity, and not between-valence or overall similarity, to allow a comparison between our data and the data reported by Unkelbach and colleagues (2008).

into one of the category boxes that just appeared in the lower third of the screen; to reassign that word, simply drag-and-drop it from its current category box to another category box". Once all 40 words were categorized, participants could finalize the categorization phase. For all categorized words, the program recorded how many of the other same-valence words (i.e., X out of 19) had been assigned to the same category. On average, spatially arranging the 40 words took less than 10 minutes, and sorting them into between 2 and 7 categories took less than 5 minutes.

Results

For reasons of direct comparability, we report all inferential tests as *F*-tests. Participants clearly distinguished between the 20 positive and the 20 negative words, as the spatially arranged between-category distance (i.e., the average distance of positive to negative words and negative to positive words) was more than twice as large as the spatially arranged within-category distance (i.e., the average distance of positive to positive words and negative to negative words), $M = 2.58$, $SD = 1.06$.

More importantly, in line with a general valence asymmetry in similarity, participants spatially arranged the 20 positive words more closely to one another than the 20 negative ones ($M_{pos} = 14.49\%$ of the screen diagonal, $SD = 5.42$; $M_{neg} = 19.07\%$, $SD = 7.50$), $F(1, 54) = 25.79$, $p < .001$, $\eta_p^2 = .32$, 90% CI [.16, .46]. A comparison of the number of category boxes that contained positive and negative words at the end of the categorization phase revealed that participants also assigned the 20 positive words to fewer categories than the 20 negative words ($M_{pos} = 3.41$ out of 7, $SD = 0.86$; $M_{neg} = 4.05$, $SD = 1.00$), $F(1, 54) = 14.20$, $p < .001$, $\eta_p^2 = .21$, 90% CI [.07, .35].

These measures of similarity correlated positively across participants, $r(53) = .27$, $p < .05$. The fewer boxes participants used to categorize positive compared to negative words, the more densely did participants spatially arrange positive compared to negative words. These results are based on a participant-level analysis. Next, we tested whether these findings are also obtained on an item-level analysis; that is, for each positive/negative word, we aggregated similarity across participants.

Similar to the participant-level analysis, on the stimulus-level of analysis the spatially arranged between-category dissimilarity distance was more than twice as large as the within-category dissimilarity distance, $M = 2.44$, $SD = 0.49$.

More importantly, the difference in spatially arranged proximity/similarity between the 20 positive and 20 negative words ($M_{pos} = 14.49\%$, $SD = 1.46$ vs. $M_{neg} = 19.07\%$, $SD = 3.04$)

was again significant, $F(1, 38) = 36.69$, $p < .001$, $\eta_p^2 = .49$, 90% CI [.29, .62]. This effect was larger than the effect reported by Unkelbach and colleagues (2008), $F(1, 38) = 17.02$, $p < .001$, $\eta_p^2 = .31$, 90% CI [.12, .47], who analyzed pairwise similarity judgments for the same 40 words. Further, on the item-level of analysis, the positive compared to negative words were also assigned to categories together with more same-valence words ($M_{pos} = 7.50$ out of 19, $SD = 0.61$; $M_{neg} = 6.93$, $SD = 1.36$), but this effect was not significant, $F(1, 38) = 2.80$, $p = .10$, $\eta_p^2 = .07$, 90% CI [.00, .22].

The observed valence asymmetry in spatially arranged proximity/similarity might be due to other factors that might be confounded with valence; for example, the positive words (e.g., love and baby) might be more intensely positive compared to the intensity of the negative words (e.g., litter and cockroach). To exclude such alternative explanations, we predicted the stimuli's spatially arranged proximity/similarity from the stimuli's effect-coded valence, and their interval-scaled valence intensity, frequency, familiarity, and concreteness in a multiple linear regression[2]. Table 1 presents the results; the only two significant predictors of similarity were valence intensity and valence. The 20 positive words were more proximal/similar than the 20 negative words even when simultaneously controlling for valence intensity, frequency, familiarity, and concreteness.

Table 1

[2] We measured the 40 words' valence intensity in terms of the absolute difference between the 40 words' mean rating on a 0-10 negative-positive scale and 5, the affectively neutral midpoint of that scale (Klauer & Musch, 1999). We measured the 40 words' frequency of occurrence in the vast Corpus of Contemporary American English (~450 million words spoken or written between 1990 and 2012; Davies, 2011). Finally, we offered 26 students of the University of Cologne (14 women and 12 men; 26 native German speakers) a pack of gummi bears to rate the 40 words in a random order on a 1-10 either unusual-familiar ("ungewohnt-vertraut" in German) or abstract-concrete (in German "abstract-concrete") scale. We calculated the 40 words' familiarity and concreteness means.

Results of a multiple regression analysis predicting mean spatially arranged proximity/similarity from valence, valence intensity, frequency, familiarity, and concreteness across the 40 German words examined in Study 1

Predictors	*ß*	*t*	*p*	*r*	*pr²*	*VIF*
Valence	.71	4.02	< .001	.70	.57	3.04
Valence intensity	.37	3.26	< .01	.25	.49	1.27
Frequency	.01	0.04	.97	.24	.01	1.37
Familiarity	.12	0.78	.44	.65	.13	2.62
Concreteness	-.10	-0.81	.42	.17	-.14	1.37

Note. r, *pr²*, and *VIF* denote zero order and partial correlation, and variance inflation factor, respectively.

To validate the SpAM version[3] used here, we correlated the 40 words' within-valence SpAM similarity with their within-valence Pairwise similarity judged on a "similar-dissimilar" scale, which is arguably the gold standard of similarity measurement. Supporting the validity of SpAM, the correlation between the 40 words' SpAM similarity and Pairwise similarity (reported by Unkelbach et al., 2008) was very high, $r(38) = .84$, $p < .001$[4].

---

[3] The difference between our spatial arrangement version and the version used by Hout and colleagues (2013) is that in our version the stimuli appeared one after another in the middle of the screen, while in their version the stimuli appeared all at once in random locations on the screen. Thus, in our (their) version participants eventually (instantly) calibrated to the full set of stimuli.

[4] Individuals' agreement about the intra-category similarity of the 40 words was high when intra-category similarity was measured based on scaled pairwise judgments (the technique used by Unkelbach et al., 2008), mean r(38) = .87, SD = .17, moderate when it was measured based on unscaled pairwise judgments, mean r(38) = .52, SD = .19, and low when it was measured based on spatial arrangement, mean r(53) = .24, SD = .31. This difference in inter-rater agreement for Pairwise and SpAM intra-category similarity is presumably due the number of trials across which it is measured. In the paper by Unkelbach and colleagues (2008), participants on average made approximately ten pairwise intra-category similarity judgments per word, whereas in Study 1, participants made only one spatial intra-category similarity arrangement per word, as rearranging a word simultaneously readjusted all similarities between that word and all other words of the same category. This is precisely the efficiency advantage of SpAM over the Pairwise method to measure intra-category similarity. This advantage comes at the cost of low inter-rater agreement about intra-category similarity.

To further explore the correlations between these psychological (i.e. subjective) measures and two ecological (i.e. more objective) measures of word similarity, we calculated how often the 40 words co-occur in two real-life word environments: the internet (Google similarity; Cilibrasi & Vitanyi, 2007), and a collection of books that is representative of the literature read by US college students (LSA similarity; Landauer & Dumais, 1999). As frequency of co-occurrence in space and time is a widely accepted proxy for inter-stimulus similarity (e.g., Griffiths et al., 2007; Jones & Mewhort, 2007), these correlations provided further insights into the validity of SpAM similarity.

Correspondence between SpAM similarity and Google similarity. In February 2013, we entered all 780 word pairs that can be formed with the 40 words into the search bar of the most widely used search engine (Google Search[5]), and we recorded the amount of search "results" (hits). More precisely, we searched for both orders of each pair (e.g., "party friends" and "friends party"), and for each pair, we averaged hits across order, resulting in 780 pairwise hits. In Google Search, a pairwise hit approximates the total number of webpages on which two words co-occur. Next, to model the 40 words as points in a geometric space in which their similarity can be reliably compared, we subjected the multiplicative inverses (i.e., 1/X) of the 780 pairwise hits to a multidimensional scaling analysis (MDS; e.g., Borg & Groenen, 2005). Using the ALSCAL procedure (Young, Takane, & Lewyckyj, 1978) provided by the SAS system, we assumed an ordinal scale and estimated coordinates for each word in ten spaces. The 1D, 2D, 3D, 4D, 5D, 6D, 7D, 8D, 9D, and 10D coordinates of the 40 words retained R^2 = 0.70, 0.73, 0.76, 0.79, 0.83, 0.85, 0.87, 0.88, 0.90, and 0.91 (stress = 0.44, 0.31, 0.23, 0.19, 0.15, 0.13, 0.11, 0.10, 0.09, and 0.08; the lower the stress the higher the scaling fit of the respective space) of the original variance of the 780 pairwise hits, respectively.

There was no elbow in the stress scree plot. Thus, we resorted to the stress interpretation guideline by Kruskal and Wish (1978), according to which stress <= .20, <= .15, <= .10, <=.05 and <= .025 may be interpreted as poor, sufficient, satisfactory, good, and excellent,

---

On a side note, for the 40 positive/negative words both SpAM (M = 2.44, SD = 0.48) and the Pairwise method (M = 2.30, SD = 0.42) produced more than twice as much between-compared to within-category dissimilarity variance. Thus, given a stimulus set composed of two obvious main categories, SpAM and the Pairwise method clearly capture the categories, and they do so to a comparable extent.

[5] We used www.google.de instead of www.google.com, as the words are German rather than English. A test with some of the target word pairs revealed that www.google.de and www.google.com return the same amount of search results. Thus, we speculate that Google returns the same amount of results when searched from different countries.

respectively. We proceeded with the 6D space, because the 6D space is the first that achieved a sufficient scaling fit (stress <= .15; to balance scaling fit and parsimony, in MDS as many as necessary and as few as possible dimensions are extracted; Jaworska & Chupetlovska-Anastasova, 2009). Following Unkelbach and colleagues (2008), we calculated the average Euclidean proximity of each word to all other same-valence words in the respective 6D space. This index (Google similarity) correlated highly with the 40 words' SpAM similarity, $r(38) = .56$, $p < .001$, and with their average pairwise similarity, $r(38) = .56$, $p < .001$. In addition, the 20 positive words are also more similar to one another than the 20 negative words in terms of how often they co-occur on webpages accessible through Google Search, $F(1, 38) = 21.15$, $p < .001$, $\eta_p^2 = .36$, 90% CI [.16, .51].

Correspondence between SpAM similarity and LSA similarity. In November 2014, we entered the 40 words into the "Matrix Comparison" application of the Latent Semantic Analysis (LSA; Landauer & Dumais, 1999) online tool provided by the University of Colorado at Boulder, US (lsa.colorado.edu). This application returned the similarity of each of the 780 word pairs that can be formed with the 40 words as the cosine of the angle between the vectors of the words in a pair in a high-dimensional semantic space derived from the frequency of co-occurrence of all 104852 words in all 942425 passages in a collection of 738 books that is representative of the literature read by US college students. We selected the topic "General Reading up to 1st Year College (300 factors)" and the comparison "Term to Term", and we left "Numbers of Factors to Use" blank to receive the 780 cosine similarities in the highest-dimensional semantic space available for this topic (i.e., 338D). Next, we calculated the average cosine similarity of each of the 40 positive/negative words to all other same-valence words (LSA similarity). The 40 words' LSA similarity correlated strongly with their SpAM similarity, $r(38) = .64$, $p < .001$, with their similarity judged in pairs (reported by Unkelbach et al., 2008), $r(38) = .73$, $p < .001$, and with their Google similarity, $r(38) = .38$, $p < .05$. Also, the positive words were more similar to one another than the negative words in terms of how often they co-occur in paragraphs in the collection of books that is representative of the literature read by US college students, $F(1, 38) = 19.88$, $p < .001$, $\eta_p^2 = .34$, 90% CI [.14, .50]. Table 2 summarizes the correlations of SpAM, Pairwise (reported by Unkelbach et al., 2008), Google, and LSA similarity; these correlations indicated a high *construct validity* of the spatial arrangement method (SpAM; e.g., Hout et al., 2013) as a measure of similarity.

Table 2

Correlations between the SpAM, Pairwise, Google, and LSA measures of similarity

	Pairwise	Google	LSA
SpAM	.84***	.56***	.64***
Pairwise		.56***	.73***
Google			.38*

Note. Correlations between the psychological and ecological measures of word similarity examined in this article: the spatial arrangement method (SpAM), classical pairwise judgment on a "similar-dissimilar" scale, and frequency of co-occurrence on webpages accessible through Google Search (Google) and in passages in a collection of books that is representative of the literature read by US college students (LSA). * p =< .05, ** p =< .01, *** p =< .001.

Predictive strength of SpAM similarity for cognitive processing. Next, we compared how well these four measures of word similarity predicted five basic aspects of cognitive processing. First and second, words that are more similar to other words are evaluated faster on a "negative-positive" scale (Klauer & Musch, 1999) and classified faster as "negative" or "positive" (Unkelbach et al., 2008), presumably because they co-activate a more comprehensive pattern of related words in the associative memory network, speeding up word recognition (Unkelbach, 2012). Third and fourth, as more similar words are co-activated more often and more strongly, they are later more likely to be mistaken as having been present (e.g., in a previous phase in a study on recognition memory), resulting in more erroneous judgments about whether they are "old" or "new" (Alves et al., 2015). And, fifth, words that are more similar to other words are more likely to be subsumed under a category (Shepard, 1987), possibly also because they are more strongly associated to one another in the associative memory network (De Deyne, Peirsman, & Storms, 2009).

We obtained the data on how fast the 40 words are evaluated on a "negative-positive" scale (Klauer & Musch, 1999), how fast they are classified as "negative" or "positive" (Unkelbach et al., 2008), how likely they are to be falsely recognized as present before when they were in fact absent (in terms of signal detection theory: their sensitivity and response bias; Alves et al., 2015), and how likely they are to be subsumed under a category (measured in the present study). Across the 40 words, we correlated these five aspects of cognitive processing with each of the four measures of similarity discussed above. Table 3 shows the respective 20 correlations.

Table 3

Predictive strength of the SpAM, Pairwise, Google, and LSA measures of Similarity

	Evaluation speed	Classification speed	Recognition response bias	Recognition sensitivity	Categorization probability
SpAM	-.62***	-.58***	-.50**	-.33*	.51**
Pairwise	-.57***	-.68***	-.60***	-.47**	.31*
Google	-.41**	-.53***	-.18	-.25	.24
LSA	-.44**	-.54***	-.53**	-.46**	.21

Note. Predicting the 40 words' evaluation speed, classification speed, response bias (greater and lower values indicate a tendency toward "no" and "yes", respectively) / sensitivity in recognition memory, and probability of being subsumed under a category based on their similarity as measured with spatial arrangement (SpAM), pairwise judgment on a "similar-dissimilar" scale (Pairwise) and frequency of co-occurrence on webpages accessible through Google Search (Google) and in passages in a collection of books that is representative of the literature read by US college students (LSA). * p =< .05, ** p =< .01, *** p =< .001.

First, across the four measures, similarity substantially predicts all five aspects of cognitive processing. Second, the SpAM similarity measure significantly predicts all five aspects to an extent that is comparable to the similarity measure derived from pairwise judgment. Third, our Google similarity index only predicted evaluation and classification speed, but not recognition sensitivity, response bias, and categorization probability. The LSA similarity index did not predict categorization probability. In conclusion, SpAM similarity is an index with high construct and substantial predictive validity that is comparable to the standard measure of similarity.

Discussion

Study 1 provided two important insights. First, spatial arrangement (SpAM; e.g., Goldstone, 1994; Hout et al., 2013) is a valid method to measure word similarity. Word similarity measured with SpAM correlated highly with word similarity judged in pairs, and with two ecological (rather than psychological) measures of word similarity, namely frequency of co-occurrence on webpages and in book passages. Importantly, word similarity measured with SpAM also correlated with performance in a variety of basic cognitive tasks (evaluation, classification, recognition, and categorization), which confirmed that spatially arranged similarity is relevant for cognitive processing (for further demonstrations of the relevance of SpAM for cognitive processing, see Berman et al., 2014; Hout & Goldinger, 2015; Hout, Goldinger, & Brady, 2014).

Second, using SpAM we re-examined the similarity of the 40 words examined by Unkelbach and colleagues (2008; see also Klauer & Musch, 1999) in a more efficient way. We replicated that the 20 extremely positive words are seen as more similar to one another than the 20 extremely negative words (we observed a large effect, $\eta_p^2 = .32$), even when controlling for the valence intensity, frequency, familiarity, and concreteness of the words. Thus, for the present sample of stimuli, Study 1 confirmed the valence asymmetry in similarity derived from the range and frequency of co-occurrence arguments presented above.

However, generalizing across similarity measurement methods does not help with generalizing across the population of positive and negative stimuli (Wells & Windschitl, 1999; Westfall et al., 2014; 2015). Yet, the cost- and time-effective SpAM method allows testing the generality of the proposed higher similarity of positive compared to negative information with a large variety of stimulus samples.

Studies 2a and 2b

To provide a strong test for the proposed generality of valence asymmetry in similarity, we aimed to sample stimuli that are representative of positive and negative from the perspective of our participants (i.e., stimuli that come to their mind as examples of the categories “positive” and “negative”). In Study 2a, *retriever* participants first freely sampled words from memory that they themselves evaluated as positive and negative and then spatially arranged these words. In Study 2b, *receiver* participants first evaluated the words selected as positive and negative by another randomly selected participant and then spatially arranged these words. If the similarity asymmetry is a general phenomenon, both retrievers and receivers should spatially arrange the positive words closer to one another, expressing that they seem more similar to one another than the negative words.

Method

Participants, design, and stimuli. We advertised an online study on Amazon's crowdsourcing platform Mechanical Turk. In Study 2a, 46 MTurkers (24 women and 22 men; 45 native English speakers) took part for $1.5, and in Study 2b, 43 MTurkers (21 women and 22 men; 41 native English speakers) took part for $1. In both studies, participants spatially arranged 20 positive words and 20 negative words. All words were generated by participants in Study 2a.

Procedure. Both studies were fully computerized. In Study 2a, the first screen slide instructed participants to generate 20 positive words ("Please enter 20 different positive nouns into the 20 text boxes displayed below") and 20 negative words ("… 20 different negative nouns …") by typing them into groups of text boxes on the left and right of the screen (or vice versa), respectively. Then, participants completed the same SpAM procedure for the 40 self-generated words as in Study 1.

In Study 2b, participants did not generate words, but first rated each word (on a 7-point "negative-positive" scale, in random order) from one randomly selected set of 40 stimuli generated by another participant in Study 2a. Then, participants spatially arranged the respective 40 words. Both studies lasted between 10 and 20 minutes.

Results

Two participants were excluded from the analyses in Study 2a, because they took excessively long to complete the task (29.63 and 30.22 minutes; $M = 8.08$, $SD = 5.52$). This exclusion of participants did not affect any statistical inferences.

Study 2a. Participants generated 1044 unique words divided into 44 unique samples of 40 words. For each participant, we averaged spatially arranged within-valence distance across the 20 self-generated positive words, and across the 20 self-generated negative words. In line with a general valence asymmetry in similarity, retrievers spatially arranged their self-generated positive words closer to one another (M_{pos} = 15.77% of the screen diagonal, $SD = 5.43$) than their self-generated negative words (M_{neg} = 16.82%, $SD = 6.20$), $F(1, 43) = 4.28$, $p < .05$, $\eta_p^2 = .09$, 90% CI [>.00, .24].

Study 2b. Receivers almost always agreed with retrievers about the valence of the words. Specifically, on the 7-point “negative-positive” scale, receivers assigned a positive rating (5, 6, or 7) to words that had been retrieved as negative in only 2.09% of all cases, and assigned a negative rating (1, 2, or 3) to words that had been retrieved as positive in only 1.27% of all cases. Receivers rated the words retrieved as positive as more positive than the words retrieved as negative (M_{pos} = 6.04, *SD* = 0.35; M_{neg} = 1.96, *SD* = 0.32), $F(1, 42) = 2189.32$, $p < .001$, $\eta_p^2 = .98$, 90% CI [.97, .99].

More importantly, receivers also spatially arranged the 20 positive words closer to one another (M_{pos} = 16.32% of the screen diagonal, *SD* = 4.61) than the 20 negative words (M_{neg} = 18.46%, *SD* = 6.50), $F(1, 42) = 5.11$, $p < .05$, $\eta_p^2 = .11$, 90% CI [.01, .26]. That is, they also saw the retrievers’ positive words as more similar to one another than the retrievers’ negative words. Further, higher SpAM similarity of positive compared to negative words on the retrievers’ side correlated with higher SpAM similarity of positive compared to negative words on the receivers’ side, $r(41) = .29$, $p = .06$.

Discussion

Studies 2a and 2b used the efficiency advantage of the spatial arrangement method (SpAM; Hout et al., 2013) to measure the similarity of a large sample of words (i.e., 1044); we believe this high number of freely selected stimuli constituted a large and arguably representative sample of what people consider as positive and negative words. Thus, consistent with the notion of representative design (Brunswik, 1955; 1956; Dhami, Hertwig, & Hoffrage, 2004), we can generalize our results to what people consider as positive and negative words.

Study 2a confirmed that self-selected positive words were represented as more similar than self-selected negative words. Moreover, Study 2b showed that words that were positive for an unknown person were also, on average, seen as more similar than words that were negative for that person. However, these effects were only medium-sized ($M\eta_p^2 = .10$) and thus much smaller than the large positive-negative difference in word similarity observed in Study 1 ($\eta_p^2 = .32$).

There are obvious explanations for this decrease in effect size: a) the 40 words in Study 1 were biased in favor of the hypothesis, b) free sampling increased error variance, and c) the online workers put less time and effort into completing the task. None of these reasons jeopardizes the support for the proposed generally higher similarity of positive compared to negative information.

A possible caveat for the generality might result from the high agreement between retrievers and receivers on the valence of the words. The high agreement might suggest that the word generation task communicated that participants should retrieve positive and negative words on whose valence participants and researchers should agree. Therefore, the higher similarity of positive compared to negative stimuli might be restricted to words of consensual valence. Study 3 therefore investigated whether the similarity asymmetry also holds for idiosyncratic valence, that is, for stimuli that only some individuals evaluate as positive and negative.

Study 3

Stimuli that are good/bad only for a given individual provide a particularly strong test of the generality of the predicted similarity asymmetry. Personal interests, preferences, and liking often result in repeated exposure, keen exploration, and thus motivated differentiation on the positive side (Smallman, Becker, & Roese, 2014; Smallman & Roese, 2008). For example, fans of ball sports might argue that football, basketball, and baseball and so forth “are all different”. Thus, stimulus words referring to concepts someone personally likes might actually appear more differentiated. Quite to the contrary, personal disinterest and disliking often result in avoidance and thus motivated summarization on the negative side (Fazio et al., 2004; Denrell, 2005). For example, people who do not like ball sports might argue that football, basketball, and baseball “are all the same”. Thus, stimulus words referring to concepts someone personally dislikes might actually appear more similar. Together, it is possible that words referring to idiosyncratically positive stimuli might be seen as less similar to one another than words that refer to idiosyncratically negative stimuli. Study 3 therefore investigated whether idiosyncrasy versus consensus moderates valence asymmetry in perceived similarity.

Participants self-selected words that are positive and negative either idiosyncratically (i.e., “for you personally”), or consensually (i.e., “for all people”). Then, as in Study 2a, participants spatially arranged the sampled words. If idiosyncratic valence leads to greater differentiation on the positive side, and to greater summarization on the negative side, we would expect an interaction of generation task (idiosyncratic vs. consensual) and stimulus valence.

Method

Participants, design, and stimuli. 110 students (86 women and 24 men; 102 native German speakers) were paid €2 to take part in the study. Similar to Study 2a, participants spatially arranged self-generated positive and negative words. We randomly assigned participants either to an *idiosyncratic* or a *consensual* valence condition. Given this sample size and an observed correlation of $r = 0.70$, $p < .001$, between the repeated measures, the statistical power to detect a small interaction effect ($\eta_p^2 = .02$) was > .95 (G*Power 3.1; Faul, Erdfelder, Buchner, & Lang, 2009).

Procedure. Procedural details were highly similar to Study 1 with small variations. Participants in the *idiosyncratic* valence condition read (translated from German) “We are interested in finding out the things that you *personally* find positive and negative. Please enter 20 positive and 20 negative words that you personally find positive and negative into the text boxes on the left and right of the screen. It is important that you type in different words that you personally find positive and negative. Please type in single words only.” The program counterbalanced whether participants entered positive (negative) information on the right or left side of the screen. Participants in the *consensual* valence condition read the same instructions, except that “you personally” was exchanged with “all people”.

Then, participants in both conditions spatially arranged the self-generated stimuli. Different from the previous studies, the 40 words appeared all together (en bloc in five columns and eight rows in the middle of the screen) instead of one after another. Thus, participants always had an overview of the 40 words while spatially arranging them. Sessions lasted between 10 and 20 minutes.

Results

Participants in the idiosyncrasy condition took an equal amount of time to generate the 20 positive and 20 negative words (M = 490s; SD = 188s) as those in the consensus condition (M = 499s; SD = 197s), $F(1, 108) = 0.05$, $p = .83$, $\eta_p^2 = .00$, 90% CI [.00, .02].

Manipulation check. Participants in the idiosyncrasy condition should generate more diverse stimuli than participants in the consensus condition. Indeed, participants in the idiosyncrasy condition generated more diverse stimuli (1139 unique stimuli out of the 55 participants * 40 stimuli = 2200 generated stimuli) than participants in the consensus condition (995 unique out of 2200 generated stimuli). This difference was significant, $X^2(1) = 9.71$, $p < .01$.

Frequency of unique words. Independent of the idiosyncrasy versus consensus manipulation, participants generated less unique stimuli for the category “positive” (946 out of

2200) compared to the category “negative” (1180 out of 2200), $X^2(1) = 27.44$, $p < .001$. This smaller diversity was apparent in both the idiosyncratic valence condition (511 unique positive words vs. 628 unique negative words), $X^2(1) = 12.01$, $p < .001$, and in the consensual valence condition (435 unique positive words vs. 560 unique negative words), $X^2(1) = 15.70$, $p < .001$.

SpAM Similarity. Table 4 displays participants’ mean SpAM similarity and standard deviations by experimental conditions. We analyzed these data with a 2 (generation task: idiosyncrasy vs. consensus) x 2 (valence: positive vs. negative) mixed ANOVA with repeated measures on the latter factor. The analysis showed main effects of the generation task, $F(1, 108) = 5.12$, $p < .05$, $\eta_p^2 = .05$, 90% CI [>.00, .12], and valence, $F(1, 108) = 37.74$, $p < .001$, $\eta_p^2 = .26$, 90% CI [.15, .36], but the interaction term was *not* significant, $F(1, 108) = 0.47$, $p = .49$, $\eta_p^2 = .00$, 90% CI [.00, .05]. Participants spatially arranged positive words closer to one another than negative words, regardless of their idiosyncratic or consensual valence. Participants also arranged the 40 words closer to one another in the consensual valence condition than in the idiosyncratic valence condition, again reflecting the manipulation’s success.

Table 4

Similarity means and F-tests for positive and negative stimuli in study 3’s idiosyncratic and consensual conditions (standard deviations in parentheses)

	Positive valence	Negative valence	*F*	*p*	η_p^2	*90% CI LB*	*90% CI UB*
Idiosyncratic valence	15.39 (6.74)	18.75 (7.91)	20.84	< .001	.28	.12	.42
Consensual valence	13.22 (4.07)	15.91 (6.02)	16.90	< .001	.24	.09	.38

Note. Values reflect the spatially arranged average pixel distance between all positive stimuli (20) or all negative stimuli (20) in relation to the diagonal of the screen.

Discussion

Participants adhered to the instructions and generated more diverse stimuli in the idiosyncratic compared to consensual valence condition. Results nevertheless showed the proposed greater similarity of positive compared to negative stimuli in both conditions, again supporting a general valence asymmetry in similarity. Although participants should know more about and differentiate more between what they personally like compared to dislike (Smallman, Becker, & Roese, 2014; Alves, Koch, & Unkelbach, 2016), they spatially arranged positive idiosyncratic stimuli more densely to one another than negative idiosyncratic stimuli. This valence asymmetry was as pronounced (η_p^2 = .28) as in the consensual valence condition (η_p^2 = .24).

The effect sizes ($M\eta_p^2$ = .26) are close to Study 1 (η_p^2 = .32), suggesting that mainly the error variance introduced by recruiting participants online was responsible for the lower effect sizes in Studies 2a and 2b ($M\eta_p^2$ = .10). Alternatively, giving participants an outright rather than gradually increasing overview of the 40 words to be spatially arranged might have decreased error variance. Of note, Study 3's spatial arrangement design follows the procedures by Hout and colleagues (2013) more closely.

Different from Studies 2a/2b, Study 3's participants (university students) sampled German rather than English words. Therefore, Study 3 additionally showed that the hypothesized valence asymmetry in similarity holds true also across different languages and different participant pools.

Studies 2a, 2b and 3 examined a large variety of stimulus words freely sampled by participants, thereby avoiding researcher-selected stimulus samples biased in favor of their hypothesis (Fiedler, 2011) and allowing generalization across the population of stimuli (Wells & Windschitl, 1999; Westfall et al., 2014; 2015). However, the free sampling process provides another alternative explanation; the observed valence asymmetry in similarity might be due to the *process of selecting positive and negative words* – that is, a valence-specific sampling bias – and less due to actual similarity differences of the retrieved stimuli (i.e., assuming a representative sample). Study 4 addressed this concern.

Study 4

Study 4 sought to rule out that the valence asymmetry in similarity observed in Studies 2 and 3 was only due to biased retrieval processes. It could be that positive and negative information are factually equal in similarity, but participants retrieved positive stimuli that are more similar to one another compared to the negative stimuli that they retrieved. For example, retrieving positive and negative words may have induced positive and negative affect (Topolinski & Deutsch, 2012; 2013), which might have modulated inclusive and exclusive thinking (Bless & Fiedler, 2006; Forgas, 2013) resulting in a tendency to select similar and dissimilar words, respectively. Or, as Fazio and colleagues (2004) suggested, positive stimuli invite exploration, while negative stimuli are abandoned.

To illustrate these principles, participants doing the positive-happy-inclusive-similar half of the word selection process might have selected “friends”, then “family”, then “partner”, then “love” and so on, exploring neighboring positive stimuli. In the negative-sad-exclusive-dissimilar half, participant might have selected “bombs”, then “lie” (rather than “war”), then “junk” (rather than “guilt”), then “depression” (rather than “germs”) and so forth, abandoning each negative stimulus without exploring the mental neighborhood further. Such an explanation would be interesting per se, but provides a clear alternative for the proposed general valence asymmetry in stimulus similarity.

To exclude the possibility of such valence-biased stimulus retrieval, Study 4 restricted the stimulus generation process to one positive and one negative stimulus per participant, thereby excluding possible explanations in terms of stimulus retrieval processes. Specifically, participants in one of two random samples were instructed to generate only one positive and only one negative word. The non-redundant positive words generated in this way were combined to form a multi-source (i.e., many participants as the source) pool of positive words whose selection was completely independent of one another, as they had been generated by as many participants as there were positive words in the pool (one positive word per participant). The non-redundant negative words were combined in the same way. Out of these two multi-source pools, different participants received 20 positive and 20 negative randomly drawn words, and then spatially arranged these. This procedure precluded explanations in terms of valence influences during retrieval or, in other words, the processing rather than meaning of positive/negative words.

Additionally, in Study 4 we wanted to exclude explanations in terms of the valence intensity, frequency, familiarity, and concreteness of positive/negative words in the same way as in Study 1 – that is, by an item-level multiple linear regression analysis. Please note that this was not possible in Studies 2 and 3; there, each participant generated/received a new set of stimuli, prohibiting such item-based analyses. Study 4 will thereby show that valence predicts spatially arranged proximity/similarity beyond alternative item characteristics in a sample of English rather than German words.

Method

Participants, design, and stimuli. 40 MTurkers were paid $0.1 to retrieve one positive and one negative word. Another 54 MTurkers (23 women and 31 men; 54 native English speakers) received a random sub-selection of 40 of the words retrieved in this way (20 out of 29 non-redundant positive words, e.g., “courage”, “happy”, “awesome” etc.; and 20 out of 35 non-redundant negative words, e.g., “boring”, “afraid”, “fat” etc.; see Appendix B), and were paid $0.8 to spatially arrange these.

Procedure. The study was fully computerized. The first 40 participants generated one positive and one negative word. Then, after filtering redundant stimuli, the second random sample of 54 participants completed the same spatial arrangement task as in Studies 1 and 2. They spatially arranged 20 positive words, randomly selected from the 29 non-redundant positive words, and 20 negative words, randomly selected from the 35 non-redundant positive words; again, these 29 positive and 35 negative words were independently generated by the 40 participants in the first sample. As in Studies 1 and 2, the 40 words appeared sequentially in the middle of the sorting screen. For the participants who generated the words, Study 4 took less than a minute. For those who spatially arranged the words, Study 4 took between 5 and 15 minutes.

Results

Supporting a general valence asymmetry in similarity, participants arranged the 20 randomly selected positive words more densely (M_{pos} = 16.93% of the screen diagonal, *SD* = 6.86) compared to 20 randomly selected negative words (M_{neg} = 19.09%, *SD* = 6.28), $F(1, 53) = 7.40$, $p < .01$, $\eta_p^2 = .12$, 90% CI [.02, .26].

As the number of positive (29) and negative (35) stimuli was fixed, we could test whether the observed asymmetry was actually due to valence. For each positive/negative word, we aggregated spatially arranged proximity/similarity across participants. On this item-level of analysis, the difference in spatially arranged proximity/similarity between the 29 positive and 35 negative words (M_{pos} = 16.39%, *SD* = 1.48 vs. M_{neg} = 19.09%, *SD* = 2.56) was again significant, $F(1, 62) = 17.03$, $p < .001$, $\eta_p^2 = .22$, 90% CI [.08, .35].

Similar to Study 1, we predicted the 64 words' spatially arranged similarity from their effect-coded valence, and their interval-scaled valence intensity, frequency, familiarity, and concreteness in a multiple linear regression[6]. Table 5 shows the results. Similar to Study 1 and the data in Table 1, the regression confirmed that the 29 positive words were more similar than the 35 negative words even when controlling for valence intensity, frequency, familiarity, and concreteness[7]. As in Study 1, the only other significant predictor of similarity was valence intensity.

[6] The analysis actually included only 63 of the 64 words, because we could not obtain a valence intensity rating for the word "myopic". We measured valence intensity in terms of the absolute difference between the 63 words' mean rating on a 1-9 "negative-positive" scale and 5, the affectively neutral midpoint of that scale (Warriner, Kuperman and Brysbaert, 2013). We measured the 64 words' frequency of occurrence in the vast and representative Corpus of Contemporary American English (~450 million words spoken or written between 1990 and 2012; Davies, 2011). Finally, we paid 50 MTurkers (22 women and 28 men; 50 native English speakers) $0.5 to rate the 64 words in a random order on a 1-10 either "unusual-familiar" or "abstract-concrete" scale. We calculated the 64 words' familiarity and concreteness means.

[7] To rule out multicollinearity between valence and familiarity, we computed the variance inflation factor (VIF) for all predictors in the multiple linear regression in Studies 1 and 4 (see Tables 1 and 5). According to Menard (1995), multicollinearity is a concern with VIFs greater than 5; according to Hair and colleagues (1995) and Mason and colleagues (1989), multicollinearity is a concern with VIFs greater than 10. None of our predictors had a VIF greater than 5.

Table 5

Results of a multiple regression analysis predicting mean spatially arranged proximity/similarity from valence, valence intensity, frequency, familiarity, and concreteness across the 64 English words examined in Study 4

Predictors	*ß*	*t*	*p*	*r*	pr^2	*VIF*
Valence	.58	2.66	.01	.46	.33	4.27
Valence intensity	.38	3.43	.001	.45	.41	1.08
Frequency	.03	0.28	.78	.13	.04	1.10
Familiarity	-.18	-0.86	.39	.35	-.11	3.88
Concreteness	.05	0.42	.68	-.12	.06	1.35

Note. *r*, pr^2, and *VIF* denote zero order and partial correlation, and variance inflation factor, respectively.

Discussion

Study 4 used stimuli randomly selected from a pool to which each participant had contributed only one positive and one negative word. This independent word generation precluded valence asymmetries in similarity due to retrieval processes, such as selecting similar and dissimilar words due to inclusive and exclusive thinking (Forgas, 2013), or more exploratory sampling for positive compared to negative stimuli (Fazio et al., 2004). Participants nevertheless spatially arranged positive compared to negative words more densely to one another.

Further, valence significantly predicted spatially arranged proximity/similarity even when the valence intensity, frequency, familiarity, and concreteness of the words was controlled for, which suggests that the spatially arranged difference in proximity/similarity between the positive and negative words was actually due to their valence, and *not* due to other features that might be confounded with valence.

The present data does not preclude that the aforementioned alternative explanations contribute to the effect in general (i.e., valence influence during stimulus retrieval). However, in the present study and in the original study by Unkelbach and colleagues (2008; Experiment 2) they could not contribute, which shows that the asymmetry persists independent of these possible contributions.

The effect (η_p^2 = .12) was about the same size as the effects observed in Studies 2a/2b ($M\eta_p^2$ = .10), suggesting that using SpAM with online participants increases error variance. Alternatively, the mode of stimulus presentation (simultaneously vs. serially) might have influenced the effect size. In any case, the asymmetry emerged for both presentation modes and for online and laboratory participants.

Study 5

The previous studies compared the similarity of words that come to mind as exemplars of the categories positive and negative. However, these words may only represent imagined, possible concepts, which are not representative of real-life experiences. People receive and thus retrieve all kind of positive and negative information that they have not experienced directly (e.g., being elected as president, winning the jackpot, staying healthy for 100 years, suffering from Parkinson’s disease, losing a child, causing a car accident, etc.). The number of such second-hand information by far exceeds the number and variety of self-experienced (positive and negative) information. Thus, it could be that the greater similarity of positive compared to negative stimuli observed in Studies 1-4 may be true for imagined, possible objects, people, and events, but does not hold for self-experienced stimuli; again, this might be because people purposefully accumulate more self-experienced stimuli on the positive side (Fazio et al., 2004; Denrell, 2005), which should lead to more differentiation and thus less similar mental representations compared to the negative side. Study 5 investigated whether positive self-experiences are seen as more similar than negative self-experiences, too.

Study 5 employed an event-sampling design; across seven consecutive days, participants named one positive and one negative “event of the day” and then – on day eight, nine, or ten – spatially arranged these real-life events. If there is a general valence asymmetry in similarity, participants should arrange their positive everyday experiences as more similar to one another than their negative everyday experiences, thereby generalizing our findings from the semantic denotations of words to connotative real-life experiences.

Method

Participants, design, and stimuli. We recruited participants via the mailing list of psychology students at a large German university, and online, via large open access Facebook groups, for example: NETT-WERK KÖLN (115,000+ members) and Neu in Köln (15,000+ members). We offered €15 for taking part in a week-long event-sampling study (Reis, Gable, & Maniaci, 2014) on work-life balance[8]. On 7 consecutive days, participants received a text message at nighttime (9PM ± 30 minutes). The links in these text messages redirected participants to the survey website on which the study was hosted. On this website, above a blank text box, participants read: "Please describe one positive event of your day using no more than three words. Your description of this positive event should be precise, so that you can recognize it at the end of the study week"; above another blank text box, they read: "Please describe one negative event of your day …" (the order of these two instructions plus text box was random).

On day 8, 168 participants who had described a positive and a negative everyday event on at least 5 out of the 7 study days received an email with instructions on how to complete the final task (see below) within 3 days. 124 participants (95 women, 29 men; 119 native German speakers) completed the final task.

Procedure. The final task was a fully computerized SpAM study. Participants first read the same SpAM instructions as in Studies 2a, 2b and 3, except that there was no mention of positive/negative, and except that "words" was replaced with "events of the day" (all Study 5 instructions presented here are English translations of the German instructions provided to participants). The next slide provided an overview of the positive and negative everyday events (arranged in two columns and 5-7 rows, and in random order) that they had experienced and described over the course of their last week. Finally, they spatially arranged the positive and negative everyday events. The stimuli appeared on-demand, in random order.

Results

A 3 (response rate: on 5 days vs. on 6 days vs. on 7 days; between participants) x 2 (everyday event valence: positive vs. negative; within participants) mixed ANOVA of everyday event similarity revealed no main effect of response rate, $F(2, 121) = 0.47$, $p = .63$, $\eta_p^2 = .01$, 90% CI [.00, .04], but a main effect of everyday event valence, $F(1, 121) = 3.92$, $p = .05$, $\eta_p^2 = .03$, 90% CI [>.00, .10]. As expected, participants spatially arranged the positive everyday events of their last week closer to one another (M_{pos} = 19.75% of the screen diagonal, SD =

8 Study 5 was part of a larger investigation of work – life tradeoffs in everyday life (Rom & Hofmann, 2015).

9.67; M_{neg} = 21.44%, *SD* = 10.21). The interaction was not significant, $F(2, 123) = 0.53$, $p = .59$, $\eta_p^2 = .01$, 90% CI [.00, .04].

We repeated this analysis with the 92 most conscientious participants who described a positive and a negative everyday event on 6 or 7 of 7 study days (71 women, 21 men), revealing the very same results: no main effect of response rate, $F(1, 90) = 0.10$, $p = .75$, $\eta_p^2 = .00$, 90% CI [.00, .01], but a main effect of everyday event valence, $F(1, 90) = 5.72$, $p < .05$, $\eta_p^2 = .06$, 90% CI [>.00, .12]. The interaction was again not significant, $F(1, 90) = 0.04$, $p = .85$, $\eta_p^2 = .01$, 90% CI [.00, .01].

Discussion

Study 5 showed that the proposed valence asymmetry in perceived similarity generalizes from the semantic meaning of positive and negative words to experience-sampled positive and negative real-life events. Participants' spatial arrangements showed that the positive everyday events of their last week were significantly more similar to one another than the negative everyday events of their last week, indicating that despite the hedonic principle (pleasures are sought and pains are avoided), pleasures are more similar than pains.

While the effect sizes (η_p^2 = .03-.06) were much smaller than the effect sizes obtained in Studies 2a/2b, 3 and 4 ($M\eta_p^2 = .15$), the possible high variety of events across seven days and the lack of experimental control might fully account for this decrease. In addition, multi-word experiences might be less easy to spatially arrange than single-word concepts, increasing error variance in SpAM similarity. Nevertheless, the results still supported the proposed similarity asymmetry.

Study 6

Studies 2-5 tested the generality of the proposed valence asymmetry in similarity in large, representative samples of words retrieved as exemplars of "positive" and "negative", and everyday life events retrieved as "positive" and "negative". However, the ratings from Study 2 locate both the negative stimuli (M = 1.95) and the positive stimuli (M = 6.04) on the extremes of a 7-point valence scale. Similarly, Study 1's 40 stimuli are the 20 most extremely positive and 20 most extremely negative stimuli from the 92 stimulus words set by Fazio and colleagues (1986). We thus cannot be reasonably sure that the proposed valence asymmetry in similarity holds true across the across the entire spectrum of valence intensity ranging from mildly to moderately to extremely positive/negative.

To explore if this is the case, and to explore if the proposed valence asymmetry in similarity generalizes from verbal to visual stimuli, Study 6 examined two large databases of extremely, moderately, and mildly valenced words and images: the ~14,000 word database by Warriner, Kuperman, and Brysbaert (2013; WKB) and the international affective picture system with 956 pictures (IAPS; Lang et al., 2005). We expected the positive WKB words and IAPS pictures to be more similar to one another than negative WKB words and IAPS pictures, respectively.

Method

Participants, design, and stimuli. We re-analyzed data on all 13,915 words that together form the WKB; these words had been selected "to collect affective ratings for a majority of well-known English content words" (Warriner et al., 2013, p. 1192). Each word had been rated by approximately 25 MTurkers. Each MTurker had used a 9-point scale to assess one of the three arguably most relevant aspects of affective impression: valence, arousal, and potency (Osgood, Suci, & Tannenbaum, 1957).

Further, we re-analyzed data on all 956 pictures that together form the IAPS in its version from 2005; these color pictures had been selected with the aim to create a "broad sample of contents across the entire affective space" (Lang et al., 2005, p.3). Each IAPS picture had been rated by approximately 100 students of the University of Florida. These participants had also used 9-point scales to assess valence, arousal, or dominance.

We divided the words and pictures into a positive and a negative half (median-split) according to their mean valence ratings. We then computed the average absolute rating difference of each word to all other same-valence words, and of each picture to all other same-valence pictures. Separately for the words and the pictures, we computed this absolute rating difference across the three rating dimensions (i.e., valence, arousal, and dominance), and also separately for each rating dimension. Operationalizing absolute rating difference as a dissimilarity measure (e.g., the valence rating of the two IAPS pictures 428 and 927 are 6.89 and 6.98; thus, these two pictures have a similarly positive valence rating), for each of the 13,915 WKB words and 956 IAPS pictures, we obtained an overall similarity index, a valence similarity index, an arousal similarity index, and a dominance similarity index. Lower values on these four indices indicate higher similarity.

Results

Given the nature of the data, we conducted the relevant analysis on the level of stimuli. Table 6 summarizes the results. As expected**,** the overall similarity of the positive words and pictures was greater than the overall similarity of the negative words and pictures, respectively. The same was true for the valence, arousal, and dominance similarity indices.

The three WKB/IAPS rating dimensions correlated with one another (WKB: valence and arousal, $r(13913) = -.19$, $p < .001$, valence and dominance, $r(13913) = .71$, $p < .001$, and arousal and dominance, $r(13913) = -.18$, $p < .001$; IAPS: valence and arousal, $r(954) = -.28$, $p < .001$, valence and dominance, $r(954) = .84$, $p < .001$, and arousal and dominance, $r(954) = -.59$, $p < .001$). To test if the positive compared to negative WKB words and IAPS pictures are seen as more similar to one another in terms of valence *independent of arousal and dominance,* in terms of arousal *independent of valence and dominance*, and in terms of dominance *independent of arousal and dominance,* we repeated the single-dimension analyses reported above, but with the unstandardized residuals of a dimension regressed on the other two dimensions. The pattern remained unchanged with one exception. The valence and dominance residual similarity of the positive words/pictures was also higher than the valence and dominance residual similarity of the negative words/pictures, respectively, but the arousal residual similarity of the positive compared to negative words/pictures was not higher (see Table 6).

Table 6

Similarity means and F-tests for different indices of WKB word and IAPS picture similarity in Study 6 (standard deviations in parentheses).

Similarity	Positive	Negative	*F*	*p*	η_p^2	*90% CI LB*	*90% CI UB*
Words							
Overall	0.59 (0.28)	0.72 (0.33)	620.32	< .001	.04	.04	.05
Valence	0.54 (0.37)	0.74 (0.46)	848.04	< .001	.06	.05	.06
Arousal	0.69 (0.53)	0.73 (0.54)	18.68	< .001	.00	.00	.00
Dominance	0.55 (0.42)	0.70 (0.50)	340.30	< .001	.02	.02	.03
Valence residuals	-0.08 (0.37)	0.08 (0.44)	603.74	< .001	.04	.04	.05
Arousal residuals	0 (0.53)	0 (0.53)	0.14	.71	.00	.00	.00
Dominance residuals	-0.04 (0.42)	0.04 (0.48)	121.09	< .001	.01	.01	.01
Images							
Overall	0.88 (0.19)	1.30 (0.29)	702.29	< .001	.42	.39	.46
Valence	0.87 (0.23)	1.28 (0.27)	610.23	< .001	.39	.35	.43
Arousal	1.19 (0.38)	1.46 (0.43)	106.30	< .001	.10	.07	.13
Dominance	0.59 (0.26)	1.17 (0.33)	910.11	< .001	.49	.45	.52
Valence residuals	-0.06 (0.26)	0.06 (0.23)	49.41	< .001	.05	.03	.07
Arousal residuals	0.03 (0.39)	-0.03 (0.32)	5.96	< .05	.01	.00	.02
Dominance residuals	-0.11 (0.30)	0.11 (0.21)	159.82	< .001	.14	.11	.18

Note. Values reflect the average absolute rating difference (on a 9-point scale) between all 6958/478 positive WKB words/IAPS pictures and all 6958/478 negative WKB words/IAPS pictures. Lower values indicate higher inter-stimulus similarity.

In sum, based on the available ratings, the positive half of the ~14,000 WKB words are more similar to one another than the negative half of all WKB words both overall and on two of three independent (i.e., residualized) rating dimensions, and the same results were obtained for the ~1,000 IAPS pictures.

Discussion

Study 6 generalized the proposed asymmetry in similarity from participant-generated words that are representative of extreme positivity and negativity to researcher-selected words that are representative of the entire spectrum of valence intensity ranging from mildly to moderately to extremely positive/negative. The ~7,000 positive WKB words were more similar to one another than the ~7,000 negative WKB words. This effect was found overall, across all rating dimensions (η_p^2 = .04), and separately for the valence ratings/residuals (η_p^2 = .06/.04) and dominance ratings/residuals (η_p^2 = .02/.01), but not for the arousal ratings/residuals (η_p^2 = .00/.00).

These effect sizes reveal that the valence asymmetry in similarity observed in Study 6 was less pronounced than the asymmetries observed in the previous studies, possibly because the difference in similarity between moderately and weakly positive and negative words is still present but not as marked as in strongly positive and negative words.

Study 6 also generalized the proposed asymmetry in similarity from words to pictures that are representative of the entire valence spectrum. The ~500 positive IAPS pictures were more similar to one another than the ~500 negative IAPS pictures, an effect that was also found across all rating dimensions (η_p^2 = .34), and separately for the valence ratings/residuals (η_p^2 = .39/.05), dominance ratings/residuals (η_p^2 = .49/.15), and arousal ratings (η_p^2 = .10), but not for the arousal residuals (η_p^2 = .01).

These effect sizes suggest that the difference in similarity between positive and negative pictures is as marked as in strongly positive and negative words. To explore reasons for the more pronounced valence asymmetry in similarity in pictures compared to words, for each WKB word and IAPS picture, we calculated the absolute rating difference between its valence and the mean valence of all WKB words and IAPS pictures, respectively. The valence rating scales of these words and pictures are identical (1-9 “unhappy-happy”) and thus comparable. The mean valence deviation of the IAPS pictures from the midpoint of the scale (*M* = 1.54, *SD* = 0.94) is stronger than the mean valence deviation of the WKB words from the midpoint of the scale (*M* = 1.03, *SD* = 0.76), $F(1, 14869) = 394.16$, $p < .001$, η_p^2 = .03, 90% CI [.02, .03]. Thus, it could be that we observed greater valence asymmetry in similarity in the IAPS pictures compared to the WKB words because the pictures are more strongly positive and negative than the words. This conclusion is further supported by the valence asymmetries in similarity observed in Studies 1-4 in which we examined mostly strongly positive and negative words (mean deviation from the midpoint of the 1-9 scale: *M* = 2.49, *SD* = 0.61[9]). The effect sizes in these studies are consistently higher than the WKB word and everyday event effect sizes in Studies 6 and 5 (experienced real-life events should be less strongly positive and negative than imagined, possible objects, people, and events, see Studies 1-4), respectively. In sum, in combination with the previous studies Study 6 suggests that valence intensity is a moderator of valence asymmetry in similarity.

Moreover, Study 6 shows that the higher overall similarity of positive compared to negative WKB words and IAPS pictures *cannot* be reduced to the positivity variance of the positive words and pictures being smaller than the negativity variance of the negative words and pictures, respectively. Instead, Study 6 shows that impressions of positive pictures are also more similar to one another than impressions of negative pictures in other relevant respects than valence, namely dominance, a finding that is consistent with the notion that there are more negative than positive basic emotions (e.g., Ekman & Friesen, 1971).

General Discussion

We started the present investigation with the density hypothesis in mind; this hypothesis states that "positive information is more similar to other positive information, in comparison with the similarity of negative information to other negative information" and "let us assume a hypothetical space in which proximity signifies similarity. Within such a spatial model, greater similarity of positive compared to negative information implies a higher density (or closeness)

---

[9] Based on the WKB database (Warriner, Kuperman, & Brysbaert, 2013), we recorded the valence of the words examined in Study 4 on a 1-9 scale. We contracted the 0-10 valence scale used in Study 1 to a 1-9 scale to enable comparisons between Studies 1, 4, and 6 (we did not collect valence ratings for the thousands of words examined in Studies 2, 3, and 5, because this would have taken a great deal of time; however, the instructions under which participants named words in Studies 2 and 3 were the same as in Studies 1 and 4, and thus the valence intensity of the words examined in Studies 1-4 is presumably the same). The mean valence deviation of the Study 1 and 4 words from 5, the midpoint of the 1-9 WKB scale (i.e., valence intensity) was M = 2.77, SD = 0.44, and M = 2.32, SD = 0.64, respectively. Across Studies 1 and 4, the mean valence intensity was M = 2.49 and SD = 0.61, and thus greater than the mean WKB words and the mean IAPS pictures valence intensity in Study 6.

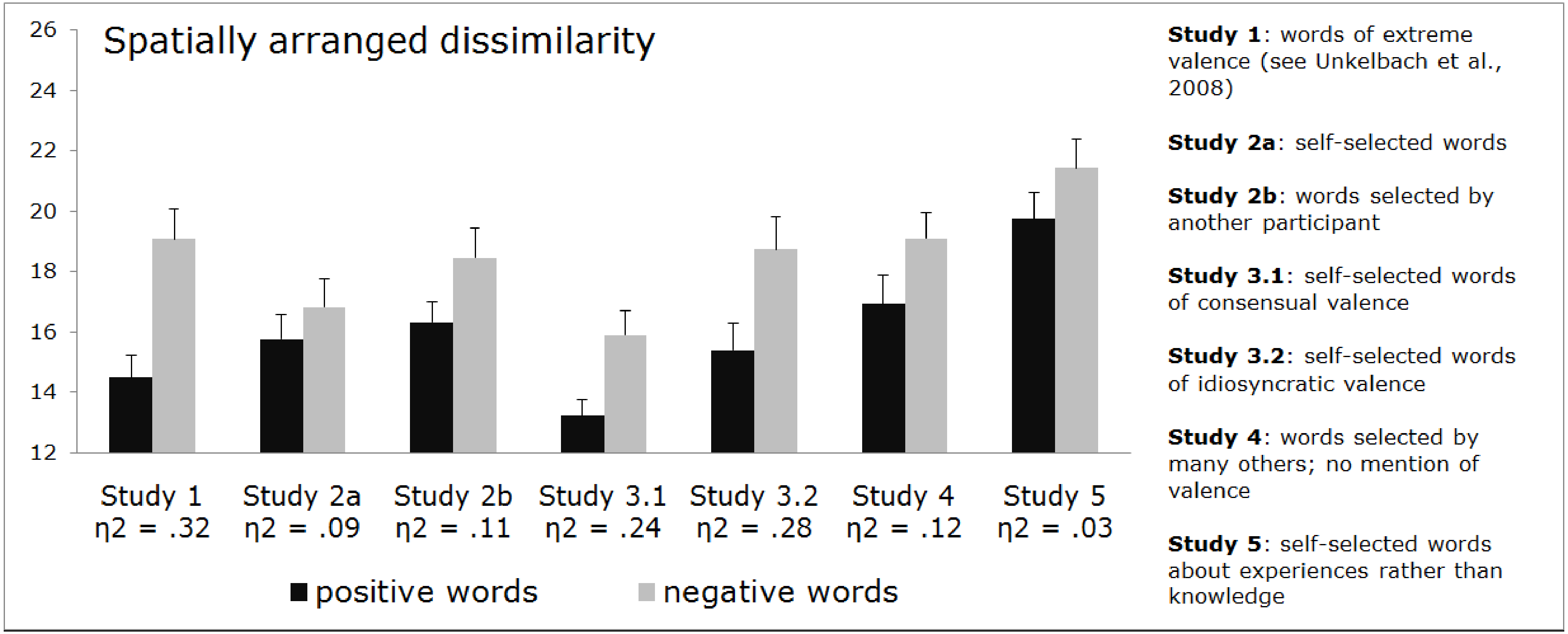
on average." (Unkelbach et al., 2008, p. 30). We argued that the available evidence for a general valence asymmetry in similarity is not convincing, because it has been directly shown only two times (Bruckmüller & Abele, 2013; Unkelbach et al., 2008), because the researcher-selected positive and negative words examined in these studies may have been biased samples (i.e., possibly not representative of positive and negative as seen from the perspective of participants; Fiedler, 2011), and because the observed asymmetry in similarity may have been due to differences in the valence intensity, frequency, familiarity, and / or concreteness of these positive/negative words rather than due to their valence. The aim of this paper was to solve these problems by repeatedly showing that the proposed valence asymmetry in similarity generalizes across large, representative samples of positive and negative stimuli, and by showing that the effect is found even when controlling for stimulus valence intensity, frequency, familiarity, and concreteness.

Testing the generality of valence asymmetry in similarity necessitated a new measure that is more efficient than pairwise judgment. Study 1 further validated such an efficient similarity measure: the spatial arrangement method (SpAM; Goldstone, 1994; Hout et al., 2013; Kriegeskorte & Mur, 2012) in which participant's task was to drag-and-drop similar and dissimilar stimuli closer together and further apart on the computer screen, respectively. Study 1 showed that SpAM similarity correlated strongly, $r = .84$, with similarity judged in pairs, and moderately, $r = .56$ and .64, with co-occurrence on webpages and in book passages (see Table 2), respectively. Thus, Study 1 generalized the construct validity of SpAM from visual and conceptually uniform verbal stimuli (see Hout et al., 2013) to conceptually diverse verbal stimuli. Further, Study 1 revealed that the predictive validity of SpAM and Pairwise similarity is comparably substantial, as both measures correlated with basic aspects of cognitive processing (i.e., evaluation speed, classification speed, and sensitivity and response bias in recognition memory; SpAM: $r = |.32|-|.62|$; Pairwise: $r = |.31|-|.68|$, Table 3).

Studies 2-5 then employed the efficiency advantage of SpAM (Hout et al., 2013) to test the generality of the proposed higher similarity of positive compared to negative stimuli in large, representative samples of participant-generated rather than researcher-selected words (see Figure 2). Study 2 generalized the proposed valence asymmetry in similarity from *self-generated, retrieved* to *other-generated, received* words, and showed that the receivers agreed with the retrievers on the valence of the positive and negative words that they had generated in > 98% of all cases. Thus, Study 2 examined words of consensual valence. Study 3 investigated whether people differentiate idiosyncratically positive stimuli while summarizing idiosyncratically negative stimuli (Denrell, 2005; Fazio et al., 2004; Smallman et al., 2014;

Smallman & Roese, 2008), which might result in a reversal of the valence asymmetry in similarity found for consensually positive and negative stimuli in Study 2. However, this reversal was not found. Instead, Study 3 generalized the proposed valence asymmetry in similarity from *consensually* to *idiosyncratically* positive and negative words. Study 4 generalized the valence asymmetry from words *generated by one other individual* to words *generated by many other people*. This result increases the range of validity of the asymmetry, as individuals receive positive and negative information from many independent rather than just one source. Further extending the validity of the asymmetry, Study 5 used a smartphone-based event-sampling method to show that it generalizes to self-experienced positive and negative everyday events.

Figure 2



Note. Spatially arranged distance of positive to other positive words and negative to other negative words in percentage of the screen diagonal, and effect sizes in Studies 1-5. Participants freely sampled positive and negative stimuli and spatially arranged them on a blank screen. Participants arranged positive words more densely (i.e. more similar) to one another than the negative words.

Finally, Study 6 operationalized dissimilarity in terms of absolute rating difference across three relevant aspects of affective impression (valence, arousal, and potency; see Osgood, Suci, & Tannenbaum, 1957), and compared the similarity of all positive and negative words in the WKB database (~14,000 items) by Warriner, Kuperman, and Brysbaert (2013), and all positive and negative pictures in the IAPS database (~1,000 items) by Lang, Bradley, and Cuthbert (2005). In contrast to Studies 2-4, these words and pictures are mainly of moderate and weak valence. Results nevertheless showed the proposed valence asymmetry in similarity. In sum, these six studies strongly supported the proposed *general* valence asymmetry in stimulus similarity.

Valence asymmetry in similarity is not a spurious effect

Affective and/or motivational influences during retrieval and spatial arrangement provide alternative explanations of the observed similarity asymmetries, which would then not be based on the factual difference in similarity between positive and negative information, but rather on psychological processes due to the information's affective/motivational potential. Across the studies, we believe there is good evidence that the similarity asymmetry exists independent of such affective and/or motivational influences.

Study 4 ruled out alternative explanations due to inclusive and exclusive sampling elicited by positive and negative affect elicited by the process of selecting several positive/negative stimuli, respectively (Bless & Fiedler, 2006; Forgas, 2013), as participants spatially arranged stimuli selected by as many retrievers as there were positive/negative words to be spatially arranged for similarity.

Moreover, the effect is unlikely to be based on an inclusive/exclusive style of spatially arranging positive/negative stimuli due to positive/negative affect. Participants in Studies 2-5 spatially arranged the positive and negative stimuli in a simultaneous fashion. With both positive and negative stimuli simultaneously in sight, rapid changes between cognitive styles (Topolinski & Deutsch, 2012; 2013) does not seem a likely explanation.

Importantly, the effect is not due to a motivation to move the aversively negative stimuli away from the attentional center and keep the pleasant positive stimuli in the center. This would create the observed pattern, as towards the edges of the screen, stimuli are, on average, further apart, and thus will be recorded as less similar to one another compared to the center of the screen. To test this possible alternative explanation, we computed the average distance of the Study 1-5 positive and negative stimuli to the center of the screen; Table 7 shows the results. As can be seen, across all SpAM studies and in each single SpAM study, participants spatially arranged the positive and negative words at equal distance to the center of the SpAM board. Participants did not position the positive information closer to the center, but positioned it closer together. In addition, Study 6 was not a SpAM study, but the higher similarity of positive compared to negative information was nevertheless found; and, as Study 1 showed, SpAM similarity correlates highly with other similarity measures, which should not be the case if our results are an artefact of the spatial arrangement method.

Table 7

Average distance of positive vs. negative stimuli from the midpoint of the spatial arrangement board in Studies 1-5 (standard deviations in parentheses).

	Positive valence	Negative valence	*F*	*p*	η_p^2	*90% CI LB*	*90% CI UB*
Overall	24.00 (7.62)	24.09 (7.57)	0.48	.83	.00	.00	,01
Study 1: Unkelbach et al., 2008	23.12 (7.45)	25.83 (7.61)	3.67	.06	.06	.00	.10
Study 2: self-generated stimuli	26.26 (8.88)	25.23 (6.73)	0.61	.44	.01	.00	.12
Study 2: other-generated stimuli	25.27 (5.97)	27.75 (6.78)	3.66	.06	.08	.00	.23
Study 3: consensual stimuli	27.03 (7.36)	25.54 (8.12)	1.97	.17	.04	.00	.10
Study 3: idiosyncratic stimuli	26.57 (7.36)	25.60 (6.55)	0.62	.43	.01	.00	.12
Study 4: independent stimuli	25.99 (6.82)	24.57 (6.83)	1.24	.27	.02	.00	.03
Study 5: real-life stimuli	19.81 (6.47)	20.13 (7.13)	0.25	.62	.00	.00	.01

Note. Values reflect the average distance of the positive or negative stimuli from the midpoint of the spatial arrangement board in relation to the screen diagonal, overall and separately for Studies 1 – 5. This positive-negative difference never reached statistical significance.

The affective and motivational potential of positive and negative stimuli has received much attention: negative stimuli are stronger (Baumeister et al., 2001), more dominant/contagious (Rozin & Royzman, 2001), and more mobilizing (Taylor, 1991) than positive stimuli. The observed valence asymmetry in similarity is not necessarily related to this valence asymmetry in affective potential. In fact, we found empirical evidence for this theoretical independence of valence asymmetry in similarity and affective potential: Studies 1, 2, 4, and 6 showed *presence of* valence asymmetry in similarity *in the absence of* valence asymmetry in rated affective potential[10] (we did not measure the valence intensity of the positive and negative stimuli examined in Studies 3 and 5). In any case, exploring the relation of similarity and affective potential is a fascinating topic for further research.

Finally, in Studies 1 and 4 we ran a regression analysis with the positive/negative words' within-valence similarity as the criterion and the positive/negative words' effect-coded valence, and their interval-scaled valence intensity, frequency, familiarity, and concreteness as predictors. In both Studies 1 and 4, results showed that valence predicted similarity even when simultaneously controlling for valence intensity, frequency, familiarity, and concreteness. These results suggest that the asymmetries in similarity observed in Studies 1-6 were actually due to valence, and *not* due to these factors possibly confounded with valence.

These alternative variables largely relate to the affective and motivational potential of evaluative information; that is, these variables should affect the processing of positive and

---

[10] First, the positive and negative words examined in Study 1 were found to be equally distant from the midpoint (5) of a 0-10 "negative-positive" scale (M_{pos} = 3.36 > 5, *SD* = 0.60; M_{neg} = 3.56 < 5, *SD* = 0.52; $F(1, 38) = 1.33$, $p = .26$, $\eta_p^2 = .03$, CI 90% [.00, .16]; Klauer & Musch, 1999). Second, in Study 2b, before spatially arranging the positive and negative words generated by another participant, receiver participants evaluated the words on a 1-7 "negative-positive" scale with the midpoint (4) labeled as "neutral". Participants provided equal distances from this midpoint for positive words (M_{pos} = 2.04 > 4, *SD* = 0.35) and negative words (M_{neg} = 2.05 < 4, *SD* = 0.35), $F(1, 42) = 0.01$, $p = .92$, $\eta_p^2 = .00$, CI 90% [.00, .01]. Further, across participants valence asymmetry in distance from the neutral midpoint did not correlate with valence asymmetry in SpAM similarity, $r(41) = .16$, $p = .29$. Third, using the database proved by Warriner, Kuperman, and Brysbaert (2013) we compared the valence of the positive and negative words examined in Study 4 (we omitted "myopic", because the WKB database does not contain this negative word). Again, the positive and negative words were found to be equally distant from the midpoint (5) of a 1-9 "unhappy-happy" scale (M_{pos} = 2.41 > 5, *SD* = 0.54; M_{neg} = 2.24 < 5, *SD* = 0.71; $F(1, 61) = 1.14$, $p = .29$, $\eta_p^2 = .02$, CI 90% [.00, .10]). And fourth, in Study 6 the positive and negative words were found to be equally distant from the midpoint (5) of the same 1-9 "unhappy-happy" scale (M_{pos} = 1.07 > 5, *SD* = 0.65; M_{neg} = 0.98 < 5, *SD* = 0.84; $F(1, 13913) = 62.24$, $p < .001$, $\eta_p^2 = .00$, CI 90% [>.00, .01]). The same was true for the positive and negative pictures examined in Study 6 (M_{pos} = 1.59 > 5, *SD* = 0.76; M_{neg} = 1.48 < 5, *SD* = 1.08; $F(1, 954) = 3.24$, $p = .07$, $\eta_p^2 = .00$, CI 90% [.00, .01]).

negative information. Showing that valence asymmetry in similarity exists independent of these influences increases our confidence in the two ecological rather than psychological explanations we proposed in the introduction. Again, we assumed that positive information is more similar to other positive information compared to negative information's similarity to other negative information, because (1) on most evaluatively relevant content dimensions positive, adequate states are flanked by both too little- and too much-negative states and thus are quantitatively more similar than negative states. Second, (2) positive information occurs more frequently ("positive events are more common (more tokens), but negative events are more differentiated (more types)", Rozin et al., 2010, p.536) and thus co-occurs more frequently compared to negative information. This ecologically higher frequency of co-occurrence leads to psychologically higher similarity via stronger association in memory. Having established that the proposed asymmetry is a general phenomenon, future research must directly test these two explanations.

Implications for cognitive processing

Similarity impacts learning, memory, and cognition in profound ways. For example, as shown in Study 1, stimuli that are more similar to one another are classified and evaluated faster, are more likely to be subsumed under a category, are more often confused with one another and thus harder to recognize (see Table 3). Also, as discussed, similar prime-target are processed faster/easier (e.g., McRae & Boisvert, 1998; Perea, Duñabeitia, & Carreiras, 2008). Prototypical stimuli (i.e., exemplars that are more similar to other exemplars of a category) are categorized more accurately (Nosoksky, 1986; 1988; Smith & Sloman, 1994), and generalizations of processing strategies, judgments and decisions to similar stimuli are more likely (Ames, 2004; Gräf & Unkelbach, 2015; Shepard; 1987; Tenenbaum & Griffiths, 2001).

The present studies showed that positive stimuli are generally more similar to one another than negative stimuli. Thus, given the broadness of similarity effects, this valence asymmetry in similarity should lead to valence asymmetries on a variety of levels of information processing, including evaluation, classification, categorization, judgment and decision making, prediction, recognition, and recall, and might provide a unitary explanation for a host of previous findings that are commonly explained in terms of the affective and motivational potential of evaluative information. And indeed, there is already evidence for valence asymmetries in cognitive processing caused by evaluative information's differential similarity (e.g., processing speed, likelihood of generalization, and memory accuracy; Alves et al., 2015; Gräf & Unkelbach, 2015; Unkelbach et al., 2008). A promising path of future research is thus to explore and reveal further valence asymmetries in cognitive processing that are due to the general valence asymmetry in similarity found here.

Conclusion

The density hypothesis (Unkelbach et al., 2008) claimed that positive information is mentally represented as more similar to one another than negative information. We investigated whether this proposed valence asymmetry in similarity is a general phenomenon. The present research provides a clear empirical answer: The proposed valence asymmetry in similarity is a general phenomenon that is reliably found for both self-generated, retrieved and other-generated, received information, for information of both consensual and idiosyncratic valence, for information received from both one and many sources, for both words and experienced everyday events, and for both verbal and visual information of strong, moderate, and weak valence. This difference in similarity is due to the valence, and not the valence intensity, frequency, familiarity, or concreteness of positive and negative stimuli. And, finally, the observed valence asymmetry in similarity may explain downstream valence asymmetries on many levels of cognitive processing.

Appendix A

German stimuli used in Study 1 with English translations

Study 1 original stimuli	Study 1 translations	Study 1 original stimuli	Study 1 translations
Positive	Positive	Negative	Negative
1. Baby	1. Baby	1. Alkoholismus	1. Alcoholism
2. Geburtstag	2. Birthday	2. Bomben	2. Bombs
3. Schmetterling	3. Butterfly	3. Krebs	3. Cancer
4. Kuchen	4. Cake	4. Kakerlake	4. Cockroach
5. Schokolade	5. Chocolate	5. Verbrechen	5. Crime
6. Blumen	6. Flowers	6. Tod	6. Death
7. Essen	7. Food	7. Krankheit	7. Disease
8. Freund	8. Friend	8. Scheidung	8. Divorce
9. Geschenk	9. Gift	9. Beerdigung	9. Funeral
10. Hawaii	10. Hawaii	10. Müll	10. Garbage
11. Urlaub	11. Holiday	11. Gewehre	11. Guns
12. Eiscreme	12. Ice cream	12. Hass	12. Hate
13. Kätzchen	13. Kitten	13. Hölle	13. Hell
14. Kino	14. Movies	14. Hitler	14. Hitler
15. Musik	15. Music	15. Abfall	15. Litter
16. Party	16. Party	16. Rezession	16. Recession
17. Pizza	17. Pizza	17. Steuern	17. Taxes
18. Erdbeere	18. Strawberry	18. Zahnschmerzen	18. Toothache
19. Sommer	19. Summer	19. Virus	19. Virus
20. Sonnenschein	20. Sunshine	20. Krieg	20. War

Note. Same stimuli as used by Unkelbach and colleagues (2008).

Appendix B

English stimuli used in Study 4 with German translations

Study 4 original stimuli	Study 4 translations	Study 4 original stimuli	Study 4 translations
Positive	Positive	Negative	Negative
1. Awesome	1. Fantastisch	1. Afraid	1. Ängstlich
2. Beautiful	2. Schön	2. Anger	2. Zorn
3. Beneficial	3. Vorteilhaft	3. Angry	3. Verärgert
4. Brilliant	4. Großartig	4. Blacklisted	4. Schwarzgelistet
5. Courage	5. Mut	5. Boring	5. Langweilig
6. Creation	6. Schöpfung	6. Capitalism	6. Kapitalismus
7. Ecstatic	7. Begeistert	7. Crazy	7. Verrückt
8. Energetic	8. Energetisch	8. Depressed	8. Depressiv
9. Enhanced	9. Verbessert	9. Depression	9. Depression
10. Excitement	10. Spannung	10. Destruction	10. Zerstörung
11. Exciting	11. Spannend	11. Disgusting	11. Ekelhaft
12. Fabulous	12. Fabelhaft	12. Dishonest	12. Unehrlich
13. Funny	13. Witzig	13. Fat	13. Fett
14. Generous	14. Großzügig	14. Harmful	14. Schädlich
15. Great	15. Toll	15. Hate	15. Hass
16. Happy	16. Glücklich	16. Horrible	16. Schrecklich
17. Healthy	17. Gesund	17. Hurt	17. Schmerz
18. Helpful	18. Hilfreich	18. Inefficient	18. Ineffizient
19. Hero	19. Held	19. Jerk	19. Trottel
20. Honest	20. Ehrlich	20. Junk	20. Schrott
21. Hope	21. Hoffnung	21. Lie	21. Lüge
22. Inspire	22. Inspirieren	22. Mean	22. Geheim
23. Love	23. Liebe	23. Murder	23. Mord
24. Morality	24. Moral	24. Mutilate	24. Verstümmeln
25. Motivated	25. Motiviert	25. Myopic	25. Kurzsichtig
26. Optimism	26. Optimismus	26. Obnoxious	26. Unausstehlich
27. Promising	27. Vielversprechend	27. Poor	27. Arm
28. Smile	28. Lächeln	28. Quit	28. Aufgeben
29. Wonderful	29. Wunderbar	29. Rude	29. Unhöflich
		30. Shallow	30. Oberflächlich
		31. Sickness	31. Krankheit
		32. Ugly	32. Hässlich
		33. Unpleasant	33. Unangenehm
		34. Wound	34. Wunde
		35. Wretched	35. Erbärmlich

Note. New stimuli generated by participants in Study 4. Redundant stimuli are not displayed.

Chapter 3 – The ABC of stereotype about groups: Agency/socio-economic success, conservative-progressive beliefs, and communion

Abstract

Previous research argued that stereotypes differ primarily on the two dimensions of warmth / communion and competence / agency. We identify an empirical gap in support for this notion. The theoretical model constrains stereotypes a priori to these two dimensions; without this constraint, participants might spontaneously employ other relevant dimensions. We fill this gap by complementing the existing theory-driven approaches with a data-driven approach that allows an estimation of the spontaneously employed dimensions of stereotyping. Seven studies (total N = 4451) show that people organize social groups primarily based on their agency / socio-economic success (A), and as a second dimension, based on their conservative-progressive beliefs (B). Communion (C) is not found as a dimension by its own, but rather as an emergent quality in the two-dimensional space of A and B, resulting in a two-dimensional ABC model of stereotype content about social groups.

Stereotypes are everywhere. To navigate their social world, people quickly group individuals in meaningful social categories based on their age, gender, ethnic origin, occupation, or interest (Brewer, 1988; Fiske & Neuberg, 1990; Tajfel, 1969). Knowledge about these categories includes what typical members of this category are like, think, feel and do, and the schematic application of this knowledge provides an economical alternative to effortful individuation (Fiske & Pavelchak, 1986; Gilbert & Hixon, 1991; Macrae, Milne, & Bodenhausen, 1994). And stereotypes matter. They allow people to go beyond the information given (Bruner, 1957), make predictions about the future behavior of individuals based on their sheer category membership (Hamilton, Sherman & Ruvolo, 1990), and they influence people's judgments, decisions and behavior in a stereotype-consistent way (Wheeler & Petty, 2001), even without being aware of this (Bargh, Chen, & Burrows, 1996).

Each stereotype consists of a more or less unique set of attributes associated with the social group: White, Black, Latino, Middle Eastern, and Asian men are "rich", "athletic", "macho", "bearded", and "intelligent", respectively. White, Black, Latino, Middle Eastern, and Asian women are "arrogant", "have an attitude", are "feisty", "quiet", and "intelligent", respectively (Ghavami & Peplau, 2012, pp. 118-120). Librarians are shy, hairdressers are flamboyant, and stock-traders are greedy. Some attributes, however, may be of greater importance for effectively coordinating social behavior than others and thus are likely to serve as content of stereotypes about many, if not all, groups. That is, some attributes may serve as fundamental dimensions of stereotype content that stretch out people's social maps on which groups can be located as a function of scoring low or high on the respective dimensions.

Warmth and competence are meaningful stereotype content dimensions

According to the stereotype content model (SCM; Fiske, Cuddy, Glick, & Xu, 2002), the most relevant criteria in intergroup interaction are the social groups members' intentions and their ability to carry out their plans. The central question is whether a group has goals compatible with the perceiver and is thus likely to help him or her, or whether it has antagonistic goals and thus might harm him or her (Fiske et al., 2002). Knowing this (i.e., a group's warmth, Fiske et al., 2002; communion, Abele & Wojciszke, 2007; morality, Wojciszke, 1994; other-profitableness, Peeters, 1983; trustworthiness, Oosterhof & Todorov, 2008), the second most relevant question has been theorized to be a group's ability to carry out their intentions (i.e., competence, Fiske et al. 2002; agency, Abele & Wojciszke, 2007; Wojciszke, 1994; self-profitableness, Peeters, 1983; instrumentality, Parson & Bales, 1955).

More than a decade of research on these two dimensions of the SCM (Fiske et al., 2002) suggests that whether a group is perceived as warm and / or competent has implications for emotional reactions to the group (Cikara & Fiske, 2012; Cuddy, Fiske, & Glick, 2007), neurological responses to the group (Harris & Fiske, 2006), people's perception of what typical group members look like (Imhoff, Woelki, Hanke, & Dotsch, 2013), as well as behavioral intentions of harming and helping (Becker & Asbrock, 2012; Cuddy et al., 2007) like invitations to a job interview (Agerström, Björklund, Carlsson, & Rooth, 2012), or support for immigration politics (Reyna, Dobria, & Wetherell, 2013). Even beyond groups, the two SCM dimensions have been employed to assess people's perceptions of brands (Aaker, Garbinsky, & Vohs, 2012; Kervyn, Chan, Malone, Korpusik, & Ybarra, 2014), exonerees (Clow & Leach, 2015), and individuals in pain (Ashton-James, Richardson, Williams, Bianchi-Berthouze, & Dekker, 2014).

Some of these studies adopt what we would call a *relational* approach and aim to explore how individuals determine their concrete behavior toward an individual from a group based on their assumptions about this group's warmth and competence (e.g., Kervyn, Dolderer, Mahieu, & Yzerbyt, 2010). Others have adopted what might be framed as a *lay sociologist* perspective, that is: on which dimensions do people identify the most relevant differences between social groups (Cuddy et al., 2007; Fiske et al., 2002; Imhoff et al., 2013). The present research addresses particularly the latter perspective.

Warmth and competence may not be the dimensions that people spontaneously use

Within the lay sociologist perspective we argue that although warmth and competence are meaningful dimensions of stereotype content, we currently lack empirical support for the notion that these are indeed the dimensions that individuals *spontaneously* employ when making sense of social groups. Spontaneously employed dimensions are the ones that come to people's mind without theoretical constraints made by the researchers. Most studies on stereotype content constrain participants to the two theoretically derived dimensions, because in most cases only these two dimensions are rated (e.g., "participants rated the 15 groups on scales of warmth, competence, status, and competition", Cuddy et al., 2009, p. 12; "participants rated the 53 categories on either competence or warmth", Durante, Volpato, & Fiske, 2010, p. 473; "participants rated the groups on scales reflecting warmth, competence, perceived status, and perceived competition", Fiske et al, 2002, p. 884; "participants rated the extent to which each group appeared warm (friendly, cold (reversed), likable […]) and competent (capable, incompetent (reversed), smart […])", Bergsieker, Leslie, Constantine, & Fiske, 2012, p. 1229). For participants, it is thus impossible to employ any other stereotype content dimensions.

Another source of constraints is the selection of groups to be rated. Although some studies sampled groups spontaneously named by participants, the instructions prompted race, gender, occupation and so forth as criteria of what constitutes groups, thereby biasing the likelihood of certain categories to be named (e.g., 'Blacks', 'women', and 'professionals', see "off the top of your head, what various types of people do you think today's society categorizes into groups (i.e., based on ethnicity, race, gender, occupation, ability, etc.)?", Fiske et al., 2002, p. 883; see also Kervyn, Fiske, & Yzerbyt, 2013, p. 676). It is conceivable and highly likely that a biased sample of certain social groups will make certain stereotype dimensions more salient than others (e.g., prompting race and gender will make dimensions associated with race and gender more salient).

To give another example, Imhoff and colleagues (2013) showed that visual facial representations of typical exemplars of two social groups pretested as differing on warmth and competence were judged by other raters as differing on both warmth and competence. While this finding supports that people are able to associate warmth and competence with facial features, it does not rule out that the space of group stereotypes also includes one, two, or more additional – and potentially more fundamental – dimensions that were not visually encoded in the faces because the two pretested groups (managers and kindergartners) were not different on them, and / or that were not decoded from the faces because the researchers never asked for ratings other than warmth and competence.

Thus, the above-mentioned studies lack representative design (Brunswik, 1955; 1956). To illustrate this important aspect, imagine one wants to find out the fundamental dimensions people spontaneously use to compare cars. A non-representative sample of cars of the same price, size, and fuel efficiency, but in different colors will probably prompt the result that the most fundamental dimension on which people spontaneously distinguish cars is their color. While that might very well be the case, the biased sampling prevented other dimensions from being detected because there was no meaningful variance on these other dimensions. Likewise, even if there is a representative sample of cars, but participants rate them only on the number of airbags and the maximum speed, this will give us a two-dimensional space on which all cars can be positioned, with one dimension being number of airbags, and the other being maximum speed. Crucially, though, we have no empirical base to judge whether these two dimensions are indeed the fundamental dimensions that individuals spontaneously employ when comparing cars even if we replicate the rating multiple times in many different environments. Without a more representative sampling approach, one cannot rule out that empirical findings are influenced by sampling biases (Fiedler, 2011). As much as we ideally draw representative participant samples from the population we aim to generalize to, a representative design also calls for an unbiased sampling of stimuli (to be able to generalize to the universe of stimuli) as well as dimensions (to generalize to the universe of attributes; see Wells & Windschitl, 1999; Westfall et al., 2014; for a more elaborate discussion of the problems of stimulus sampling and generalization).

In summary, we believe the available evidence for the nature of stereotype content dimensions about social groups suffers from a) a non-representative sampling of social groups, which prevents generalization to the population of groups, and b) a non-representative sampling of rated attributes, which prevents generalization to the population of all conceivable attributes.

How to explore the nature of spontaneous stereotype content about groups

To gain insights into the fundamental, spontaneously employed dimensions of stereotypes about groups, one thus needs a different approach that more closely follows the ideal of a representative design (Brunswik, 1955, 1956). In such a design, a sample of participants organizes a random (i.e., without any theoretical constraints) sample of stimuli on dimensions without being constrained in what these dimensions are. Sampling of groups can be achieved by asking people to name groups and selecting the most frequently named ones. In doing so, we avoid theory-driven a priori assumptions about the most relevant criteria for segmenting society into groups, such as age, sex, race, occupation, ability etc.

Assessing fundamental dimensions on which people align social groups without influencing participants by naming theoretically derived candidate dimensions requires more effort. Here, we rely on a data-driven strategy; such data-driven methods have proven to be extremely successful tools to identify fundamental dimensions of social perception with as little bias as possible in areas like face and gender perception (Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; Deaux & Lewis, 1984; Ghavami & Peplau, 2012; Todorov, Dotsch, Wigboldus, & Said, 2011; Williams & Best, 1990).

One well-established data-driven method is multidimensional scaling based on global dissimilarity estimates (Nosofsky, 1992; Schiffman, Reynolds, & Young, 1981). In this approach, participants merely provide estimates of the similarity / dissimilarity between social groups. Importantly, they are free to rely on any dimension that spontaneously comes to their mind and seems most diagnostic to them for that decision. When judging for instance the similarity between lawyers, nurses, and maids, individuals could resort to relatively consensual impressions of warmth and thus see lawyers and maids as similar (cold), but both different from nurses who are seen as warm (see Fiske & Dupree, 2014). If competence, however, is the most salient and subjectively diagnostic dimension, participants should see lawyers and nurses as similar compared to the dissimilar (relatively incompetent) maids. Finally, it is conceivable that people make use of completely different characteristics and see assumed gender as more central, with nurses and maids as occupations typically perceived to be female-dominated compared to lawyers evoking associations with men.

Exploring the dimensionality of stimulus spaces in this way is well established in the social psychology of personality impressions (good-bad x hard-soft; Rosenberg, Nelson, & Vivekananthan, 1968), emotions (valence x intensity; Russell, 1980; Shaver, Schwartz, Kirson, & O'Connor, 1987), animals (size x ferocity; Henley, 1969), power strategies (rationality x directness; Falbo, 1977), and responses to dissatisfaction in the job and one's relationship (active-passive x constructive-destructive; Farrell, 1983; Rusbult & Zembrodt, 1983). More relevant to the focus of the present paper, Pattyn, Rosseel, and van Hiel (2013) recently asked participants to complete a hierarchical sorting task to estimate dissimilarities between individuals who belonged to predefined social groups. Across three studies they reported converging support for five to six meaningful dimensions of the social group space (conventional vs. alternative, old vs. young, male vs. female, cognitive vs. physical, deviant vs. non-deviant, and to a lesser extent: cold vs. warm).

Although these results are thought-provoking as they suggest very different dimensions than the well-received stereotype content model, a closer look at their stimulus sampling procedure indicates that, as in all previous work, biased sampling might have again played a major role in producing these findings (for another example of the large impact of stimulus sampling on results see Frable, 1993; Jones & Ashmore, 1973). Specifically, the researchers searched for pictures of (male and female) individuals who belonged to a predefined set of social groups, among them 'punk', 'hippie', 'yuppie', 'typical woman', and 'senior citizen'. Accordingly, two of the central dimensions turned out to be conventional (typical woman) versus alternative (punks, hippies) as well as old versus young and a similar argument can be made for the other dimensions. This study thus illustrates how stimulus sampling may influence the inferred underlying dimensions.

The present research

This paper aims to investigate the fundamental, spontaneously employed dimensions of stereotype content about social groups. To achieve this aim, we followed the proposed data-driven research strategy. We asked participants in two cultural contexts (U.S.-based MTurkers and German students) to name examples of what constitutes groups without biasing the selection by any examples or criteria. The groups that were most frequently named and appeared most often in contemporary mass media were then judged on dissimilarity to one another in order to compute stereotype maps of groups with multidimensional scaling. The dimensions of the emerging scaling solutions were then interpreted via property fitting analyses (Chang & Carroll, 1969) with a variety of candidate stereotype content dimensions on which the groups had been judged by independent raters. As these candidate dimensions may constitute an experimenter influence, we finally asked participants to label all rotated content dimensions that run through the origin of the groups' stereotype maps. Other independent raters confirmed that these labels did not reflect a dimension that was not included in our selection of candidate stereotype content dimensions. We believe this strategy avoided biases due to selective sampling of stimuli and / or dimensions and allowed participants to spontaneously employ any dimension they saw as important to distinguish between the groups that they saw as important to distinguish. In a total of seven studies with 4451 participants, we found, confirmed, and generalized what we refer to as the 2D ABC model of spontaneous stereotypes about groups. According to the data, people distinguish groups based on differences in agency / socio-economic success (A; 'powerless-powerful', 'poor-wealthy', 'low status-high status', 'dominated-dominating', 'unconfident-confident', and 'unassertive-competitive') and conservative-progressive beliefs (B; 'traditional-modern', 'religious-science-oriented', 'conventional-alternative', and 'conservative-liberal'). Further, the groups' communion / warmth (C; 'cold-warm', 'untrustworthy-trustworthy', 'dishonest-sincere', 'repellent-likable', 'threatening-benevolent', and 'egoistic-altruistic') emerges as a function of centrality in the stereotype map spanned by A and B. That is, groups that appear average on both dimensions appear to be warm, trustworthy, sincere, likable, benevolent, and altruistic. Just like the stereotype content model by Fiske and colleagues (2002; see also Cuddy et al., 2007), the 2D ABC model addresses consensual rather than idiosyncratic group stereotypes.

We conducted five more studies within this project that we do not report for reasons of brevity. All studies consistently supported the pattern of results reported in this manuscript. These twelve studies represent the full set of all studies we have conducted up to this point to explore the number and nature of the stereotype content dimensions that people spontaneously employ to distinguish large sets of social groups sampled without bias in favor of a specific stereotype content model.

Study 1

We first generated a large sample of social groups by asking people to name groups and then selected the most frequently named ones (consensus > 10%). Then, new participants judged the dissimilarity between each group and each other group, allowing participants to spontaneously choose dimensions on which they base their judgment (Forgas, 1976; Rosenberg et al., 1968). Dissimilarity per se is unspecific and open to idiosyncratic interpretation – that is, it needs to be construed in one or another respect (Medin, Goldstone, & Gentner, 1993). The chosen dimensions might be different for each participant, but highly idiosyncratic approaches will be filtered out by aggregation across individuals so that the average pairwise estimates of the dissimilarities between the groups will reflect a consensual view. The dimensions might be different for each pairwise comparison and each dissimilarity rating might be a judgment based on the integration of many dimensions. However, as long as all participants employ more or less identical dimensions in making the dissimilarity judgments, the multidimensional scaling (MDS; for a review, see Borg & Groenen, 2005) algorithm will compute a multidimensional social space in which the groups' coordinates retain almost all the variance contained in the original dissimilarity judgments.

If there are fundamental stereotype content dimensions, then the next question is their nature, which can be addressed with a property fitting analysis (ProFit, Chang & Carroll, 1969; e.g., Pattyn et al., 2013) during which rating dimensions are sought that can be best predicted by the social groups' MDS coordinates. This approach is ideal to "help systematize data in areas where organizing concepts and underlying dimensions are not well-developed" (Schiffman, Reynolds & Young, 1981, p. 3, see also Giguère, 2006). The properties to be fitted were 24 trait dimensions ('unfriendly-friendly', 'incompetent-competent', etc.) that were identified as possible candidates of being fundamental to stereotype content, both in light of the data as well as established theories. While we diverge here from a purely data-driven approach, 24 dimensions present a much larger sample of possible candidates than in previous studies on the dimensionality and nature of spontaneous stereotype content about groups (e.g., Fiske et al., 2002). Studies 5 and 6 will solve this deviation from a purely data-driven approach and show that our selection of candidates included all stereotype content dimensions that participants employed to distinguish between groups. At this point we refrained from making predictions regarding the existence, number, and nature of the fundamental stereotype content dimensions for groups.

Methods and results

To avoid having an overly homogenous sample of undergraduate students, we recruited a more diverse sample in terms of educational and professional background as well as age, via Amazon's crowdsourcing platform Mechanical Turk.

Study 1a: Naming social groups. We paid 213 people (101 women, 112 men; M = 34.41 years, SD = 11.02) $1.5 to "name 40 social groups". Importantly, we refrained from recommending sampling strategies to get at people's naive understanding of groups (for a different approach, see Fiske et al., 2002). In the upper half of the screen, people read "Dear participant, each society is not only made up by the individuals that live in the society, but these individuals also constitute what we call 'social groups'. People belong to social groups either because they have a specific characteristic that is seen as typical for a social group or because they have chosen to become part of a social group. Thus, some social groups are based on how people are, while others are based on how people behave or see the world. (These groups do not have to be mutually exclusive in the sense that being part of one social group means one cannot also be part of another social group.) Although this definition may sound very abstract to you, you probably have examples of social groups in your mind. We ask you to name 40 social groups that spontaneously come to your mind. Just think for a moment of the groups that structure society and name 40 of them." In the bottom half, people entered 40 groups into 40 text boxes.

Table 1 shows all 80 social groups named by more than 10% of people in Study 1a. Apparently people selected groups based on race or ethnicity (Whites, Blacks, Asians), social class (Poor, Middle class, Rich), and political or religious beliefs (Democrats, Atheists, Republicans, Christians). The combination of these 80 groups results in 3160 possible pairs for which we collected dissimilarity judgments in Study 1b.

Table 1

Most Frequently Named Social Groups in the U.S. (Consensus > 10%) in Study 1a.

1st - 20th most frequent	21st - 40th most frequent	41st - 60th most frequent	61st - 80th most frequent
Whites (66%)	Teenagers (28%)	Buddhists (19%)	Upper class (14%)
Democrats (51%)	Muslims (27%)	Working class (19%)	Military (14%)
Blacks (48%)	Politicians (27%)	Young (19%)	Religious (14%)
Poor (47%)	Catholics (26%)	Elderly (18%)	Techies (14%)
Middle class (45%)	Gays (26%)	Hipsters (18%)	Sports fans (13%)
Asians (45%)	Men (25%)	Actors (18%)	Heterosexuals (13%)
Rich (44%)	Teachers (25%)	Homeless (17%)	Lower class (13%)
Atheists (42%)	Children (25%)	Libertarians (17%)	Drug users (12%)
Republicans (41%)	Goths (24%)	Independents (17%)	Employed (12%)
Christians (37%)	Jocks (22%)	Mexicans (17%)	Hindu (12%)
Liberals (36%)	Parents (22%)	Businesspeople	Lawyers (12%)
Conservatives (35%)	Hippies (22%)	Educated (16%)	Straight (12%)
Nerds (34%)	Doctors (21%)	White collar (16%)	Families (12%)
Students (33%)	Adults (21%)	Indians (16%)	Lesbians (12%)
Athletes (31%)	Blue collar (21%)	Old (16%)	Skaters (12%)
Jews (30%)	Geeks (21%)	Bisexuals (14%)	Stoners (12%)
Hispanics (30%)	Preps (21%)	Criminals (14%)	Agnostics (11%)
Women (30%)	Scientists (20%)	Homosexuals (14%)	Latinos (11%)
Artists (29%)	Americans (19%)	Immigrants (14%)	Rednecks (11%)
Musicians (29%)	Gamers (19%)	Unemployed (14%)	Tea Party (11%)

Note. Percentage in parentheses is proportion of participants who spontaneously named this group as a social group that is representative of the structure of U.S. society.

Study 1b: Multidimensional scaling of 80 groups. We paid 843 other people (420 women, 423 men; M = 36.33 years, SD = 12.65) \$0.6 to “rate the similarity-dissimilarity of 80 pairs of social groups”. Multidimensional scaling operates on the stimulus level (here: groups); as it was not feasible to do all pairwise dissimilarity comparisons, we presented each participant with 80 randomly selected pairs of stimuli out of the full 3160 pairs of stimuli and averaged the ratings on the stimulus level. On the first screen slide, they read “Dear participant, please rate the similarity-dissimilarity of these two social groups”. Below, they used a 9-point ‘very similar-very dissimilar’ scale to rate the two randomly selected groups. On the next screens, people rated 79 other randomly selected pairs of groups.

On average, each of the 3160 dissimilarities was judged by M = 20.94 participants, SD = 4.77. We subjected the full matrix of 3160 mean pairwise dissimilarities to multidimensional scaling (MDS; for a review, see Borg & Groenen, 2005). We used the ALSCAL procedure (Young, Takane, & Lewyckyj, 1978); assuming an interval scale, we estimated coordinates for the 80 social groups in dissimilarity spaces in which Euclidean distances can be interpreted as dissimilarity. The further apart two groups are in these spaces, the more dissimilar people judged them to be. We estimated coordinates for six MDS solutions, varying from a one-dimensional to a six-dimensional dissimilarity space.

There are two indicators of goodness of scaling fit: scaling stress (S; should be preferably low) and the proportion of original dissimilarity variance accounted for by the scaling solution (R^2; should be preferably high). Table 2 shows S and R^2 for the six scaling solutions (1D, 2D, 3D, 4D, 5D, and 6D). Balancing goodness of scaling fit and ease of interpretation (Jaworska & Chupetlovska-Anastasova, 2009), we proceeded with the social groups' 1D, 2D, and 3D dissimilarity spaces. The scree plots of S and 1-R^2 showed that extracting a fourth, fifth, and sixth dimension only slightly improved S and R^2. Next, we inspected the corresponding scatter plots, searching for and selecting a number of candidate stereotype content dimensions deemed suitable to interpret the 1D, 2D, and 3D space. These data-driven candidates were augmented with candidate dimensions derived from the main theories of stereotype content (e.g., Fiske et al., 2002). Appendix A shows all 24 candidate dimensions.

Table 2

Goodness of 1D, 2D, 3D, 4D, 5D, and 6D Scaling Fit in Studies 1-6

	Method	People	Groups	*1D*	*2D*	*3D*	*4D*	*5D*	*6D*
Study 1	sequential dissimilarity judgment	U.S.	80	.23 0.57	.19 0.73	.16 0.81	**.15** 0.84	**.14** 0.86	**.13** 0.88
Study 2	simultaneous dissimilarity arrangement	U.S.	80	.16 0.62	**.13** 0.78	**.11** 0.85	**.10** 0.89	**.09** 0.91	**.08** 0.92
Study 3	simultaneous dissimilarity arrangement	German	76	.16 0.75	**.12** 0.87	**.10** 0.91	**.09** 0.93	**.08** 0.95	**.07** 0.96
Study 4	simultaneous rating on pre-specified	U.S.	80	**.14** 0.72	**.13** 0.80	**.12** 0.84	**.11** 0.86	**.10** 0.88	**.10** 0.89
Study 5	simultaneous dissimilarity arrangement	U.S.	42 minimal.	**.12** 0.75	**.09** 0.87	**.07** 0.93	**.06** 0.95	**.04** 0.97	**.04** 0.98
Study 5	simultaneous dissimilarity arrangement	U.S.	61 natural.	**.13** 0.70	**.09** 0.86	**.08** 0.91	**.06** 0.94	**.06** 0.95	**.05** 0.96
Study 6	simultaneous dissimilarity arrangement	U.S.	42 minimal.	**.14** 0.85	**.10** 0.93	**.08** 0.95	**.07** 0.97	**.06** 0.98	**.05** 0.98
Study 6	simultaneous dissimilarity arrangement	U.S.	61 natural.	.16 0.71	**.13** 0.82	**.10** 0.88	**.09** 0.90	**.09** 0.92	**.08** 0.93

Note. Upper values indicate scaling stress (for a review, see Borg & Groenen, 2005). Lower values indicate percent of original variance retained in the scaling solution. According to Kruskal and Wish (1978), stress <= .20, <= .15, <= .10, <=.05 and <= .025 may be interpreted as poor, sufficient, satisfactory, good and excellent, respectively. Bold values are stress =< .15, which are sufficient. In all studies except Study 1, the 2D scaling solution achieved a sufficient low stress.

Study 1c: Disambiguating the dissimilarity ratings. Finally, 620 people (275 women, 336 men and 9 unassigned; *M* = 34.94 years, *SD* = 12.17) were paid $1 to "rate 80 social groups on a stereotype dimension" (e.g., 'unfriendly-friendly'). On the first screen slide, they read "Dear participant, some kind of people in our society are [friendly], while other kind of people in our society are [the opposite stereotype; unfriendly]. Please rate the following 80 social groups according to how [friendly] or [unfriendly] they are". People then used 0-100 slider scales to rate the groups in a random order, one below the other on the same screen slide. There were between 22 and 27 raters per candidate stereotype content dimension. Raters' agreement about the groups was very high, *ICC(2,k)* >= .84, for all 24 candidate stereotype content dimensions (McGraw & Wong, 1996; Shrout & Fleiss, 1979).

To facilitate the interpretation of the social groups' 1D, 2D, and 3D dissimilarity space, we ran principal component analyses (PCA; Jolliffe, 2002) on the 24 candidate stereotype content dimensions, using varimax rotation. First, we determined the number of components to be extracted. The first, second, third, fourth, fifth and sixth component explained 35%, 28%, 15%, 5%, 4%, and 3% of the total variance, respectively. Based on the scree plot, we proceeded with the extraction of three components. Aiming for simple structure, we omitted all eight candidate stereotype content dimensions that had no primary factor loading of ≥ .75 and / or a cross-loading of ≥ .45 on any of the three components. The eight omitted dimensions were: 'incompetent-competent', 'unintelligent-smart', 'masculine-feminine', 'communal-individualistic', 'typical (in the U.S.)-unusual (in the U.S.)', 'unfriendly-friendly', 'intolerant-tolerant', and 'unable-skillful'. The third step validated the simple structure and no more omissions of candidate stereotype content dimensions were necessary.

Table 3 shows the varimax rotated component loadings of the 16 candidate stereotype content dimensions retained in this solution. Based on these component loadings, we composed the three combined candidate stereotype content dimensions *agency / socio-economic success* (A; 'powerless-powerful', 'poor-wealthy', 'low status-high status', 'dominated-dominating', 'unconfident-confident', and 'unassertive-competitive'; α = .955), *conservative-progressive beliefs* (B; 'traditional-modern', 'religious-science-oriented', 'conventional-alternative', and 'conservative-liberal'; α = .900), and *communion* (C; 'cold-warm', 'untrustworthy-trustworthy', 'dishonest-sincere', 'repellent-likable', 'threatening-benevolent', and 'egoistic-altruistic'; α = .953)[1]. For short, the analysis yielded the dimensions A, B, and C. A and B were

[1] Each subdimension of our 3 *combined* candidate stereotype content dimensions was rated by different people. It might be argued that valid estimates of the social groups' A, B, and C requires judging the groups on all subcomponents of A, B, and C at once, as the whole is more than the sum of its parts. To address this possibility, 79 MTurkers (36 women, 43 men; *M* = 32.46 years, *SD* = 10.04) were paid 1$ to rate the 80 groups on *compound* items of A, *n* = 25, *ICC(2,k)* = .96, B, *n* = 27, *ICC(2,k)* = .95, or C, *n* = 27, *ICC(2,k)* = .92. Each of the corresponding 0-100 slider scale items was anchored with a meaning cloud of all sub-dimensions that are included in the combined items. All compound items showed very high convergence with the combined items, *r*s ≥ .97, *p*s < .001, and all analyses reported below led to identical conclusions if compound rather than combined items were used.

almost but not entirely orthogonal, $r = -.29$, $p < .01$; A and C were orthogonal, $r = .07$, $p = .55$; B and C were orthogonal, $r = -.01$, $p = .90$.

Table 3

Factor Loadings and Interpretation of the 16 Retained Dimensions in Study 1c

Candidate stereotype content dimension	1st component: Agency (A)	2rd component: Beliefs (B)	3nd component: Communion (C)
Powerless-Powerful	**.940**	-.097	.102
Dominated-Dominating	**.928**	-.205	-.150
Low status-High status	**.924**	-.097	.284
Poor-Wealthy	**.905**	-.015	.019
Unconfident-Confident	**.873**	-.174	.034
Unassertive-Competitive	**.808**	.032	-.247
Traditional-Modern	-.124	**.964**	.143
Religious-Science-oriented	.313	**.855**	.044
Conventional-Alternative	-.417	**.819**	-.234
Conservative-Liberal	-.445	**.815**	.141
Untrustworthy-Trustworthy	.081	-.014	**.953**
Dishonest-Sincere	-.022	.025	**.936**
Repellent-Likable	.226	.033	**.913**
Threatening-Benevolent	.167	.103	**.910**
Cold-Warm	-.178	.068	**.909**
Egoistic-Altruistic	-.417	-.067	**.790**

Note. Bold factor loadings are significant at p =< .001.

To compare the suitability of A, B, and C for interpreting the social groups' 1D, 2D, and 3D dissimilarity space, we carried out a series of nine multiple linear regressions with the groups' mean A, B, and C as criterion and their x-, x- / y-, and x- / y- / z-coordinates in the 1D, 2D, and 3D space as predictors, respectively (Forgas, 1976; Rosenberg et al., 1968; Shaver et al., 1987). Figuratively speaking, each of these nine property fitting analyses (ProFit; Chang & Carroll, 1969; e.g., Pattyn et al., 2013) finds out how much the location of groups in the 1D, 2D, or 3D dissimilarity spaces can be mapped onto either A, B, or C by means of rotating the dissimilarity spaces. Ideally, in the 1D group space, consisting of one axis, the groups' coordinates (consisting of scores on the single axis) correlate as high as *R(1D axis)* = 1 with only A, B, or C. For example, if *R(1D axis)* = 1 for A, but *R(1D axis)* = 0 for B and C, then A is maximally suitable as an axial interpretation of the groups' 1D space, accounting for 100%

of the dissimilarity variance in this space. To account for 100% of the dissimilarity variance in the groups' 2D space, consisting of two axes, two *orthogonal* candidate stereotype content dimensions with *R(2D axis)* = 1 need to be fitted, because each axis should map onto one of the two stereotype content dimension. If so, these two can be interpreted as the two primary independent stereotype content dimensions on which people spontaneously judged the dissimilarities between the 80 social groups. This reasoning can be generalized to higher dimensions (e.g., 3 axes).

Table 4 shows the results[2]. The higher a multiple correlation *R(1D axis), R(2D axis),* and *R(3D axis),* the more suitable was the corresponding candidate stereotype content dimension as an *axial* interpretation of the 1D, 2D, and 3D social group space, respectively. In Study 1, A, *R(2D axis)* = .72, $p < .001$, and B, *R(2D axis)* = .91, $p < .001$, were almost maximally suitable axial interpretations of the 2D group space (see also Figure 1), whereas C was not suitable as an axial interpretation of the 2D space, *R(2D axis)* = .23, $p = .13$.

[2] The results of (24 * 3 =) 72 multiple linear regressions with the social groups' means on each of the 24 candidate stereotype content dimensions as the criterion and the groups' coordinates in their 1D, 2D and 3D dissimilarity spaces as predictors are shown in Appendix A, and are consistent with Table 4.

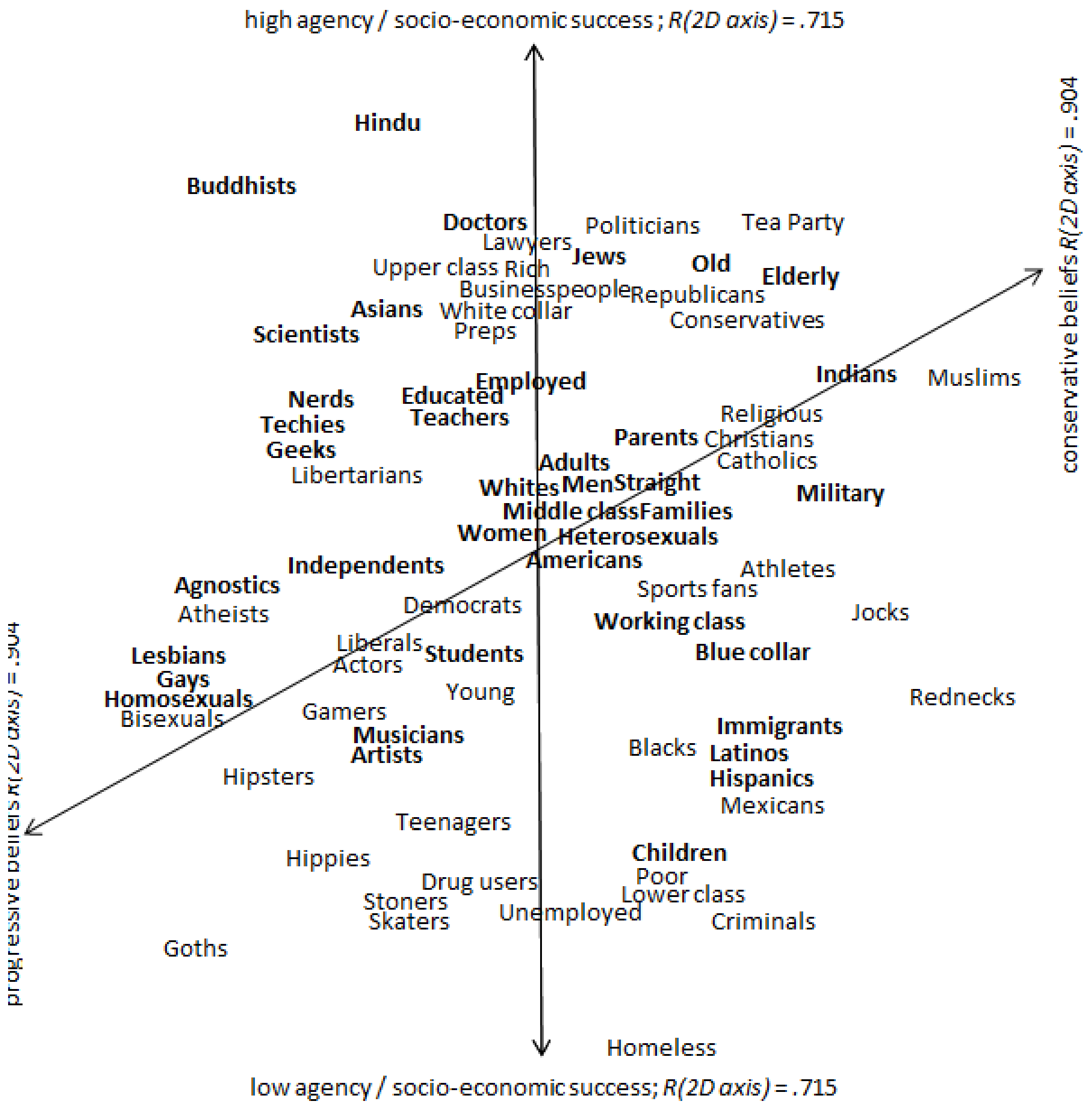
Table 4

Property Fitting Results for Studies 1-4

	Group sample	Stereotype content	*R(1D axis)*	*R(2D axis)*	*R(3D axis)*	*r(1D pole)*	*r(2D pole)*	*r(3D pole)*
Study	80 U.S.	Agency (A)	**.615**	**.715**	**.842**	**.419**	**.456**	**.426**
		Beliefs (B)	**.677**	**.904**	**.935**	-.116	-.060	-.037
		Communion (C)	.162	.229	**.468**	**.358**	**.406**	**.415**
Study	80 U.S.	Agency (A)	**.766**	**.812**	**.898**	.069	.183	.234
		Beliefs (B)	**.720**	**.812**	**.954**	-.035	.076	.017
		Communion (C)	.071	.175	.236	**.562**	**.580**	**.581**
Study	76 German	Agency (A)	**.821**	**.903**	**.909**	.283	.272	.334
		Beliefs (B)	.257	**.857**	**.831**	**-.372**	-.187	-.174
		Communion (C)	**.502**	**.464**	**.479**	**.783**	**.745**	**.727**
Study	80 U.S.	Agency (A)	**.856**	**.893**	**.893**	.221	.208	.222
		Beliefs (B)	.339	**.848**	**.890**	.015	.009	.010
		Communion (C)	.205	.375	**.622**	**.482**	**.506**	**.477**

Note. *R(1D-3D axis)* indicate the maximal correlations between the 80/76 U.S./German social groups' agency / socio-economic success, conservative-progressive beliefs, and communion ratings and their projections on an axis rotated around the origin of their 1D, 2D, and 3D dissimilarity space in Studies 1-4; *r(1D-3D pol)* indicate correlations between the groups' A, B, and C ratings and their proximity to the origin of these spaces. Bold correlations are significant at p =< .001.

Figure 1



Note. Study 1 (U.S. participants and target social groups): This 2D space of 80 representatively sampled social groups was computed based on pairwise dissimilarity ratings, and can be interpreted by agency / socio-economic success and conservative-progressive beliefs. Communion emerges within these two dimensions. Groups that are average on A and B are perceived as more communal (the 40 most communal social groups are bold), whereas groups that are extreme on A and B are perceived as less communal (the 40 least communal social groups are not bold).

Table 4 also shows the correlations *r(1D pole)*, *r(2D pole)* and *r(3D pole)*, the extent to which the 80 social groups' proximity to the origin of their 1D, 2D and 3D space related to their score on A, B, and C, respectively. The closer to the origin a group is positioned, the higher that group scores on the respective dimension, resulting in a positive correlation. In the 2D space, to some extent this was the case for A and C, but not for B. Especially C was interesting, because it was not a maximally suitable axial interpretation of the 2D space. So, to some extent C was suitable as a polar interpretation of the 2D space, $r(2D\ pole) = .41$, $p < .001$, emerging from the two axes that represent A and B. That is, the more average a group on A and B, the more communal it was stereotyped to be; the more extreme a group on A and B, the less communal it was stereotyped to be.

Discussion

We refrained from pre-selecting candidate stereotype content dimensions as well as social groups because we wanted to identify the dimensions that people spontaneously use to distinguish between groups sampled without bias. The results showed that the first two of these spontaneously used stereotype content dimensions can be interpreted as *agency / socio-economic success* (A) and *conservative-progressive beliefs* (B), because the statistical fit of A and B modelled as axes of the 2D group space was almost maximal and far better than the statistical fit of C as an axis, and A, B and C as poles at the origin of the 2D space. Unexpectedly, this data-driven A and B space (see Figure 1) was different from the warmth and competence space (Fiske et al., 2002).

Although one of the identified principal components, which we labeled agency / socio-economic success, seemed to align somewhat with the competence dimension in the SCM (Fiske, Cuddy, & Glick, 2007), we decided against labelling it that way. Recent research suggested that stereotypic competence and stereotypic agency are distinct, and that agency is more related to socio-economic success than to competence in the sense of ability (Carrier, Louvet, Chauvin, & Rohmer, 2013). Indeed, the items that loaded on the principal component in question (i.e., power, dominance, status, wealth, confidence, and competiveness) seemed to reflect agency better than competence. The items that reflected competence in the sense of ability (i.e., smartness, skill, and competence) either did not load strongly on this component or showed substantial cross-loadings on other components and were thus excluded. In other words, a janitor might be very smart and highly skilled, but would lack status and wealth. Conversely, a manager has high status and wealth, but might not be smart and skilled.

This alone, of course, can be an artefact of the item list we started from. However, a property fitting analysis with the 24 single items also suggested that wealth, power, dominance, and status are statistically better fitting single candidate stereotype content dimensions than smartness, skill, and competence, and that confidence and competitiveness are statistically better fitting candidate stereotype content dimensions than skill and competence (see Appendix A). Based on these results we propose agency / socio-economic success and not competence as one of the fundamental stereotype content dimensions on which people distinguish social groups.

We labeled the second dimension conservative-progressive beliefs. Judgments of how traditional versus modern, how conventional versus alternative, how conservative versus liberal, and how religious versus science-oriented the groups are loaded high on this dimension; we thus concluded that it captures socially shared convictions about groups' conservative-progressive beliefs. The discovery of this dimension underlines the usefulness and necessity of data-driven approaches because few theories have previously addressed what a group believes in as relevant for stereotyping. Participants seem to systematically differentiate groups on the basis of them either striving to keep up traditions / preserving the status quo (e.g., conservatives, religious, Republicans) or striving to overcome traditions / altering the status quo (e.g., gays, atheists, liberals). In a sense, much like warmth in the stereotype content model is conceptualized as informative of mainstream society's views about a group's intention to help / care versus harm / neglect, conservative-progressive beliefs are informative of mainstream society's views about a group's intention to preserve versus change the status quo.

This finding is in line with Jones and Ashmore (1973) and Pattyn and colleagues (2013), who found a similar dimensions (modern – backward and alternative – conventional, respectively) using a theory-driven selection of stimuli (e.g., an image of a punk and an elderly person). Our findings are thus the first to establish the centrality of this dimension for distinguishing between representatively sampled social groups.

The dimension of conservative-progressive beliefs is also compatible with fundamental dimensions from other areas of psychology. On the level of personality traits, the Big 5 factor openness to experience (McCrae & Costa, 1987; McCrae & John, 1992) taps into a similar construct, and this personality trait has been identified as one of the central predictors of political conservatism (Jost, Glaser, Kruglanski, & Sulloway, 2003). Moreover, one of the two central dimensions on which human values can be positioned is openness to change (self-direction, stimulation) vs. conservation (security, conformity, tradition; Schwartz, 1994; Schwartz & Bilsky, 1987; 1990). In U.S. society this dimension has received increasing attention over the recent years, leading some scholars to speak of a divide or even polarization between liberal and conservative camps (Brewer, 2005; Haidt, 2012).

Of the three combined candidate stereotype content dimensions that we composed, communion (Abele & Wojciszke, 2014; highly akin to the dimension warmth in the SCM; Fiske et al., 2007) did not appear to be one of the two stereotype content dimensions that participants most often used to judge the dissimilarities between the 80 social groups. Importantly, this was not due to the fact that the groups' C (trustworthiness, sincerity, warmth, benevolence, likeability, and altruism) ratings were unreliable. In fact, the reliability was very high for all sub-dimensions of stereotypic C, *ICC(2,k)* >= .84 (McGraw & Wong, 1996; Shrout & Fleiss, 1979). This contrasts with the pivotal role of communion/warmth in existing theories of stereotype content (Fiske et al., 2002, Cuddy et al., 2007), and also in theories of social perception in general (Abele & Wojciszke, 2014).

Despite this lack of support for C as one of the first two spontaneously employed stereotype content dimensions, we found support for C as emerging from the first two spontaneously employed stereotype content dimensions. Specifically, groups positioned closer to the origin of the 2D A and B space were judged as relatively more communal. Therefore, highly communal groups were seen as neither too rich, nor too poor, as well as neither too conservative, nor too progressive. Less communal groups were peripheral (see groups marked in blue in Figure 1), whereas more communal groups were central (see groups marked in red in Figure 1). This finding reconciles our 2D solution with existing models that consider communal attributes to be fundamental to stereotype content: Study 1 suggests that communion is encoded by the two stereotype content dimensions that we refer to as A and B in a non-linear way.

In sum, Study 1 suggests that fundamental stereotype content about social groups can be described by a two-dimensional space spanned by A and B from which communion emerges as a function of centrality within that space (see Figure 1). At this point, we cannot be certain that this 2D ABC (agency / socio-economic success, conservative-progressive beliefs, and communion) stereotype content model provides a full description of the dimensions that people spontaneously used for judging the dissimilarities between the representatively sampled groups. Most problematically, the goodness of scaling fit of the groups' 3D coordinates was more than slightly better than the scaling fit of the groups' 2D coordinates, suggesting that the 2D ABC model misses a third spontaneously used stereotype content dimension. Based on that C was a moderately suitable axial interpretation of the social groups' 3D space, this third dimension could be C (note that C can at the same time be a third independent dimension and emerge from centrality on the first two). Further, according to Kruskal and Wish (1978), a scaling fit of *S* <= .20, <= .15, <= .10, <= .05 and <= .025 is poor, sufficient, satisfactory, good and excellent, respectively. Using these criteria we have to concede that neither the 2D nor the 3D solution showed a sufficient fit. Although the 4D space met this standard (*S* = .15), the improvement was only marginal compared to the 3D solution. These findings suggest two aspects. First, people based their group dissimilarity judgments primarily but not solely on A and B. Second, the dimensions they employed additionally are not consensually shared to the degree that they form more than one orthogonal dimension that explains a noteworthy increase in explained dissimilarity variance.

To a certain degree, this was a consequence – and an advantage – of our design. Each participant judged the dissimilarity of only ~ 2.5% of all unique pairs of social groups, and thus each participant judged dissimilarity in a highly different context. This might have added additional noise to the data because dissimilarity (i.e., the way it is construed) varies as a function of context of judgment (Goldstone, Medin, & Halberstadt, 1997; Krumhansl, 1978; Tversky, 1977). The sequential, pairwise mode of dissimilarity judgment also could have encouraged people to switch between many circumstantial stereotype content dimensions rather than to stick with the essential ones. The advantage lies in that the context of judgment is so variable across participants that it could not have constrained the outcome of judgment to any dimension. The fact that we nevertheless obtained at least two meaningful dimensions speaks to the centrality of these.

Another factor that might have contributed to the non-optimal scaling fit might lie in the repetitive nature of making 80 pairwise dissimilarity judgments sequentially. It is conceivable that the repetitive nature of the task tempted our online participants to pay increasingly less attention and thus added noise to the data. In our next study, we aimed to ameliorate these problems by employing a more stimulating research design in which participants judged the dissimilarities between large arrays of social groups simultaneously. Such an alternative to the classic pairwise method has recently been proposed as the *spatial arrangement method* (SpAM; Hout et al., 2013).

Study 2

Consistent with the geometric model of similarity (Carroll & Wish, 1974; Nosofsky, 1992; Torgerson, 1965), the spatial arrangement method (SpAM; Hout et al., 2013; see also

Goldstone, 1994; Koch et al., 2016a; Kriegeskorte & Mur, 2012) rests on the assumption that people can reliably and validly sort attitude objects in a way that more dissimilar attitude objects are located further apart. To illustrate, Goldstone (1994) presented participants with multiple variants of the letter A (in different font styles) all at once and in random locations on the computer screen. Their task was to use the computer mouse to “move the letters around so that letters that are similar to each other are close. The more similar two letters are, the closer they should be” (Goldstone, 1994, p.382). The distances between the spatially arranged letters correlated strongly with sequential, pairwise dissimilarity judgments collected from a different sample of people (see also Hout et al., 2013; Koch et al., 2016a). Thus, sequential, pairwise judgment and SpAM seem to be equally effective ways to measure inter-stimulus dissimilarity. The advantage of SpAM is that it is a lot more efficient, because the dragging-and-dropping of a single attitude object simultaneously adjusts the distances between that attitude object and all other attitude objects on the dissimilarity map. In fact, with the help of SpAM, people are able to assess the entire pattern of dissimilarities between dozens of attitude objects in a quick, easy, and re-adjustable way, because all attitude objects can be moved to a different location on the dissimilarity map at all times during the task. Thus, SpAM is ideal to improve on the design of Study 1. Based on the results of Study 1, we hypothesized that people would spontaneously use the stereotype content dimensions A and B to spatially arrange the dissimilarities between the 80 social groups. We expected C to – again – emerge as a function of centrality in the two-dimensional space spanned by the other two dimensions. Together, these findings would further support our 2D ABC model of stereotype content.

Methods

Participants and stimuli. We paid 131 MTurkers (67 women, 64 men; $M = 34.74$ years, $SD = 11.84$) $1 to “sort 40 social groups on the computer screen”. They received a random sample of 40 out of the 80 groups chosen as representative of U.S. society in Study 1.

Procedure. On the first screen slide, participants read "Dear participant, your task is to sort 40 social groups based on how similar / dissimilar they are. The social groups will appear in the middle of the screen one at a time, and you can drag-and-drop them at any time to change their location on the screen. Please sort the social groups in such a way that more similar social groups are more close to each other, while more dissimilar social groups are more distant to each other. That is, please use the social groups to draw a map in which greater proximity means greater similarity, and in which greater distance means greater dissimilarity”. After clicking on an “I understand” button, the button disappeared, and a randomly selected group appeared in the middle of the screen. Once that group was dragged to another location on the screen, the button reappeared as a "Next social group" button in the center of the screen bottom, and with a click on the button the next randomly selected group appeared in the middle of the screen, and the button disappeared again. This procedure was repeated until all 40 groups were positioned on the screen. After the 40th group was arranged on the dissimilarity map, the button changed to "I finished" (see Appendix K). Upon clicking this button, the dissimilarity distances between the groups were recorded as proportions of the greatest possible distance – the screen diagonal.

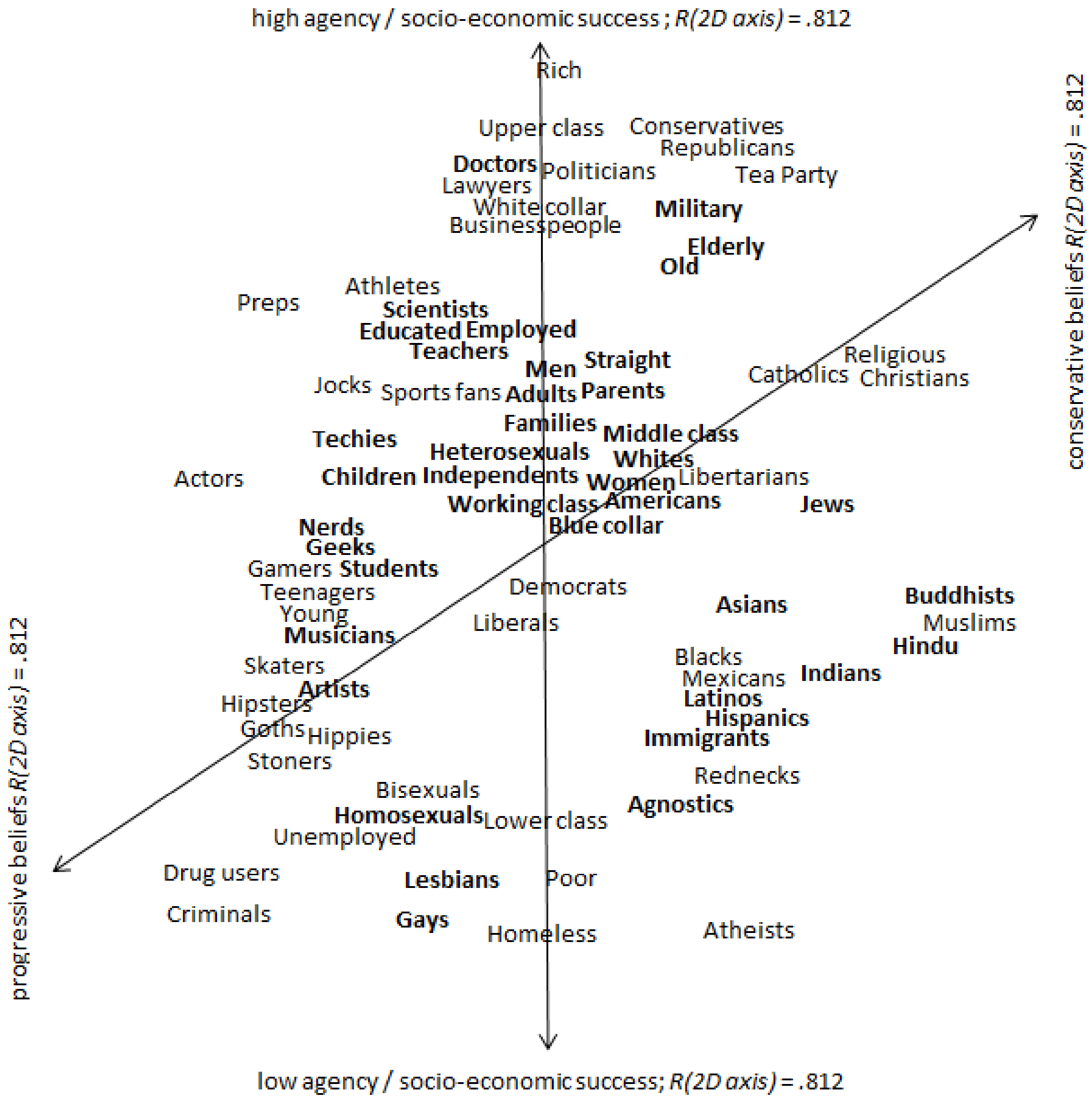
Results

First, we computed the mean distance – that is, the mean dissimilarity – for each of the 3160 unique pairs of social groups across all people who had dragged-and-dropped that pair (*M* = 32.19, *SD* = 4.94). Next, we subjected the mean dissimilarities to MDS (with the same parameter values as in Study 1. The goodness of fit of the 1D, 2D, 3D, 4D, 5D, and 6D scaling solutions are shown in Table 2. Balancing goodness of scaling and ease of interpretation, as in Study 1, we proceeded with the 1D, 2D, and 3D group spaces.

To compare the suitability of the social groups' stereotypic A, B, and C for interpreting the groups' 1D, 2D, and 3D space, as in Study 1, we again carried out nine multiple linear regressions with the groups' means on A, B, and C from Study 1 as the criterion and the groups' coordinates in the 1D, 2D, and 3D space as predictors, respectively. The results are shown in Table 4[3]. As in Study 1, A, *R(2D axis)* = .81, $p < .001$, and B, *R(2D axis)* = .81, $p < .001$, were the most and almost maximally suitable axial interpretations of the 2D space (see Figure 2), whereas C was not a suitable axial interpretation of the 2D space, *R(2D axis)* = .18, $p = .30$. Table 4 also shows the linear relations between the groups' A, B, and C and the groups' proximity to the origin of their 1D, 2D, and 3D space. As in Study 1, C was a suitable polar interpretation of the 2D space, *r(2D pole)* = .58, $p < .001$, whereas A and B were not. The same pattern of results was found for the groups' 3D space.

Figure 2

[3] The results of separate analyses for the 24 candidate stereotype dimensions are shown in Appendix B, and are consistent with Table 4.



Note. Study 2 (U.S. participants and target social groups): This 2D space of 80 representatively sampled social groups was computed based on spatially arranged dissimilarity distances, and can be interpreted by agency / socio-economic success and conservative-progressive beliefs. Communion emerges within these two dimensions. Groups that are average on A and B are perceived as more communal (the 40 most communal social groups are bold), whereas groups that are extreme on A and B are perceived as less communal (the 40 least communal social groups are not bold).

Discussion

In Study 2, participants spatially arranged sequentially appearing social groups on the two-dimensional screen (more dissimilar groups had to be positioned further apart). Our results show that given two dimensions to distinguish between the groups, people used A and B, but not C; as in Study 1, C again emerged as a function of centrality within the stereotype content space spanned by A and B (see Figure 2). Thus, our results provided further support for the 2D ABC model identified in Study 1. First, the suitability of A and B as axial interpretations of the 2D group space was almost maximal, while C was not a suitable axial interpretation of the 2D space. Second, the suitability of C as a polar interpretation of the 2D space was substantial and higher than in Study 1, while A and B were not suitable as polar interpretations of the 2D space. Third, the suitability of A and B as axial interpretations of the 2D space was higher than the suitability of C as a polar interpretation of the 2D space. And fourth, the scaling fit of the groups' 2D coordinates was sufficient (S <= .15; see Table 2).

As in Study 1, the scaling fit of the social groups' 3D coordinates was better than the scaling fit of the social groups' 2D coordinates, suggesting that the 2D ABC model missed a third independent stereotype content dimension (in the 2D ABC model, C is not an independent dimension, because it emerges as a function of centrality within the stereotype content space spanned by A and B). However, in contrast to Study 1, Study 2 showed no evidence that this third independent dimension might be described as C. Based on our data, we could not adequately interpret the third independent dimension (if there is any). Therefore, the more parsimonious 2D ABC model was the best available interpretation of the stereotype content dimensions that people spontaneously used to spatially arrange the dissimilarities between the 80 representative U.S. groups. In other words, Study 2 confirmed that the two most fundamental dimensions of stereotype content can be described as A and B, and that C can be described to emerge from A and B, and not vice versa. Specifically, the more average a social group on A and B, the more communal it was stereotyped to be; in contrast, the more extreme a social group on A and B, the less communal it was stereotyped to be.

Both Study 1 and 2 recruited Amazon.com Mechanical Turk workers as participants, because their demographics have repeatedly been shown to be relatively more population-representative than the demographics of other convenience samples such as university students (Berinsky, Huber, & Lenz, 2012; Buhrmester, Kwang, & Gosling, 2011; Casler, Bickel, & Hackett, 2013; Mason & Suri, 2011; Paolacci, Chandler, & Ipeirotis, 2010). Nevertheless, one could argue that the results might be specific to our participant sample. Despite their greater representativeness in terms of age, education, and income, MTurkers might constitute a biased sample in terms of other variables like affinity with computers. Even more relevant, the population of the U.S. is not representative of other nations. The strong topicality of the divide between two political camps in the U.S. (Brewer, 2005; Layman & Carsey, 2002; McCarty, Poole, & Rosenthal, 2006) might have increased the salience and accessibility of the conservative-progressive beliefs dimension. We thus sought to bolster the generalizability of our 2D ABC model by replicating it in a different culture.

Study 3

Study 3 replicated Study 2 with German rather than U.S. American participants.

Methods and results

Study 3a: Naming social groups in Germany. We collected data from 178 online participants contacted through an e-mail list of individuals interested in participating in studies at the University of Cologne (119 women, 53 men; *M* = 26.35 years, *SD* = 6.11). They were offered a chance to win one of five vouchers (€20) for a large online retailer. In the top half of the first screen slide, they read the same instructions (in German) as the people who named social groups in Study 1. In the bottom half, they entered 40 social groups into 40 text boxes. Table 5 shows all 76 social groups named by more than 10% of all people.

Table 5

Most Frequently Named Social Groups (Consensus ≥ 10%) in Germany in Study 3a

1st - 20th most frequent	21st - 40th most frequent	41st - 60th most frequent	61st - 76th most frequent
Students (70%)	Christians (27%)	Adults (20%)	Germans (14%)
Children (58%)	Foreigners (27%)	Drug addicts (20%)	Goths (13%)
Employed (56%)	Religious (27%)	Catholics (19%)	Alcoholics (13%)
Unemployed (56%)	Academics (26%)	Conservatives (19%)	Single parents (13%)
Young (47%)	Homosexuals (26%)	Self-employed (18%)	Rightists (12%)
Pupils (46%)	Musicians (26%)	Welfare recipients	Sick (12%)
Pensioners (44%)	Jews (24%)	Criminals (18%)	Nazis (12%)
Muslims (38%)	Trainees (24%)	Lower class (16%)	Blue collar (11%)
Officials (37%)	Parents (23%)	Upper class (16%)	Hip-Hopper (11%)
Workers (36%)	Vegans (22%)	Leftists (16%)	Emos (11%)
Athletes (34%)	Hipsters (22%)	Rural (16%)	Scientists (11%)
Politicians (33%)	Singles (22%)	Economic-liberals	Right-wing extremists
Migrants (33%)	Teachers (21%)	Employers (16%)	Rockers (11%)
Artists (31%)	Atheists (21%)	Car drivers (15%)	Managers (11%)
Middle class (31%)	Vegetarians (20%)	Nerds (15%)	Bicycle drivers (10%)
Punks (30%)	Poor (20%)	Educated (15%)	Soccer players (10%)
Elderly (30%)	Urban (20%)	Buddhists (15%)	
Disabled (29%)	Doctors (20%)	Hippies(15%)	
Rich (29%)	Heterosexuals (20%)	Environmentalists	
Homeless (28%)	Families (20%)	Celebrities (14%)	

Note. Percentage in parentheses is proportion of participants who spontaneously named this group as a social group that is representative of the structure of German society.

Study 3b: Dissimilarity arrangement of 76 groups. Another 69 students were recruited on the campus of the University of Cologne (47 women, 22 men; M = 23.37 years, SD = 4.53) to participate in a lab study for a small monetary compensation (€2). Their instructions were the same as in Study 2 (in German), namely to spatially arrange a random sample of 50 of the 76 social groups. More similar groups had to be placed more close to one another, and more dissimilar social groups had to be placed further apart from one another.

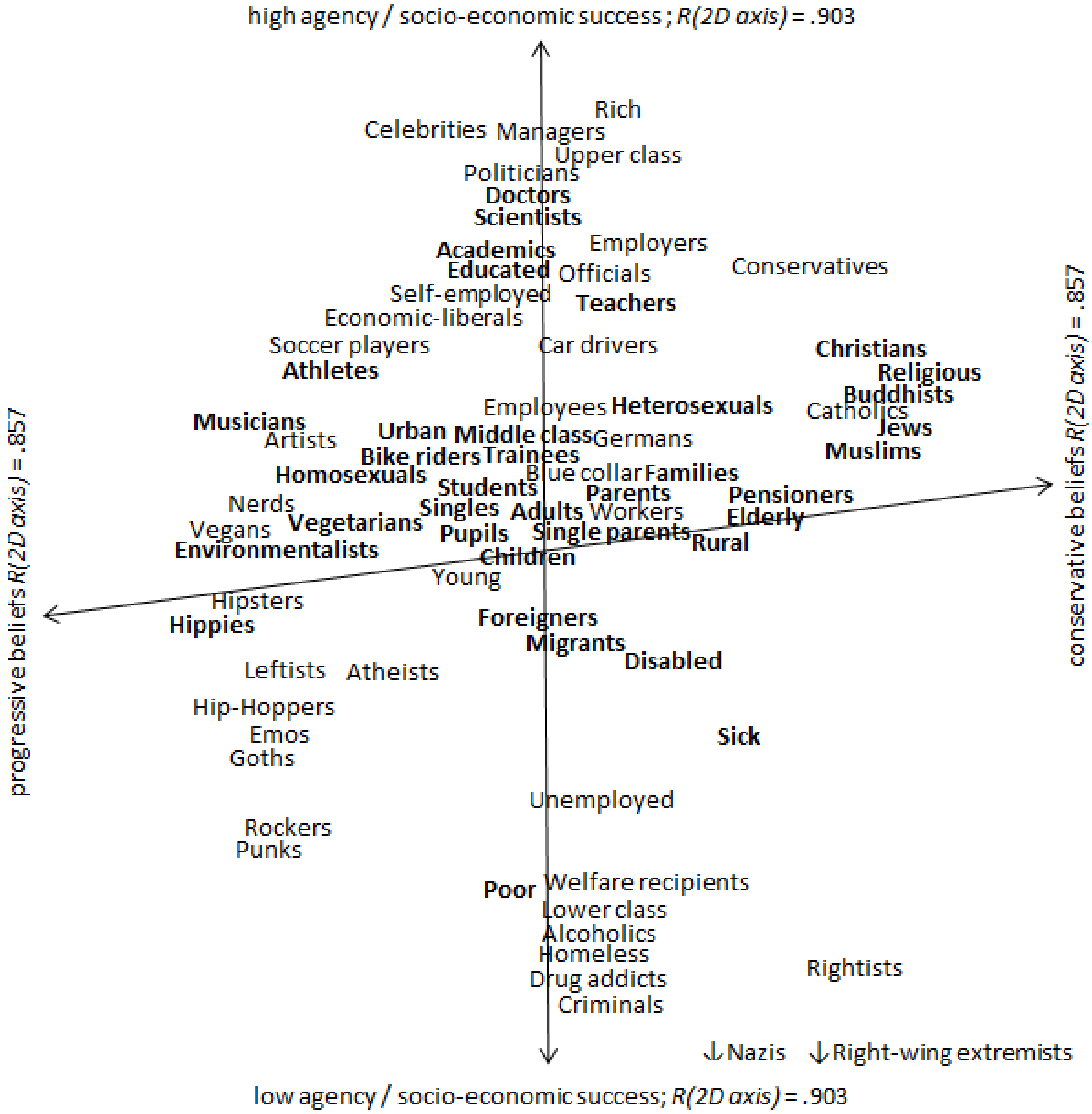
As in the previous spatial arrangement study we subjected the mean distances for each of the 2850 unique pairs of social groups (average number of raters per pair M = 29.23, SD = 5.15) to MDS (same settings as in Studies 1 and 2). The goodness of fit of the 1D, 2D, 3D, 4D, 5D and 6D scaling solutions are shown in Table 2. Balancing goodness of scaling fit and ease of interpretation, we extracted, analyzed, and interpreted only the 1D and 2D social group spaces, because the scree plots of S and $1\text{-}R^2$ showed that extracting a third and higher dimensions did only to a slight degree improve S and R^2. Nevertheless, to better compare

Study 3 with Studies 1 and 2, we extracted and analyzed a third dimension, but refrained from interpreting the respective 3D group space.

Study 3c: Disambiguating the dissimilarity arrangement. Finally, 60 other participants recruited on the campus of the University of Cologne (41 women, 19 men; $M = 22.55$ years, $SD = 4.55$) received a piece of candy to participate in a lab-study to rate the 76 social groups on compound A, B, or C. Twenty participants rated all groups on a slider scale ranging from 1 ('low power / low status / low dominance / low confidence') to 100 ('high power / high status / high dominance / high confidence'), measuring stereotypic A, $ICC(2, k) = .965$, twenty different participants rated all social groups' B on an identical scale with the anchors 'traditional / religious / conservative / conventional – modern / faithless / liberal / alternative'; $ICC(2, k) = .912$, and a third group rated all social groups' C ('low trustworthiness / low sincerity / low benevolence / low likability – high trustworthiness / high sincerity / high benevolence / high likability'), $ICC(2, k) = .952$. A and C were correlated, $r = .32$, $p < .01$; A and B were orthogonal, $r = -.10$, $p = .41$; and B and C were orthogonal, $r = .08$, $p = .49$.

Employing the same property fitting strategy as in the previous studies suggested that A, $R(2D\ axis) = .90$, $p < .001$, and B, $R(2D\ axis) = .86$, $p < .001$, were far better axial interpretations of the 2D social group space than C, $R(2D\ axis) = .38$, $p < .001$ (see Table 4 and Figure 3). As in Study 2, C was a suitable polar interpretation of the 2D space, $r(2D\ pole) = .75$, $p < .001$, whereas A and B were not. The same pattern of results was found for the groups' 3D space.

Figure 3



Note. Study 3 (German participants and target social groups): This 2D space of 76 representatively sampled social groups was computed based on spatially arranged dissimilarity distances, and can be interpreted by agency / socio-economic success and conservative-progressive beliefs. Communion emerges within these two dimensions. Groups that are average on A and B are perceived as more communal (the 40 most communal social groups are bold), whereas social groups that are extreme on A and B are perceived as less communal (the 40 least communal social groups are not bold).

Discussion

Study 3 was set in another national context (Germany), and we used another type of sample (mostly university students) and another research setting (Studies 3b and 3c were conducted in the lab rather than online). The scaling fit of the social groups' 2D coordinates was again sufficient ($S <= .15$) and did not markedly improve with the addition of another dissimilarity dimension. Thus, modelling a third independent stereotype content dimension was nonessential. Overall, Study 3 supported the 2D ABC stereotype content model identified in Studies 1 and 2 (see Figure 3), and also refined it, potentially due to reduced noise in the data. Specifically, A was found to be more fundamental than B, as in the groups' 1D space (already accounting for 75% of the original spatially arranged dissimilarity variance) A was a suitable axial interpretation, whereas B was not. B only accompanied A as a suitable axial interpretation in the 2D space. Because A and B were more suitable axial interpretations of the 2D space compared to the suitability of C as a polar interpretation of the 2D space, the primary and secondary fundamental stereotype content dimensions should be interpreted as A and B, respectively. Further, the suitability of C as a polar interpretation was already maximal in the 1D space, which, in contrast to Studies 1 and 2, suggests that C is primarily a function of not being too high or too low on A (i.e., C can be inferred from A, but not from B; see blue and red social groups in Figure 3).

Study 3 thus refined the conclusions drawn from Studies 1 and 2 and at the same time strengthened the empirical base of the proposed 2D ABC model of stereotype content. Our results became increasingly clear from Study 1 to Studies 2 and 3, speaking to the usefulness of employing the spatial arrangement method. In particular, if participants are confined to a two-dimensional space, they use A and B to organize the social groups, while C emerges as a non-linear function of A but not B.

Verticality, horizontality, and centrality metaphors as possible alternative explanations

Despite this, there may also be pitfalls in using this method. It is conceivable that semantic concepts are intrinsically associated with spatial locations. One example is that the concept of power is commonly found to be intuitively represented vertically with high power on the top and low power on the bottom (Schubert, 2005; see also Meier & Robinson, 2004; Slepian, Masicampo, & Ambady, 2015). Specifically, words representing high vs. low power (e.g., employer vs. employee) can be more rapidly identified as connoting high vs. low power when they are presented in the metaphorically corresponding area of a screen (top for high power, bottom for low power; Schubert, 2005). Likewise, people presented at the top of the screen are perceived to be more powerful (Meier, Hauser, Robinson, Friesen, & Schjeldahl, 2007; Giessner & Schubert, 2007). Thus, the vertical nature of the spatial arrangement board could have primed people to construe the dissimilarities between the social groups with respect to their stereotypic agency / socio-economic success.

A similar argument could be made for the belief dimension. Going back to the seating arrangement in legislative bodies during the French revolution era, progressives are often referred to as left-wing, whereas conservative beliefs are referred to as right-wing. Importantly, this is not only an abstract reference but horizontal positions on the left or the right are intuitively connected with the corresponding political attitudes (Farias, Garrido, & Semin, 2013; Oppenheimer & Trail, 2010; van Elk, van Schie, & Bekkering, 2010). The horizontal nature of the spatial arrangement board could thus have primed people to construe the dissimilarities between the social groups with respect to their conservative-progressive beliefs and sort the groups accordingly.

Finally, even the position of high communal groups in the center and low communal groups near the margin of the screen could be construed as merely reflecting an instance of embodied semantics or metaphors. The center often stands for relevant ingroups, whereas outgroups are labelled as (in this case literally) peripheral. Moreover, motivated explanations are conceivable. Pushing unpleasant (i.e., unfriendly and cold) groups to the margin of the screen and keeping the friendly and warm groups in the center of frequent attention could be experienced as more pleasant than the opposite. Alternatively, being motivated to keep the most frequent, typical, and familiar groups in the center of frequent attention could explain why high and low communal groups were positioned in the center and at the margins, respectively, because "what is typical is good" (Sofer, Dotsch, Wigboldus, & Todorov, 2015).

To test whether these confounds are, at least in principal, a problem, we correlated the social groups' spatially arranged average verticality (higher values: more upwards), horizontality (higher values: more rightwards), and centrality (higher values: more inwards) with the social groups' A, B, and C ratings, respectively. In line with spatial metaphors, more agentic social groups were positioned further upwards, $r_{Study\ 2} = .74$, $p < .001$, $r_{Study\ 3} = .60$, $p < .001$; and, more communal social groups were positioned further inwards, $r_{Study\ 2} = .52$, $p < .001$, $r_{Study\ 3} = .69$, $p < .001$. Contradicting a spatial metaphor, more liberal (conservative) social groups were not positioned further leftwards (rightwards), $r_{Study\ 2} = .03$, $p = .78$, $r_{Study\ 3} = -.23$, $p = .04$.

Although the social groups' spatially arranged coordinates cannot explain our empirical support for the B dimension, and although the spatial A, B, and C metaphors account cannot explain the highly convergent findings of Study 1 that did not use a spatial arrangement task, in Study 4 we sought to rule out spatial metaphors of A, B, and C and motivational forces as alternative explanations for the results obtained in Studies 2 and 3.

Another caveat that we sought to address in Study 4 refers to how participants decided on the criterion to estimate the dissimilarities between the social groups. Participants relied on group differences that map well on the groups' A and B. This can be interpreted as evidence that A and B are the two fundamental and thus most important stereotype content dimensions. It is also possible that the employed dimensions are not the most important but merely the most convenient, either because they are most accessible (which could be seen as an indirect indicator of importance), or because their metaphorical spatial representation conveniently maps on the spatial nature of the arrangement task. We sought to address this caveat by introducing a more explicit judgment of dimension importance. Specifically, in Study 4, we explicitly asked participants to label the first and second most fundamental stereotype content dimensions prior to the spatial arrangement task. This also helped addressing the issue of order of importance.

Study 4

Study 4 presented people with a sample of 40 groups, and asked them to specify the two person characteristics (i.e., stereotype content dimensions) that they thought best capture the dissimilarities between the social groups. Asking them to prioritize between the two allowed us to get an empirical hold on which dimension is seen as more primary than the other. Then, people judged the social groups precisely on the two dimensions that they had just selected as the most important ways to stereotypically compare the groups by spatially arranging them. Importantly, the first named dimension always had to be arranged on the horizontal dimension of the screen, thereby undermining spontaneous mapping on metaphorically corresponding dimensions (assuming that A would be most primary and thus mapped horizontally instead of vertically as in the previous studies). Based on the 2D ABC model, we expected people to specify stereotype content dimensions with a high relation to both A and B, and with a low relation to C.

Method

Participants and stimuli. We paid 66 MTurkers (31 women, 35 men; *M* = 32.86 years, *SD* = 10.81) $1 to "rate 40 social groups on 2 dimensions of your choice". People spatially arranged a random sample of 40 of the 80 social groups that are representative of the structure of the U.S. society (see Study 1).

Procedure. In the middle of the first screen slide, people were presented with a table that showed 40 social groups in 10 rows and 4 columns. At the top, they read: "Dear participant, please name the person characteristic that best describes the differences and similarities between these 40 social groups. Ideally, you should be able to divide the 40 social groups into 10+ low scorers, 10+ average scorers and 10+ high scorers on this characteristic. Please enter this characteristic in the text box below [this instruction]". At the bottom, they read: "Now, please name another person characteristic that well describes the differences and similarities between these 40 social groups. Again, you should be able to divide the 40 social groups into 10+ low scorers, 10+ average scorers and 10+ high scorers on this other characteristic. Please enter this other (= not the same as above!) characteristic in the text box below [this instruction]". On the second screen slide, participants read: "Dear participant, your next task is to position the 40 social groups on the computer screen. More specifically, on the next slide the 40 social groups will appear in the middle of the screen, and your task is to drag-and-drop each social group to a different position on the screen. Please make use of the entire screen and position the social groups as follows. For the first person characteristic that you specified (→ [whatever they had typed in first]): position the low scorers on the left of the screen, position the average scorers in between the left and the right, and position the high scorers on the right of the screen. For the second person characteristic that you specified (→ [whatever they had typed in second]): position the low scorers at the bottom of the screen, position the average scorers in between the bottom and the top, and position the high scorers at the top of the screen. In sum, your task is to plot the 40 groups according to how they differ on [whatever they had typed in first] and [whatever they had typed in second]. If you want to exchange these person characteristics, you may do so at any time during the positioning task by rephrasing the text boxes at the screen edges". On the third slide the 40 groups appeared in 40 adjacent labels in the screen middle. The appearance of the labels and the full screen background was the same as in Studies 2 and 3, except for a horizontal axis labeled in accordance with the characteristic that the participant had specified first, and a vertical axis labeled in accordance with the characteristic that the participant had specified second (see Appendix L). The labeling of the axes ("low scorers on …" and "high scorers on …") could be changed at any time during the rating phase. Once all groups were dragged-and-dropped to a different position on the screen, an "I finished" button appeared. Upon clicking this button, the computer program rescaled the groups' positions to a quadratic 2D space, and then, as in Studies 2 and 3, the distances between the groups were recorded as a proportion of the greatest possible distance – the diagonal of this 2D space.

Results

We first computed the mean rating distance – that is, the mean dissimilarity – for each of the 3610 unique pairs of social groups across all people who had rated that pair (M = 16.02, SD = 3.83). These mean dissimilarities were subjected to MDS with the same settings as in Studies 1-3. The goodness of fit of the 1D, 2D, 3D, 4D, 5D, and 6D scaling solutions are shown in Table 2. As in Study 3, to ease comparing Study 4 to the other studies, we extracted and analyzed three dimensions, but interpreted only the 1D and 2D group spaces, as the scree plots of S and 1-R^2 showed that extracting a third or higher dimensions did only slightly improve S and R^2 (i.e., modelling a third independent stereotype content dimension was nonessential.). Extracting a second dimension did not substantially improve S, but substantially improved R^2, and thus we interpreted the 1D and 2D social group space.

Identical property fitting analyses as in the previous studies were carried out and replicated the findings of Study 3 that A, *R(2D axis)* = .89, p < .001, and B, *R(2D axis)* = .85, p < .001 (see Table 4[4]), were far better axial interpretations of the 2D social group space compared to C, *R(2D axis)* = .38, p = .003. Replicating Studies 1-3, C was a substantially suitable polar interpretation of this 2D space, *r(2D pole)* = .51, p < .001 – A and B were not.

More importantly, we sought to address whether the spatial nature of our task might have prompted participants to employ metaphorically corresponding stereotype content dimensions (A for the vertical dimension, B for the horizontal dimension, C for centrality). As participants had to specify two stereotype content dimensions before learning about the spatial nature of the rating task, the task could only have influenced participants' choice of dimensions if participants changed the two dimensions once realizing that they had to spatially arrange social groups. Tracking such changes in the self-selected dimension labels revealed that this happened for only very few cases (around 7%).

To categorize the self-selected labels according to unambiguous fit with our combined candidate stereotype content dimensions, 40 additional MTurkers (18 women, 22 men; M = 33.10 years, SD = 8.45) were paid $0.5 to "assign 66 person characteristics" one after the other, and they read: "please select the category to which this person characteristic [e.g., wealthy] fits best. If this person characteristic does not fit well to any of the categories, check

[4] Separate property fitting analyses for the original 24 candidate stereotype dimensions show results that are highly consistent with Table 4, and that can be seen in Appendix C.

'no match'". The seven categories available for selection were 'no match' plus the polar opposites on A, B and C (A+: "powerful / dominating / high status / wealthy / confident / competitive", A-: "powerless / dominated / low status / poor / unconfident / unassertive", B+: "modern / science-oriented / alternative / liberal", B-: "traditional / religious / conventional / conservative", C+: "trustworthy / sincere / likable / benevolent / warm / altruistic", and C-: "untrustworthy / dishonest / repellent / threatening / cold / egoistic").

Participants assigned either the 66 labels chosen by the original participants as the "best" description of the differences and similarities between the groups (i.e., the first dimension chosen, N = 22), or the 66 labels chosen by the original participants as their second dimension (N = 18). For each of the 132 labels, we averaged percentage of assignment to categories A±, B±, C±, and 'no match', a measure of the labels' relatedness to A, B, C, and something else, respectively. Consistent with the property fitting analyses reported in Studies 1-4, the labels related to A (M = 28.35%, SD = 31.21) and B (M = 31.15%, SD = 30.83) to an equal extent, $F(1, 131) = 0.37$, $p = .54$, $\eta_p^2 = .00$, 90% CI [.00, .04], related at least by trend more to A than to C (M = 21.03%, SD = 22.65), $F(1, 131) = 3.69$, $p = .057$, $\eta_p^2 = .03$, 90% CI [.00, .09], and related more to B than to C, $F(1, 131) = 7.43$, $p < .01$, $\eta_p^2 = .05$, 90% CI [.01, .13].

Discussion

In Study 4, people first named the two stereotype content dimensions that they thought best describe the similarities and differences between 40 randomly selected social groups from our full set of 80 groups from Study 1. A relatively large number of given dimensions did not fit either of our combined ABC candidate dimensions unambiguously. Yet, for those that did, agency / socio-economic success was consensually seen as the most fundamental dimension to describe similarities and differences between groups, and conservative-progressive beliefs was named as the second most important dimension. Subsequently, people spatially arranged the groups on a rating board with x- and y-axes labeled precisely according to the two stereotype content dimensions that they had just named. During this task, people were free to re-label one or both of the stereotype content dimensions that they had named *before*. The very low frequency of re-labeling showed that most participants stuck to the originally named dimensions, which thus cannot be prompted by the spatial nature of the sorting task. Nevertheless, the two most suitable axial interpretations of the 2D social group space were again A and B, suggesting that in Studies 2 and 3 spatial A, B, and C metaphors (verticality, horizontality, and centrality, respectively) do not provide a sufficient explanation for the empirical support for our 2D ABC model of stereotype content. Further, in Study 4, the

2D group space had a sufficient scaling fit (*S* <= .15), and the ProFit analyses clearly confirmed the more fine-grained results of Study 3 that A best described the primary stereotype content dimension, followed by B as the best description for the secondary dimension, and that C emerged as a non-linear function of A but not B.

Study 4 also addressed another issue. The most noteworthy difference between our data-driven stereotype content model and existing theory-driven models lies in the role of communion or, alternatively, warmth. Whereas dominant theories not only suggest that warmth is a central dimension on which social groups are compared and judged, but also postulate a primacy of this dimension (for an overview, see Abele & Wojciszke, 2014), we only found support for an emergent nature of C. More precisely, Studies 1-4 suggest people compare and judge groups on A and B, and that C is encoded by A, with average and extreme A implying high and low C, respectively. One explanation for this difference in conclusions might be that the current studies tap into spontaneously employed dimensions, whereas previous work explicitly instructed participants to judge the C or warmth of groups. Another explanation might be that in our studies participants refrained from employing C-related stereotypical information, because they saw it as relatively more socially undesirable to denigrate groups on stereotypic C (e.g., “lawyers are dishonest, homeless are repellent, military are threatening, and punks are untrustworthy”) compared to stereotypic A and B (“lawyers are overconfident, homeless are powerless, military are overly conservative, and punks are too alternative”). Such hesitancy would in all likelihood emerge most strongly if we explicitly encourage participants to label the dimensions they employ as we did in Study 4. An inspection of the results, however, suggests that the opposite was true: the suitability of C as a third axial interpretation of the 3D social group space was never high, but more pronounced in Study 4 than in any of the previous studies. This speaks against the idea that participants responded in a socially desirable way that prohibited the expression of perceived group differences in communion.

In sum, the studies so far support a two-dimensional model of the mental organization of social groups. In marked contrast to previous theorizing about the nature of these dimensions, the data suggest that the primary and secondary mapping principles are agency / socio-economic success (i.e., the groups’ perceived wealth, status, power, dominance, confidence and competitiveness) and conservative-progressive beliefs (i.e., the groups’ perceived position on a continuum ranging from liberal, alternative, science-oriented and modern to traditional, conventional, religious and conservative), respectively (see Figures 1-3). Communion (i.e., perceived trustworthiness, sincerity, warmth, benevolence, likeability and altruism) was not used as a criterion for distinguishing between the representatively sampled social groups, but emerged for those social groups who are stereotyped as average on A but not so much B[5].

Despite the consistency of our findings so far, there remain some caveats that require further attention. Specifically, despite adherence to a data-driven strategy, every study required some decisions. Our decisions of a) how to sample social groups, b) how to instruct the arrangement of these groups, and c) which candidate dimensions to use for our property fitting analyses may have biased our results, leading to the observed 2D ABC model. For example, our instruction of how to sample groups might have favored groups defined by their socio-political ideology, our arrangement instruction might have made similarity in socio-political opinion particularly salient and there might be further oblique candidate dimensions that are equally fitting candidates that we never collected ratings for. Study 5 sought to address these caveats in a comprehensive design.

Study 5

Study 5 sought to generalize the 2D ABC model of stereotype content beyond our previous approaches to sampling (see Studies 1a and 3a), comparing (see Studies 1b, 2, 3b and 4), and rating (see Studies 1c and 3c) social groups. This was done to rule out that the consistent results

---

[5] It is possible that the empirical support for the 2D ABC model obtained in Studies 1-4 hinges on our criterion for defining a social group as ‘representative of the U.S. / German society’. This criterion was that at least 10% of the participants named the group as ‘representative of the U.S. / German society’ in Study 1 / 3. The relatively large number of groups that reached this criterion (80 and 76 compared to less than 30 in other research; e.g., Fiske et al., 2002) necessarily introduces variance on several dimensions. Such variance is a prerequisite for people to place groups on simplifying stereotype content dimensions (Ford & Stangor, 1992; Nelson & Miller, 1995). This begs the question whether a more strict criterion of what constitutes a group leads to a less diverse sample, and thus to less and / or different stereotype content dimensions, if any at all. In another study, we sought empirical support for that our 2D ABC model of stereotype content is not an artefact of the relatively infrequently named groups. This additional study was identical to Study 4, except that participants spatially arranged only the 40 groups that people saw as most representative of U.S. society in Study 1 (i.e., Table 1’s left columns; in Studies 1-4 participants spatially arranged random samples of 40 out of the 80 most frequently named groups). This study provided results highly consistent with Studies 1-4: the scaling fit of the 2D group space was satisfactory ($S = .07$). A, $R(2D\ axis) = .88$, $p < .001$, and B, $R(2D\ axis) = .91$, $p < .001$ (see Appendices D and E), were again far better axial interpretations of the 2D space compared to C, $R(2D\ axis) = .53$, $p = .002$, and C was a suitable polar interpretation of the 2D space, $r(2D\ pole) = .33$, $p = .04$, whereas A and B were not. These results ruled out that the consistently found 2D ABC model of stereotype content was an artefact of overly inclusive sampling of social groups.

we obtained in the first four studies were due to unduly influences of top-down decisions we made in pursuing our bottom-up, data-driven approach. Below, we outline four potential sources of bias in stimulus sampling, dissimilarity arrangement, and property fitting and how Study 5 addressed those.

Stimulus Sampling

First, the relatively abstract instructions according to which our participants named social groups (“name … groups that structure society”) might have primed social categories related to low-high agency / socio-economic success and / or conservative-progressive beliefs rather than different levels of communion and / or other stereotype content dimensions. Particularly our definition that social groups “… are based on how people behave or see the world …” might be interpreted as referring to religious and political ideology or lifestyle (e.g., Christians, Muslims, Republicans, Tea Party, conservatives, hippies, hipsters, goths, Democrats, liberals, independents etc.). This would artificially increase the salience of conservative-progressive beliefs over other stereotype content dimensions. Further, we forced participants to name 40 groups, which might be more than what is typically sufficient / necessary to mentally represent and organize society. Possibly, it is mainly this surplus of groups that relates to different levels of agency / socio-economic success and conservative-progressive beliefs.

A truly data-driven approach might require instructions under which different types of groups, including social categories, task groups (e.g., clubs, committees) and primary groups (e.g., family, friends), are equally accessible in memory. To this end, in Study 5a we used a minimalist definition of groups and minimalist naming instructions that did not prime certain kinds of groups. Additionally, we allowed participants to name any number between 3 and 30 social groups (e.g., Fiske et al., 2002). Consistent with Studies 1a and 3a, the selection that we used included all groups named by at least 10% of all participants.

To further validate the sampled groups, we created another selection. First, we selected all groups that were named at least twice. Of those, we selected the groups that appeared most frequently in a multi-billion word text corpus that contains a vast variety of digitalized books published in recent years (e.g., Akpinar & Berger, 2015; Michel et al., 2011). Thus, the relevance of groups in our second, naturalistic selection was not determined by participants but based on frequency of appearance in cultural products like books.

Dissimilarity Arrangement

Second, the dissimilarity rating, arrangement, and labeling tasks might have been too broad and abstract. That is to say, participants may have used information that goes well beyond typical stereotypic comparisons based on character traits and personal encounters (Fiske et al., 2002; Koenig & Eagly, 2014). Specifically, it is conceivable that participants also based their dissimilarity ratings, arrangements, and labels on the degree to which members of the groups typically (dis)agree with each other in the social and political arena. In Study 5b, we tested whether the 2D ABC model of stereotype content is valid even if people are instructed to compare groups based on the characters of / personal encounters with typical group members.

Property Fitting

Third, perhaps our empirical support for the 2D ABC model of stereotype content is contingent on the instructions according to which participants rated the social groups on agency / socio-economic success, conservative-progressive beliefs, and communion. In particular, we asked participants for their personal belief about the groups' A, B, and C rather than to ask for the groups' A, B, and C "as viewed by society" (Fiske et al., 2002, p.884; Cuddy et al., 2007; Kervyn et al., 2013; Kervyn, Fiske, & Yzerbyt 2015). Society's view of groups is closer to the definition of stereotypes as socially shared views (Fiske et al., 2002). Thus – to make sure that we measure A, B, and C stereotypes – in Study 5c we asked for A, B, C ratings "as viewed by society". In addition, we employed a different measure of communion that has been reported to better capture its essence (Kervyn et al., 2015).

And fourth, despite the good statistical fit of A and B as almost entirely orthogonal axes of the 2D social group spaces extracted in Studies 1-4, it is conceivable that there are other pairs of equally orthogonal and well-fitting stereotype dimensions that we overlooked. That is, our candidates possibly did not reflect the full range of stereotype dimensions that our participants used to mentally organize the social groups, and thus we cannot be sure that A and B is the best model of the two most important stereotype dimensions that people spontaneously use to distinguish between groups. To show that this is the case, in Study 5d we asked a new sample of participants to label nine rotated, equidistant axes that run through the origins of the 2D group spaces extracted in Study 5b, and we asked another sample of participants to categorize the generated axes labels as fitting well to either A, B, or C as defined in Study 1, or as "no match" if a label "does not fit well" to any of A, B, and C. If the 2D group spaces entail stereotype dimensions that we overlooked so far, we would expect sizable amounts of "no match" responses at the corresponding rotation angles. If, however, most spontaneously generated labels for virtually all rotation angles are categorized as fitting well to A or B (i.e., as being synonyms of A and B), this provides strong support that no other, oblique two-dimensional space provides a better description of spontaneously activated stereotype content about groups.

Our hypotheses for Studies 5a-5d were that the 2D ABC model of stereotype content holds true (1) for the new minimalist sampling instruction and the new naturalistic sample of social groups, (2) for similarity-, character- and personal encounter-based comparisons of social groups, (3) for social groups rated on the relevant dimensions "as viewed by society" rather than single persons, and that (4) there will be no evidence for overlooking an alternative model.

Methods and results

Study 5a. Creating a minimalist and a naturalistic sample of social groups. We paid 100 MTurkers (39 women, 61 men; M = 32.21 years, SD = 10.89) $0.5 to "name up to 30 social groups". Participants read these minimalist instructions: "off the top of your head, what various types of people do you think today's society categorizes into groups?" These were the exact same instructions as in Fiske and colleagues (2002, p. 883; see also Kervyn et al., 2013; 2015), except that we dropped "(i.e., based on ethnicity, race, gender, occupation, ability, etc.)" to avoid priming groups defined by the ethnicity, race, gender, occupation and / or ability of their members. Participants had the possibility to list up to 30 groups, although a minimum of three was required. On average, participants named 14.61 groups (SD = 9.32).

Table 6 shows all 42 groups named by more than 10% of all participants[6]. 40 of these 42 groups had also been named by 10% of participants in Study 1a. Further, the frequency with which the 40 groups had been named in Study 1a substantially predicted the frequency with which they were named in Study5a, r = .83, p < .001. Thus, the minimalist social group sample in Study 5a was very similar to the social group sample in Study 1a. If anything, in the minimalist sample extreme scorers on A ('poor', 'middle class', 'rich') and B ('Democrats', 'Republicans', 'gays', 'Christians', 'liberals', 'conservatives') were named more, not less, frequently. Thus, in hindsight the instructions employed in Study 1a do not seem to have biased the group sample in a way that social categories defined by their religious / political ideology / lifestyle were named disproportionally often.

[6] We added the frequency of naming for the synonymic social groups Blacks / African Americans, Church / Christians, rich / wealthy, Hispanics / Latinos, elderly / old / seniors, upper class / elites, athletes / sportsmen, and atheists / non-religious.

Table 6

Most Frequently Named Social Groups in the U.S. (Consensus > 10%) in Study 5a.

1st - 21th most frequent minimalist groups	22st - 42th most frequent minimalist groups	1st - 31th most frequent naturalistic groups	32st - 61th most frequent naturalistic groups
Blacks (50%)	Athletes (15%)	Children (13.12m)	Professionals (0.94m)

Whites (41%)	Parents (15%)	Women (10.80m)	Muslim / -s (0.93m)
Poor (37%)	Nerds (14%)	Old (10.51m)	Conservative / -s
Middle class (34%)	Hippies (14%)	Family (10.43m)	Scientists (0.87m)
Rich (33%)	Immigrants (14%)	Men (9.05m)	Tall (0.86m)
Hispanics (31%)	Atheists (13%)	White / -s (7.91m)	Republican / -s (0.84m)
Asians (29%)	Blue collar (13%)	Black / -s (7.88m)	Artists (0.84m)
Democrats (29%)	Religious (13%)	Christians (7.58m)	Lesbian / -s (0.70m)
Republicans (29%)	Men (12%)	Students (7.49m)	Actors (0.69m)
Gays (27%)	Teenagers (12%)	Young (6.70m)	Immigrants (0.60m)
Christians (26%)	White collar (12%)	Short (4.68m)	Hispanic / -s (0.60m)
Liberals (26%)	Politicians (12%)	Parents (4.35m)	Farmers (0.59m)
Conservatives (26%)	Jocks (11%)	Poor (3.83m)	Teenagers (0.52m)
Working class (22%)	Hipsters (11%)	Jewish (3.47m)	Educated (0.51m)
Transgender (21%)	Celebrities (11%)	Friends (3.37m)	Elites (0.50m)
Elderly (20%)	Drug addicts (11%)	Military (3.36m)	Democrat / -s (0.41m)
Students (19%)	Homosexuals (10%)	Religious (3.08m)	Clubs (0.38m)
Lesbians (17%)	Homeless (10%)	Americans (2.43m)	Homosexual / -s
Women (16%)	Jews (10%)	Rich (2.30m)	Politicians (0.38m)
Upper class (15%)	Goths (10%)	Gay / -s (2.21m)	Musicians (0.36m)
Muslims (15%)	Lower class (10%)	Europeans (2.19m)	Activists (0.30m)
		Chinese (2.05m)	Minorities (0.28m)
		Indian / -s (1.89m)	Law enforcement
		Straight (1.83m)	Alcoholics Anonymous
		Adults (1.80m)	Entrepreneurs (0.24m)
		Athletes (1.43m)	Catholics (0.24m)
		Writers (1.26m)	Homeless (0.24m)
		Ethnic (1.21m)	Mexicans (0.20m)
		Asian / -s (1.15m)	Rebels (0.18m)
		Employed (1.04m)	Middle class (0.17m)
		Liberal / -s (0.94m)	

Note. Left sample: percentage in parentheses is proportion of participants who spontaneously named this group as part of today's U.S. society. Right sample: number in parentheses is millions of occurrences in contemporary (i.e., 2000-2009) American English literature (i.e., ~27 billion words) according to the Google Books Corpus (Davies, 2011).

To create the naturalistic sample, we recorded how often the 136 social groups[7] that were named by at least two participants appear in the Google Books Corpus (Michel et al., 2011), the world's largest collection of digitized and searchable books (> 5 million containing > 500 billion words). To measure these groups' frequency of occurrence in contemporary American English literature, we searched only within the 2000-2009 publication period of the

[7] We added the frequency of occurrence for the synonymic social groups children / kids, old / elderly / seniors, Blacks / African Americans, Church / Christians, military / veterans, rich / wealthy, Indians / Native Americans, athletes / sports, Hispanics / Latinos, elites / upper class, entrepreneurs / business owners, atheists / non-religious, and boy / girl scouts.

American English section of the Google Books Corpus (= 26.9 billion words) provided by Davies (2011; see http://googlebooks.byu.edu/). Taking the average between the Study 1a sample (*N* = 80) and the minimalist sample (*N* = 42), the naturalistic sample included the 61 groups that we found to be most prevalent in this collection of texts (see Table 6). 46 of these 61 groups had also been named by 10% of participants in Study 1a. However, the frequency with which the 46 groups had been named in Study 1a did not predict their frequency of occurrence in contemporary American English literature, $r = .17$, $p = .25$. Noteworthy, the most frequently occurring groups in the naturalistic sample did not score either high or low on either A or B, but rather reflected differences in race, sex and age ('children', 'women', 'old', 'men', 'Whites', 'Blacks').

Study 5b. Character- / personal encounter-based arrangement of 42 / 61 groups. We paid 378 MTurkers (148 women, 230 men; *M* = 33.94 years, *SD* = 10.74) $0.75 to "sort 42 social groups on the computer screen". They arranged either the 42 minimalist groups, or a random selection of 42 of the 61 naturalistic groups. The arrangement task was the same as in Studies 2 and 3 – with two exceptions. First, to give participants an overview of the groups, they appeared all at once in a random order of four columns and eleven rows in the middle of the screen. More importantly, there were three different arrangement instructions. In the similarity (control) condition, participants read: "… social groups whose typical members are similar should be placed closer together, while social groups whose typical members are different should be placed further apart". In the character condition, they read: "… social groups whose typical members have similar characters should be placed closer together, while social groups whose typical members have different characters should be placed further apart". And in the personal encounter condition, they read: "… social groups for which personal encounters with their typical members are similar should be placed closer together, while social groups for which personal encounters with their typical members are different should be placed further apart"[8] (for the minimalist sample, there were between 49 and 51 participants per condition; for the naturalistic sample there were between 74 and 77 participants per condition). As in Studies 2-4, the arranged distances between the groups were recorded as proportions of the screen diagonal.

[8] To ensure that participants follow these different instructions, we presented them not only before, but also during the spatial arrangement phase, namely in abbreviated from (e.g., "similar character -> closer together; different character -> further apart") left to a "Continue" button at the bottom of the screen.

Study 5c: Testing the validity of the 2D ABC model of stereotype content. Next, 201 MTurkers (70 women, 131 men; *M* = 33.03 years, *SD* = 11.07) were paid $0.6 to "rate about 50 social groups on a stereotype dimension". They rated the 42 minimalist groups or the 61 naturalistic groups on either A, B, or one of two versions of C. Each of the eight corresponding 0-10 slider scale items was anchored with a meaning cloud (for an example, see Appendix M; meaning clouds accurately measure groups' A, B, and C, see Footnote 1) of all sub-dimensions of either A, B, C (see Table 3), or C2. Based on recommendations in the literature (Kervyn et al., 2015), the second version of communion was anchored with "Not at all …-Extremely friendly / sincere / sociable / well-intentioned". Above the slider scale items, participants read: "As viewed by society, how … [e.g., friendly, sincere, sociable, and well-intentioned] are members of these groups?" There were between 22 and 28 raters per stereotype dimension, and as in the previous studies, raters' agreement about the groups was very high, all *ICC(2,k)*s >= .85, (McGraw & Wong, 1996). While the expected correlations between the two versions of communion in the minimalistic sample (r = .87), and in the naturalistic sample (r = .86) were large, ps < .001, there was another moderate but statistically significant correlation between agency / socio-economic success and progressive beliefs in the minimalistic sample, r = .34, p < .05, all other $|r|$s < .24, ps > .05[9].

Next, we computed the mean distance between each pair of spatially arranged groups, separately for the minimalist and the naturalistic sample, and separately for the similarity-, character- and personal encounter-based instructions. For the minimalist sample, these mean distances correlated highly across the three different spatial arrangement instructions, mean r = .90, *SD* = .06, and the same was true for the naturalistic sample, mean r = .80, *SD* = .06. Thus, we collapsed mean intergroup distance across the three different spatial arrangement instructions, separately for the minimalist and the naturalistic sample of groups[10].

---

[9] The main difference between the group rating instructions in Studies 1 and 5 was that in Study 5 participants rated the groups "as viewed by society" rather than the self. If participants perceived mainstream society as relatively conservative, then, assuming that society trusts and likes society, participants would have rated conservative groups as scoring higher on C than progressive groups. If so, then this effect should be less pronounced in Study 1, where participants rated the groups as viewed by the self rather than society. However, the groups' B and C correlate r = -.01 in Study 1 and r = .04 (minimalist sample) and r = .05 (naturalistic sample) in Study 5, suggesting that participants in Study 5 perceived mainstream society as neither conservative nor progressive.

[10] Separate property fitting analyses for the 1D-3D spaces extracted from the similarity-, character- and personal encounter-based mean inter-group distances yielded almost identical results to the property fitting analyses for the 1D-3D spaces extracted from the mean inter-group

The mean distances between the groups were subjected to MDS (separately for the minimalist and the naturalistic sample) with the same settings as in the previous studies. As in Studies 3 / 4, based on the goodness of fit of the 1D, 2D, 3D, 4D, 5D and 6D scaling solutions (Table 2), we extracted and analyzed three dimensions, but proceeded with interpreting only the 1D and 2D space of the minimalist and the naturalistic groups. Property fitting analyses confirmed the validity of the 2D ABC model of stereotype content. Agency / socio-economic success, *R(2D axis)* = .93, *p* < .001, and conservative-progressive beliefs, *R(2D axis)* = .94, *p* < .001, were far better *axial* interpretations of the minimalist groups' 2D space than communion, *R(2D axis)* = .13, *p* = .72, and the second version of communion, *R(2D axis)* = .20, *p* = .47. The same was true for the 2D space of the naturalistic groups; A: *R(2D axis)* = .81, *p* < .001; B: *R(2D axis)* = .86, *p* < .001; C: *R(2D axis)* = .12, *p* = .67, and C2: *R(2D axis)* = .15, *p* = .52 (see Table 7; see also Figures 4 and 5). Further, as in the previous studies C and C2 were suitable *polar* interpretations of the minimalist groups' 2D space, *r(2D pole)* = .60, *p* < .001 and *r(2D pole)* = .50, *p* < .001, respectively, and the naturalistic groups' 2D space, *r(2D pole)* = .61, *p* < .001 and *r(2D pole)* = .74, *p* < .001, respectively. In contrast, A and B were not suitable as polar interpretations of these spaces.

---

distances collapsed across the these three spatial arrangement instructions, as shown in Appendix F (minimalistic groups) and Appendix G (naturalistic groups).

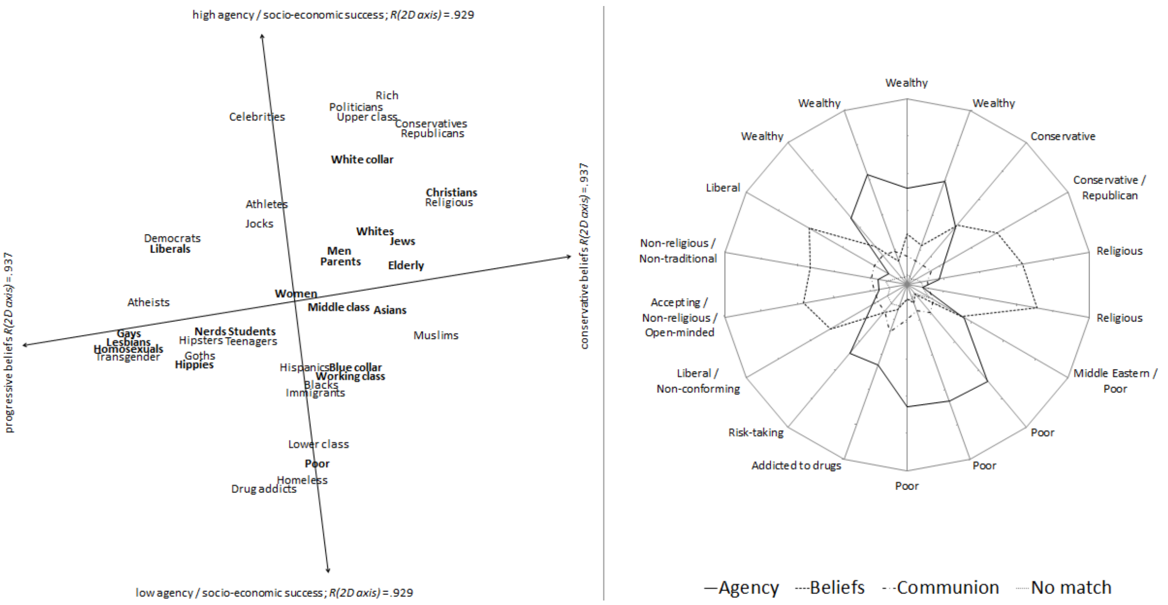
Table 7

Property Fitting Results for Studies 5 and 6

	Group sample	Stereotype content	*R(1D axis)*	*R(2D axis)*	*R(3D axis)*	*r(1D pole)*	*r(2D pole)*	*r(3D pole)*
Study	42 minimalist U.S. groups	Agency (A)	**.777**	**.929**	**.909**	-.188	-.093	-.097
		Beliefs (B)	**.783**	**.937**	**.935**	.136	.143	.159
		Communion (C)	.050	.129	.247	**.551**	**.595**	**.590**
		Communion (C2)	.149	.196	.272	**.521**	**.502**	**.545**
Study	61 U.S. groups	Agency (A)	**.765**	**.806**	**.803**	.072	.086	.116
		Beliefs (B)	.354	**.863**	**.890**	**.487**	.195	.261
		Communion (C)	.097	.118	**.526**	**.432**	**.607**	**.540**
		Communion (C2)	.113	.150	**.569**	**.640**	**.737**	**.665**
Study	42 minimalist U.S. groups	Agency (A)	**.826**	**.917**	**.923**	-.184	-.100	-.078
		Beliefs (B)	**.708**	**.936**	**.945**	.186	.222	.218
		Communion (C)	.058	.055	.084	**.640**	**.642**	**.668**
		Communion (C2)	.181	.194	.327	**.567**	**.567**	**.625**
Study	61 U.S. groups	Agency (A)	**.691**	**.741**	**.906**	-.064	.136	.094
		Beliefs (B)	**.657**	**.901**	**.945**	.391	.339	.328
		Communion (C)	.015	.102	.381	**.480**	**.518**	**.507**
		Communion (C2)	.054	.236	.447	**.619**	**.613**	**.604**

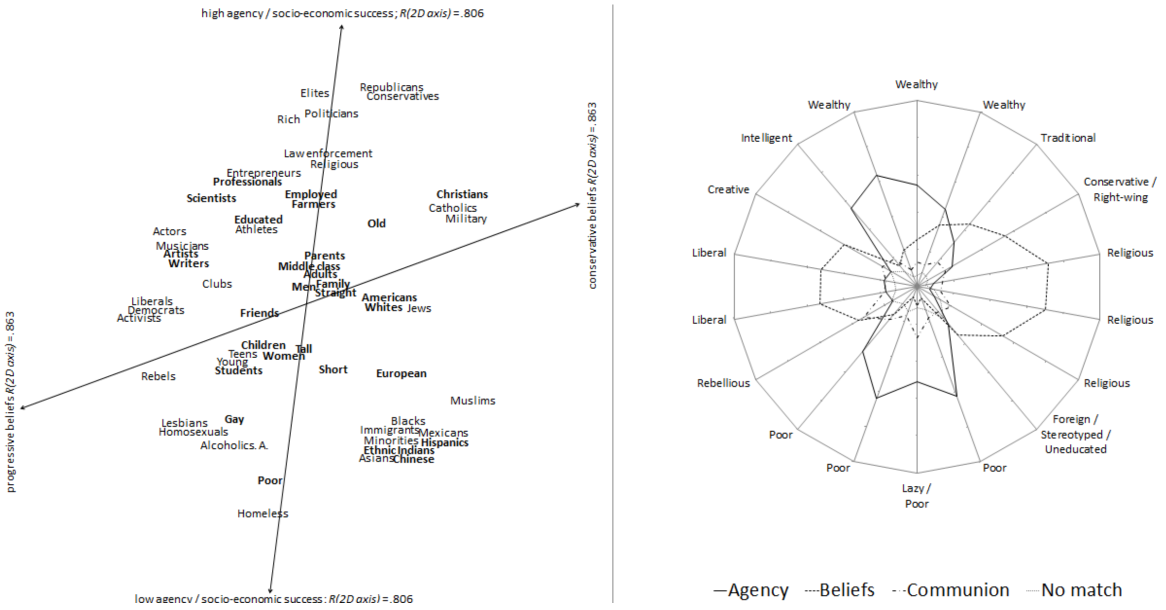
Note. *R(1D-3D axis)* indicate the maximal correlations between the 42 / 61 minimalist / naturalistic U.S. social groups' agency / socio-economic success, conservative-progressive beliefs, and communion (standard and alternative operationalization) ratings and their projections on an axis rotated around the origin of their 1D, 2D, and 3D space; *r(1D-3D pol)* indicate correlations between the social groups' A, B, and C (standard and alternative operationalization) ratings and their proximity to the origin of these three spaces. Bold correlation coefficients are significant at p =< .001.

Figure 4



Note. Study 5: The left side illustrates that the minimalist groups' 2D stereotype space is made up of the axes agency / socio-economic success and conservative-progressive beliefs and the centrally located pole communion (21 most / least communal groups = bold / not bold). The right side plots the most frequent labels for the nine pairs of reversed axes of this space, and the percentage of labels for these axes that were assigned to A, B, C, and 'no match'. All axes mainly reflect A or B at the angle where these two run through the space (see property fitting results on the left side).

Figure 5



Note. Study 5: The left side illustrates that the naturalistic groups' 2D stereotype space is made up of the axes agency / socio-economic success and conservative-progressive beliefs and the centrally located pole communion (most / least communal groups = bold / not bold). The right side plots the most frequent labels for the nine pairs of reversed axes of this space, and the percentage of labels for these axes that were assigned to A, B, C, and 'no match'. All axes mainly reflect A or B at the angle where these two run through the space (see property fitting results on the left side).

Study 5d: Ruling out alternative 2D models of stereotype content. 180 additional MTurkers (82 women, 98 men; M = 31.65 years, SD = 9.38) were paid \$0.75 to "identify 9 person characteristics". We rotated the 2D coordinates of the 42 minimalist groups clockwise around the origin of their space. We rotated in 18 steps of 20° (= a full rotation of 360°). At each rotation step, we formed a 1D ranking based on the groups' current x-coordinates (i.e., after nine rotation steps – a half rotation of 180°– each of these 18 rankings was reversed). We presented participants with the group rankings of nine consecutive rotation steps, one at a time and in random order. These nine rankings represented nine axes that run through (the origin of) the minimalist groups' 2D space in such a way that any so far overlooked stereotype dimension would have a maximal distance of 10° from one of these nine axes. A rotation angle of 10° corresponds to a correlation of r > .98, and thus the nine axes included all fundamental stereotype dimensions that we have overlooked in our previous analyses – if there are any. If agency / socio-economic success and conservative-progressive beliefs are the only two stereotype dimensions encoded in the minimalist groups' 2D space, then the collection of labels for all pairs of reversed axes (i.e., two axes between which the rotation angle is 180°) should reflect A or B and not C or something else.

To understand the task, participants were first presented with an example in which the animal characteristic based on which "giraffe, elephant, horse, deer, dog, mouse, and bee" were ranked one atop the other was labeled as "tall" and / or "big". Then, before labeling each of the nine target axes, participants were presented with the minimalist groups one atop the other in the order of their x-coordinates on the corresponding axis, and they read: "Your task is to identify X. X is the person characteristic based on which the social groups are ranked. As viewed by society, groups at the top of the ranking are extremely X. …groups above the center of the ranking are above-averagely X. …groups at the center of the ranking are averagely X. … groups below the center of the ranking are below-averagely X. … groups at the bottom of the ranking are not at all X. Please enter the person characteristic X (an adjective) in the textbox below. If you have no idea about X, enter "I don't know"" (for an example, see Appendix N). Participants generated a total of 521 labels for the nine pairs of reversed axes of the minimalist groups' 2D space (due to redundancy 274 unique labels; "I don't know" = 38.27% of all cases). We repeated this axes labeling procedure for the 61 naturalistic groups' 2D space. Other participants generated a total of 516 labels for the nine pairs of reversed axes of the naturalistic groups' 2D space (265 unique labels; "I don't know" = 33.33% of all cases).

108 additional MTurkers (47 women, 61 men; *M* = 30.47 years, *SD* = 9.05) were paid $0.5 to "assign 100 person characteristics" one after the other, and they read: "please select the category to which this person characteristic [e.g., wealthy] fits best. If this person characteristic does not fit well to any of the categories, check 'no match'". The seven categories available for selection were 'no match' plus the polar opposites on A, B and C (A+: "powerful / dominating / high status / wealthy / confident / competitive", A-: "powerless / dominated / low status / poor / unconfident / unassertive", B+: "modern / science-oriented / alternative / liberal", B-: "traditional / religious / conventional / conservative", C+: "trustworthy / sincere / likable / benevolent / warm / altruistic", and C-: "untrustworthy / dishonest / repellent / threatening / cold / egoistic") – the candidate stereotype dimensions examined in Studies 1-5. Participants assigned either 100 random of the 274 different labels generated for the (9 pairs of reversed axes =) 18 axes of the minimalist groups' 2D space, or 100 random of the 265 different labels generated for the 18 axes of the naturalistic groups' 2D space. On average, each label generated for an axis of minimalist and naturalistic groups' 2D space was assigned by 21.10 (*SD* = 3.68) and 18.86 (*SD* = 3.41) participants, respectively. For each of the 512 and 516 labels generated

for one of the axes of the minimalist and naturalistic groups' 2D space, respectively, we recorded the percentage of assignments to categories A±, B±, C± and 'no match', a measure of the labels' relatedness to A, B, C, and something else (rel:A, rel:B, rel:C, and rel:Else), respectively. Finally, for each of the nine pairs of reversed axes of the minimalist and naturalistic groups' 2D space, we averaged rel:A, rel:B, rel:C, and rel:Else across all labels generated for that pair of axes.

Table 8 shows mean relatedness of the participant-generated labels to A, B, C, and something else (rel:A, rel:B, rel:C, and rel:Else), separately for the nine pairs of reversed axes of the minimalist groups' 2D space and the naturalistic groups' 2D space. All pairs of axes in both 2D group spaces predominantly related to agency / socio-economic success or conservative-progressive beliefs rather than communion or something else. Therefore, according to participants in the label generation and assignment studies, the 2D space of both the minimalist and the naturalistic groups does *not* encode a fundamental, spontaneously employed stereotype dimension other than A and B.

Table 8

Label Assignment Results for Studies 5 and 6

Relatedness to stereotype content	*Axes 0/ 180°*	*Axes 20/ 200°*	*Axes 40/ 220°*	*Axes 60/ 240°*	*Axes 80/ 260°*	*Axes 100/ 280°*	*Axes 120/ 300°*	*Axes 140/ 320°*	*Axes 160/ 340°*
Study 5									
Minimalist 2D space									
Agency (A)	**59**	**52**	45	23	17	12	22	**57**	**65**
Beliefs (B)	18	18	30	**52**	**60**	**63**	49	18	12
Communion (C)	17	21	16	16	14	16	19	20	17
Something else	7	9	9	9	9	10	10	6	6
Naturalistic 2D space									
Agency (A)	**53**	**63**	44	14	12	13	18	39	**54**
Beliefs (B)	18	14	22	**48**	**62**	**61**	44	31	20
Communion (C)	20	13	19	22	15	16	26	20	14
Something else	9	9	15	16	11	9	12	10	11
Study 6									
Minimalist 2D space									
Agency (A)	**58**	**53**	**49**	23	19	12	11	44	**58**
Beliefs (B)	7	16	17	**47**	**46**	**53**	53	14	14
Communion (C)	27	21	26	24	26	27	30	36	20
Something else	8	9	8	6	9	8	7	6	7
Naturalistic 2D space									
Agency (A)	**50**	**43**	36	21	11	19	19	24	35
Beliefs (B)	18	19	19	35	**57**	**55**	**52**	**49**	34
Communion (C)	21	22	25	29	21	17	20	18	20
Something else	12	15	20	15	11	8	9	8	10

Note. *Axes 0°/180°-160°/340°* indicate the mean percentage of participants who assigned the labels generated for the respective pair of reversed axes of the respective 2D group space to the categories agency / socio-economic success, conservative-progressive beliefs, communion and something else. Participants saw all nine pairs of reversed axes of both 2D group spaces as related to A or B rather than communion or something else. Because the nine pairs of reversed axes include all stereotype dimensions encoded in the respective 2D group space, these results add to the corresponding property fitting results that A and B are the *only* suitable axial interpretations of both 2D group spaces. Bold numbers indicate paired t-tests of the highest against the second highest percentage that are significant at $p =< .001$.

Figures 4 (minimalist groups) and 5 (naturalistic groups) illustrate this. The left side of these figures shows the groups' coordinates in their 2D space plus the two axes that best represent A and B, which together form the best available explanation of the variance contained in this space (see the property fitting results in Study 5). The right side shows the same space (aligned in the same direction) plus the 18 axes for which we collected labels and label assignments to the stereotype dimensions A, B, C, and something else (= 'no match'). For each of these 18 axes, the far end shows the label most often generated for that axis. All these most consensual axes labels are relatable to A or B. More importantly, for each of the 18 axes, the stretch from the spaces' origin to the far end of that axis indicates the percentage of participants who assigned the labels generated for that axis to A, B, C, and something else that "does not fit well" to A, B, and C. As is immediately evident, all nine pairs of reversed axes reflect A or B rather than C or something else, and the axes that reflect A and B are orthogonal to one another and run through the space more or less exactly at the angles that best represent A and B (see the left side of the figures) according to the property fitting analyses in Study 5. That is, in Study 5 we took a data-driven approach not just to scaling the groups' 2D space, but also to interpreting this space, and results showed that A and B is the one and only pair of orthogonal stereotype dimensions that underlies this space.

Discussion

To rule out that the 2D ABC model of stereotype content is limited to the detailed and thus possibly biased instructions under which participants in Studies 1 and 3 named groups, in Study 5 participants received the minimalist naming instructions used by Fiske and colleagues (2002, p. 883; see also Kervyn et al., 2013; 2015), except that we dropped "(i.e., based on ethnicity, race, gender, occupation, ability, etc.)" to avoid priming groups defined by the ethnicity, race, gender, occupation and / or ability of their members. Moreover, to omit groups that are not essential to participants' view of society, participants were free to name any desired number between 3 and 30 groups (see Kervyn et al., 2013; 2015). Both the groups in this 'minimalist' sample and their frequency of naming were highly similar to the Study 1 sample of groups, and Study 5 shows that the 2D ABC model generalizes well from the Study 1 sample to the minimalist sample of groups.

To generalize the model to a naturalistic sample of groups, we recorded the frequency with which all groups that were named at least twice in Study 5 appear in a large text corpus that is arguably representative of contemporary American English literature (Davies, 2011; Michel et al., 2011; for another example of such a linguistic approach to personality and social psychology research, see Akpinar and Berger, 2015). The groups mentioned most often between 2000 and 2009 formed our 'naturalistic' sample, which is somewhat different from the

Study 1's sample of groups. The results of Study 5 showed that the 2D ABC model generalizes to this naturalistic sample as well.

To rule out that the 2D ABC model is limited to the instructions under which participants in Studies 1-4 compared groups, in Study 5 we instructed participants to spatially arrange the minimalist or the naturalistic groups either based on the global dissimilarity of their typical members (see Studies 1-4), based on the dissimilarity of the character of their typical members, or based on the dissimilarity of personal encounters with their typical members. The latter two types of instructions may better reflect the essence of stereotypic social group comparisons (Fiske et al., 2002; Koenig & Eagly, 2014). However, the three types of instructions yielded almost identical group comparisons for both the minimalist and the naturalistic sample, and thus the validity of the 2D ABC model of stereotype content generalizes from unspecified to character- and personal encounter-based group comparisons (see also online supplementary material, Tables osm.6 / osm.7).

To rule out that the 2D ABC model is limited to the instructions under which participants in Studies 1-4 rated groups, in Study 5 participants rated the minimalist and the naturalistic groups' stereotypic agency / socio-economic success, conservative-progressive beliefs and communion as viewed by society (Fiske et al., 2002, Cuddy et al., 2007; Kervyn, Fiske, & Yzerbyt, 2013; 2015) rather than themselves. Rating groups from the perspective of society rather than the self is arguably closer to the definition of stereotypes as socially shared views, and additionally circumvents single individuals' tendency to respond in a socially desirable way that eliminates meaningful variance on valence-related stereotype dimensions such as communion (Fiske et al., 2002). Further, in Study 5 participants also rated both the minimalist and the naturalistic groups on another set of communion subscales that has been reported to better reflect its essence, namely 'friendly / sincere / sociable / well-intentioned'. Results showed that this alternative our version of communion are almost identical in meaning (rs $>=$ .86), and that the ABC model generalizes from groups' A, B, and C as viewed by the self to their A, B, and C / C2 "as viewed by society" (Fiske et al., 2002, p.884)[11].

---

[11] 40 minimalist and 46 naturalistic groups are also part of the Study 1 sample, which allows correlating these groups' stereotypic A, B, and C "as viewed by society" (measured in Study 5) with their A, B, and C as viewed by single persons (measured in Study 1). For the minimalist groups, these A, B, and C correlations are r = .98, .97, and .92, respectively. For the naturalistic groups, the correlations are r = .98, .97, and .88, respectively. Thus, it does not make a difference whether group stereotypes are measured as viewed by society or by single persons.

Furthermore, we aimed to rule out that there are other models of stereotype content that we might have overlooked when selecting candidate stereotype dimensions based on a visual inspection of the 2D group space computed in Study 1. To that end, we asked new participants to label the stereotype dimensions that underlie different group rankings that together reflect all axes (i.e., stereotype dimensions) that run through the origin of the groups' 2D space. Yet other participants then assigned the generated stereotype dimension labels to our candidates A, B, or C as defined in Study 1, and they were instructed to select 'no match' if a label "does not fit well" to A, B, or C. For each axis / stereotype dimension of the groups' 2D space, our results showed that participants predominantly assigned the labels generated for that axis / stereotype dimension to A or B rather than C or something else (= 'no match'), a pattern that was found for both the 2D space of the minimalist and the naturalistic groups (see Table 8 and Figures 4 and 5). Thus, participants spontaneously used A and B (not C or something else) to stereotypically compare the groups.

Study 6

Study 6 addresses another caveat. The 2D ABC model of stereotype content may only apply to distinguishing between the entirety of groups that together form society. In Studies 1-5, participants always compared either all groups in the respective sample, or a randomly drawn set of groups that is more or less representative of all groups in the respective sample. Thus, for their comparisons participants had to spontaneously select stereotype dimensions on which all groups in the respective sample can be meaningfully placed. However, in real life people likely compare self-selected rather than representative or complete sets of groups, maybe because they want to compare some groups on a stereotype dimension on which only those and not all groups can be placed well, for example because they have no idea about the other groups' position or construe them as highly heterogeneous regarding this dimension. In principle it may be that participants predominantly process communion/warmth information when stereotyping groups but forcing them to rate a large number of groups for which they have no clear stereotype about their communion undermines spontaneously employing this dimension. Thus, by omitting a phase in which participants self-select groups to be stereotypically compared, in Studies 1-5 we might have artificially limited the range of stereotype dimensions that participants could spontaneously select to only those dimensions that can be meaningfully applied to all social groups.

Study 6 therefore added a phase in which participants could freely choose the groups that they would subsequently compare. If participants decide to compare different groups on different stereotype dimensions, the scaling of 1D-3D group spaces to be interpreted will entail a poor statistical fit, and these group spaces will not entail meaningful stereotype dimensions that can be interpreted based on candidate stereotype dimensions (see property fitting analyses in Studies 1-5). If, however, participants decide to compare different groups on more or less the same few stereotype dimensions, we will be able to reveal the nature of these fundamental stereotype dimensions if they are among our candidates A, B, and C. We hypothesized that participants decide to compare different groups on A and B rather than C and / or something else.

Methods and results

We paid 751 MTurkers (240 women, 411 men; M = 32.30 years, SD = 10.45) $0.75 to “select and sort 21 social groups on the computer screen”. Participants were presented with either the 42 minimalist groups or 42 random naturalistic groups (out of 61), and were instructed to select at least 21 of these groups to spatially arrange them based on either the global dissimilarity of their typical members, the dissimilarity of the character of their typical members, or the dissimilarity of personal encounters with their typical members. We set a minimum of 21 groups, because selecting half of the available groups holds a balance between increasing the number of stereotype dimensions on which the groups can be placed (i.e., compared) and decreasing the number of participants required to obtain reliable dissimilarity estimates for all possible pairs of groups (861 and 1830 in the minimalist and naturalistic sample, respectively), a necessity for an accurate scaling of the 1D-3D group spaces to be interpreted (Borg & Groenen, 2005). On average, participants selected 21.80 minimalistic and 21.58 naturalistic groups (for frequency of selection of all minimalist and naturalistic groups averaged across the three spatial arrangement instruction conditions, see Appendix H). For the minimalist groups, frequency of selection correlated with frequency of naming in Study 5, r = .40, p < .01; for the naturalistic groups, frequency of selection correlated with frequency of appearance in contemporary (2000-2009) American English literature according to the Google Books Corpus (Davies, 2011), r = .42, p = .001. Apparently, in both the minimalist and the naturalistic condition participants most often selected groups defined by the race, sex and age of their members, and groups perceived as either high or low on either A or B.

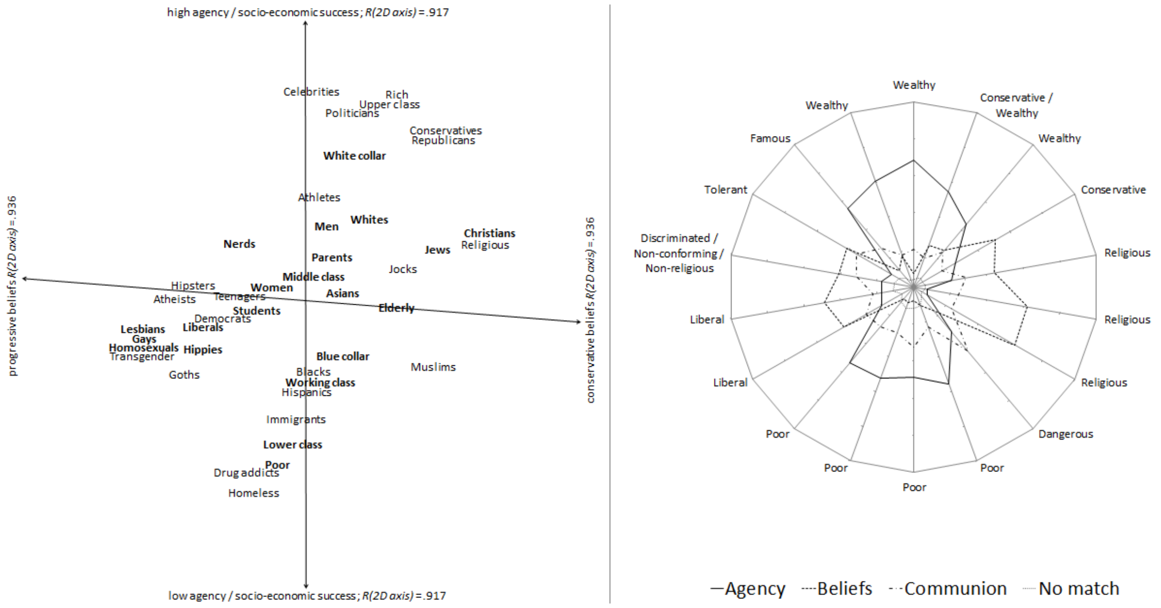
Next, participants spatially arranged the self-selected groups (the instructions and procedure were the same as in Study 5; between 97 and 100 participants per condition for the minimalist sample and between 151 and 154 participants per condition for the naturalistic sample). As in Studies 2-5, the spatially arranged distances between the groups were recorded as proportions of the screen diagonal.

Next, we computed the mean distance between each pair of spatially arranged groups, separately for the minimalist and the naturalistic sample, and separately for the similarity-, character- and personal encounter-based instructions. As in Study 5, these mean distances correlated highly across the three different spatial arrangement instructions (mean $r = .82$, $SD = .09$ for the minimalist groups, and mean $r = .62$, $SD = .03$ for the naturalistic groups), and thus we collapsed mean intergroup distance across the three different spatial arrangement instructions, separately for the minimalist and the naturalistic sample of groups[12].

The mean distances between the groups were subjected to MDS (separately for the minimalist and the naturalistic groups) with the same settings as in the previous studies (Table 2 shows the goodness of fit of the 1D, 2D, 3D, 4D, 5D and 6D scaling solutions). Property fitting analyses confirmed the validity of the 2D ABC model of stereotype content. A, *R(2D axis)* = .92, $p < .001$, and B, *R(2D axis)* = .94, $p < .001$, were far better *axial* interpretations of the minimalist groups' 2D space than C, *R(2D axis)* = .06, $p = .94$, and the second version of C, *R(2D axis)* = .19, $p = .47$. The same was true for the 2D space of the naturalistic groups; A: *R(2D axis)* =.74, $p < .001$; B: *R(2D axis)* = .90, $p < .001$; C: *R(2D axis)* = .10, $p = .74$, and C2, *R(2D axis)* = .24, $p = .19$ (see Table 7, see also Figures 6 and 7). Further, as in the previous studies, C and C2 were suitable *polar* interpretations of the minimalist groups' 2D space, *r(2D pole)* = .64, $p < .001$ and *r(2D pole)* = .57, $p < .001$, respectively, and the naturalistic groups' 2D space, *r(2D pole)* = .52, $p < .001$ and *r(2D pole)* = .61, $p < .001$, respectively. In contrast, A and B were not suitable as polar interpretations of these spaces.

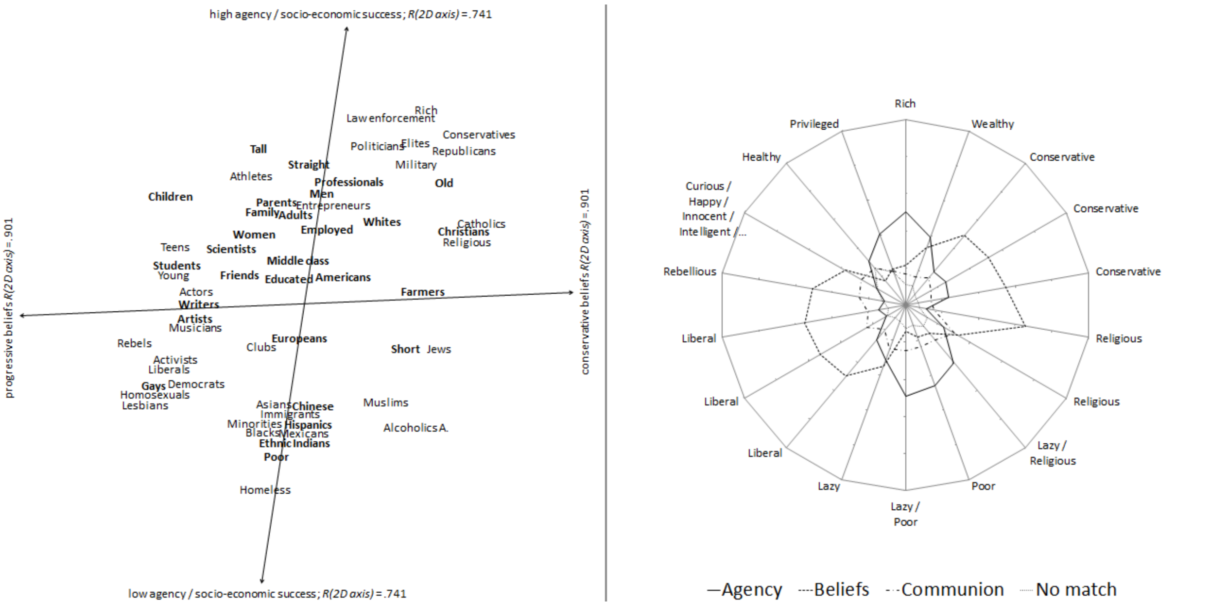
Figure 6

[12] Separate property fitting analyses for the 1D-3D spaces extracted from the similarity-, character- and personal encounter-based mean inter-group distances yielded almost identical results to the property fitting analyses for the 1D-3D spaces extracted from the mean inter-group distances collapsed across these three spatial arrangement instructions, as shown in Appendix I (minimalistic groups) and Appendix J (naturalistic groups).



Note. Study 6: The left side illustrates that the minimalist groups' 2D stereotype space is made up of the axes agency / socio-economic success and conservative-progressive beliefs and the centrally located pole communion (21 most / least communal groups = bold / not bold). The right side plots the most frequent labels for the nine pairs of reversed axes of this space, and the percentage of labels for these axes that were assigned to A, B, C, and 'no match'. All axes mainly reflect A or B at the angle where these two run through the space (see property fitting results on the left side).

Figure 7



Note. Study 6: The left side illustrates that the naturalistic groups' 2D stereotype space is made up of the axes agency / socio-economic success and conservative-progressive beliefs and the centrally located pole communion (most / least communal groups = bold / not bold). The right side plots the most frequent labels for the nine pairs of reversed axes of this space, and the percentage of labels for these axes that were assigned to A, B, C, and 'no match'. All axes mainly reflect A or B at the angle where these two run through the space (see property fitting results on the left side).

219 additional MTurkers (88 women, 131 men; M = 31.35 years, SD = 9.51) were paid $0.75 to "identify 9 person characteristics". As in Study 5, to find out if our candidates A, B, and C do not include one or more fundamental stereotype dimensions encoded in the groups' 2D space, we asked participants to label nine pairs of reversed group rankings that represent nine axes that run through (the origin of) this space in such a way that one of them correlates at least r = .98 with any stereotype dimension not included in A, B, and C. If, as predicted, A and B are the only two stereotype dimensions encoded in the groups' 2D space, then the collections of labels for all nine axes should predominantly reflect A and B.

Participants generated a total of 758 labels for the nine axes of the minimalist groups' 2D space (due to redundancy 347 unique labels; "I don't know" = 22.25% of all cases), and other participants generated a total of 730 labels for the nine axes of the naturalistic groups' 2D space (332 unique labels; "I don't know" = 33.16% of all cases).

120 additional participants (54 women, 66 men; M = 33.02 years, SD = 10.82) were paid $0.5 to "assign 100 person characteristics" to A, B, C, or 'no match' (see Studies 4 / 5).

Participants assigned either 100 random of the 347 different labels generated for the nine axes of the minimalist groups' 2D space, or 100 random of the 332 different labels generated for the nine axes of the naturalistic groups' 2D space. On average, each label generated for an axis of minimalist and naturalistic groups' 2D space was assigned by 16.92 (*SD* = 3.39) and 18.36 (*SD* = 3.41) participants, respectively. For each of the 758 and 730 labels generated for one of the axes of the minimalist and naturalistic groups' 2D space, respectively, we recorded the percentage of assignments to categories A, B, C and 'no match', a measure of the labels' relatedness to A, B, C and something else, respectively. Finally, for each of the nine axes of the minimalist and naturalistic groups' 2D space, we averaged relatedness to A, B, C and something else across all labels generated for that axis.

Table 8 shows mean relatedness of the participant-generated labels to A, B, C and something else, separately for the nine axes of the minimalist groups' 2D space and the naturalistic groups' 2D space. As in Study 5, all axes in both 2D group spaces predominantly relate to agency / socio-economic success or conservative-progressive beliefs rather than communion or something else (see Figures 6 and 7). Thus, according to the participants in the label generation and assignment studies, the two 2D spaces do *not* encode a fundamental stereotype dimension other than A and B.

Discussion

Study 6 examined if the 2D ABC model generalizes from comparing all or representative samples of groups to individually tailored samples of groups. In real life people stereotypically compare self-selected rather than representative or complete samples of groups. In doing so, they may compare different group samples on different stereotype dimensions. If this had been the case, then the MDS algorithm (Young, Takane, & Lewyckyj, 1978; see also Borg & Groenen, 2005) applied throughout this paper would have resulted in a statistically poor-fitting and thus uninterpretable 2D group space.

Quite to the contrary, the results of Study 6 showed that people stereotypically compare different selections of groups on the same dimensions, namely agency / socio-economic success and conservative-progressive beliefs, and that communion again emerges as centrality in the well-fitting and thus interpretable 2D stereotype space spanned by A and B. Moreover, Study 6 fully replicated Study 5, providing further empirical support for our conclusions that the 2D ABC model is valid across different approaches to sampling groups (from memory vs. text corpora), comparing groups (globally vs. character- vs. personal encounter-based), and rating groups (as viewed by society vs. single persons). Finally, as in Study 5, we applied a data-driven approach to scaling the groups' 2D space (see left side of Figures 6 and 7), and to interpreting this space. Results show that there is no other 2D model of stereotype content that we have overlooked (see Table 8 and the right side of Figures 6 and 7; see also the Discussion of Study 5).

Before we will elaborate on these results in detail, we put the model to a test that goes beyond the description of a similarity structure and shows that the groups' positions on the two fundamental stereotype content dimensions A and B have downstream consequences.

Study 7

Stereotypes about groups are an effective and efficient tool to plan social interactions (Gilbert & Hixon, 1991; Hamilton, Sherman & Ruvolo, 1990; Pattyn et al., 2013; Sherman, Lee, Bessenoff, & Frost, 1998; Unkelbach, Forgas, & Denson, 2008). If the 2D ABC model of stereotype content is valid, then people should rely predominantly on A and B to make predictions about the states and dynamics of their social environment. In Study 7, we explored how people make predictions about the likelihood of members of different social groups being in the same place at the same time (judgments of time-space proximity), and about the likelihood of members of different groups being friends with one another (judgments of interpersonal liking). Particularly the latter constitutes a critical test of the relatively greater weight of A and B compared to C. Interpersonal liking is clearly a judgment about a communal aspect of interpersonal behavior, allowing the assumptions that it is particularly likeable people who like each other. If, however, interpersonal liking is seen as a function of A and / or B this would translate into participants' estimation that people like each other when they share power, status, and dominance, and / or conservative or progressive beliefs.

Method

Participants and stimuli. We paid 214 MTurkers (84 women, 130 men; M = 34.72 years, SD = 11.55) $0.6 to "sort 40 social groups into 5 categories of social groups". As in Studies 2 and 4, people were presented with a random sample of 40 of the 80 social groups that had been named by at least 10% of all people who had named U.S.-representative social groups in Study 1. Participants were randomly assigned to one of two conditions. In one condition, participants sorted groups according to their *time-space proximity*, while in the other condition, participants sorted social groups according to their *interpersonal liking*.

Procedure. On the first screen slide, people read: "… please drag-and-drop each of these 40 social groups into one of the category boxes presented below." Thereafter, in the time-space proximity condition, 112 participants read on: "Members of social groups that are likely to be in the same place at the same time should be placed into the same category box. Members of social groups that are unlikely to be in the same place at the same time should be placed into different category boxes." In the interpersonal liking condition, 102 participants read on: "Members of social groups that like one another should be placed into the same category box. Members of social groups that do not like one another should be placed into different category boxes." Below, on the left side of the screen were 40 groups below one another in random order. On the rights side were five unlabeled category boxes below one another. To finish the task, participants were instructed to sort all 40 groups into any number between two and all five category boxes.

Results

First**,** we calculated the likelihood of being sorted into the same category separately for each unique pair of social groups (n = 3160) across all participants who had sorted that pair, and separately for the 112 and 102 people who had categorized the groups based on space-time proximity (M = 27.64 participants per pair, SD = 3.89) and interpersonal liking (M = 25.15 participants per pair, SD = 3.69), respectively. Second, based on the data from Study 1c, we calculated the absolute A, B, and C difference for all unique pairs of groups (e.g., the agency of students and scientists is 41.62 and 67.75 on a 0-100 scale, respectively, and thus their absolute agency difference is 26.13). And third, to test whether people sorted groups with similar and different A, B, and C ratings into the same and different time-space proximity and interpersonal liking categories, respectively, we ran two multiple linear regressions with the 3160 group pairs' likelihood of being sorted into the same time-space proximity and interpersonal liking category as criterion and with the group pairs' absolute A, B, and C differences as predictors.

The results displayed in Table 9 (all regression weights were highly significant due to the high number of degrees of freedom) showed that similarity in both A and B was used as an indicator of a high likelihood of time-space proximity in real life. Thus, participants estimated that groups that share a position on either of the two dimensions are also particularly likely to meet in real life. Importantly, both effects were additive in that they explained unique variance. C did have a comparably small impact on whether groups were seen as particularly likely to be in the same place at the same time. Results for the second criterion, the likelihood that persons from the 80 groups were judged to like one another, mirrored the results for time-space proximity: similarity in A and B strongly and independently influenced whether group members were judged to like one another (i.e., the more similar two social groups are on these two dimension, the more likely their members are judged to like one another). Similarity in C again had a much smaller influence on judgments of interpersonal liking.

Table 9

Study 7: Pairwise Time-Space Proximity and Interpersonal Liking Simultaneously Predicted by Pairwise Absolute Rating Distance on Agency / Socio-Economic Success, Conservative-Progressive Beliefs, and Communion

Predictors	*ß*	*t*	*p* <=	*r*	pr^2	sp^2
Criterion: time-space						
Distance on Agency (A)	-.431	-28.74	.001	-.421	-.456	-.426
Distance on Beliefs (B)	-.361	-24.34	.001	-.341	-.398	-.360
Distance on Communion (C)	-.038	-2.56	.01	-.105	-.046	-.038
Criterion: interpersonal liking						
Distance on Agency (A)	-.365	-23.39	.001	-.359	-.384	-.361
Distance on Beliefs (B)	-.344	-22.30	.001	-.327	-.369	-.344
Distance on Communion (C)	-.059	-3.78	.001	-.116	-.067	-.058

Note. 3160 unique U.S. social group pairs' likelihood of being sorted into the same time-space proximity or interpersonal liking category simultaneously predicted by the group pairs' absolute rating distance on A, B, and C, our candidate stereotype content dimensions; *r*, pr^2, and sp^2 denote zero order, partial, and semi-partial correlation.

Discussion

As predicted, participants predominantly relied on stereotypic A and B, and substantially less on C to predict the time-space proximity and interpersonal liking of members of 80 representatively sampled social groups. More precisely, participants thought that members of different groups can likely be found in the same place at the same time if the groups are similar in A and B. Further, they thought that members of different groups are likely to like one another depending on the groups' similarity regarding A and B as well.

This is consistent with the similarity breeds liking literature, which shows that similarity in values and attitudes (i.e., B) is particularly relevant for interpersonal liking (e.g., Byrne, 1971; Collisson & Howell, 2014). Likewise, assortative mating (i.e., a mating pattern in which similar partners mate more frequently with one another that would be expected by chance) has been shown to occur not only along dimensions of physical traits (e.g., attractiveness, height), but also along socio-economic status, intelligence, religious beliefs, and political ideology (Kail & Cavanaugh, 2010).

In sum, Study 7 provided further support for A and B as fundamental dimensions that people spontaneously employ to distinguish between social groups.

General Discussion

Previous research shows that people are able to employ estimates of warmth and competence to distinguish between social groups (e.g., Bergsieker et al., 2012; Cuddy et al., 2007; Fiske et al., 2002; Kervyn et al., 2013; 2015), and that a group's position on these two stereotype dimensions matters, because it has consequences for people's emotional and behavioral reactions and responses to that group (Becker & Asbrock, 2012; Cikara & Fiske, 2012; Cuddy et al., 2007). However, regardless of their undisputed importance, there is no evidence that warmth and competence are fundamental in the sense that people *spontaneously* use these two and not other stereotype dimensions to distinguish between groups. To test this, people need to be free to use any stereotype dimensions that they want to use. This paper presents such a data-driven approach to the assessment of the dimensionality and content of spontaneous stereotypes about groups.

Complying with Brunswik's (1956) call for representative designs, in our studies participants freely selected not only stereotype dimensions, but also groups, because stimulus samples selected by researchers are often biased towards their theories (Fiedler, 2011). To implement this research design, we asked participants to name a fixed or self-chosen amount of types of people that today's U.S. / German society categorizes into groups (Fiske et al., 2002). In 6 Studies, other participants spatially arranged either the most frequently named groups, or the groups that appear most often in contemporary American English literature on stereotype dimensions of their free choice. According to our multidimensional scaling analyses, the spatially arranged mean distances between these groups could best be described by a two-dimensional space. According to our property fitting analyses as well as several studies in which other participants generated and categorized labels for virtually all axes that run through the origin of this space, the single best pair of more or less orthogonal stereotype axes that underlie the space was *agency / socio-economic success* (A; powerless-powerful, dominated-dominant, low status-high status, poor-wealthy, unconfident-confident, and unassertive-competitive) and *conservative-progressive beliefs* (B; traditional-modern, religious-science-oriented, conventional-alternative, and conservative-liberal). In other words, variation in spontaneous stereotype content about groups could best be described by A and B. We obtained evidence for this two-dimensional model from U.S. online and German lab samples and from various data-driven approaches to measure spontaneous representations of groups: sequential similarity judgment, simultaneous spatial arrangement with respect to global, character-, and personal encounter-based similarity, as well as spatial arrangement with prior labeling of two similarity axes. Moreover, our data suggests that of the two fundamental dimensions A is primary, and B is secondary. Across studies, agency / socio-economic success was regularly the best interpretation of a one-dimensional group distances scaling solution (i.e., a one-dimensional stereotype space), and was most often named as most important for distinguishing between the groups.

Speculating why people compare groups in terms of their stereotypic A and B, not C

Why agency / socio-economic success? Social hierarchies are millions of years old and even today ubiquitous, not just in adults, but also in children and species other than humans. Social hierarchies satisfy people's need for structure, stability, identity, and safety (Jost & van der Toorn, 2012), and satisfy people's need to maintain a shared reality that coordinates social interaction for the common good (Magee & Galinsky, 2008). Specifically, keeping track of social rank is instructive about who needs to be concerned with whose perspective and feelings (Galinsky, Magee, Inesi, & Gruenfeld, 2006; van Kleef, De Dreu, Pietroni, & Manstead, 2006), about who is constrained and who is free to do and speak their mind (Berdahl & Martorana, 2006), about who speaks and who listens (DePaulo & Friedman, 1998), and about who tells whom what to do.

Perhaps more importantly, people keep track of social rank, because doing so is critical for their individual good. Social groups that are higher in rank hold the key to what people need and want – be it health (e.g., doctors), wealth (e.g., managers), entitlement (e.g., lawyers), insight (e.g., teachers), or voice (e.g., politicians). Thus, to reach their goals, people must keep track of and connect well with groups of higher rank. Also, people want to rise in social rank to have greater access to what they need and want, and to increase their influence on other groups. In a nutshell, distinguishing between groups based on their A might be essential for feeling secure, for managing cooperation, for reaching goals, and for climbing up the social ladder by approaching, attaching to, and blending in with groups of higher rank (Magee & Galinsky, 2008).

Previous research on fundamental dimensions of social perception has come to similar conclusions: A or competence (which is correlated with A, but distinct) is an integral part of virtually any such model, be it under these labels or under labels like instrumentality (Parsons & Bales, 1955), intellectual desirability (Rosenthal et al., 1968), self-profitability (Peeters, 1992), or self-enhancement (Schwartz, 1994). Ultimately, A is considered to be functional both evolutionarily and culturally (Anderson, Hildreth, & Howland, 2015; Cheng, Tracy, Foulsham, Kingstone, & Henrich, 2013; Fiske et al., 2007).

Why conservative-progressive beliefs? Less consistent with previous research is our finding that the second fundamental dimension on which people spontaneously distinguish social groups is whether they are engines of change or preservers of the status quo – that is, their position on the dimension of *progressive-conservative beliefs*. We speculate that knowing whether the ideological beliefs of a group are conservative or progressive comes with a lot of valuable insights about the ways in which that group intends to use the influence that it has on other groups, and about the ways in which members of that group think, feel and behave. In line with the idea that humans are intention detectors and often prioritize intentions over outcomes (e.g., Ames & Fiske, 2013, 2015), B may inform individuals about the general intentions of groups at a societal level. Generally speaking, conservative groups (e.g., Christians, Republicans, elderly and the military) want things to be uniform and stay the way they are, and thus they emphasize religion, traditions, conventions, and conformity. Interacting with conservatives provides people with feelings of stability, predictability, control, safety, comfort and belonging (for a review, see Jost, Federico, & Napier, 2009; Schwartz & Bilsky, 1987; 1990). In contrast, progressive groups (e.g., techies, actors, hipsters, and homosexuals) want things to change and diversify, and thus they emphasize freedom, autonomy, creativity, innovation, (technological, economic, legal etc.), modern subculture (art, music, literature etc.) and media, and alternative views and lifestyles. Interacting with progressives provides people with feelings of curiosity, stimulation, expansion, entertainment, and distinctiveness (Leonardelli, Pickett, & Brewer, 2010; Schwartz, 1994). Thus, keeping track of the ideological beliefs of groups might serve at least two functions: it helps people to anticipate and handle the form and content of social interactions (e.g., politely agreeing with somebody versus dressing up in an outrageous way), and it enables people to strike a balance between cognitive, emotional, and behavioral exploitation (conservative groups) and exploration (progressive groups; Jost et al., 2009).

Managing the trade-off between exploiting available resources of certain quality and quantity and exploring alternative resources of uncertain quality and quantity is fundamental to self-regulatory success (Cohen, McClure, & Yu, 2007; Hills, Todd, & Goldstone, 2010; Inzlicht, Schmeichel, & Macrae, 2014) both culturally (i.e., in the last couple of millennia) and evolutionarily (i.e., since the beginning of life). That is, adults, children, other primates and many other beings have always had to choose between current and alternative habitats, shelters, occupations, foods, mates etc., and these choices have always been important to survive and thrive. Based on our results it could be argued that in today's society it is the conservative and the progressive groups that provide access to current, certain and alternative, uncertain resources, respectively. Therefore, to successfully manage the ancient and ubiquitous exploitation-exploration trade-off, today's citizens might mentally organize groups along the stereotype dimension of conservative-progressive beliefs.

To further explore whether B is in fact a fundamental stereotype dimension that informs individuals about group-specific opportunities for exploitation and exploration, we asked additional participants to rate the groups that we examined in Studies 1-4 on seven stereotype dimensions that map onto exploitation-exploration[13]. With one exception ('prevention-promotion'), these stereotype dimensions ('familiarity-novelty', 'safety-freedom', 'comfort-stimulation', 'loyalty-autonomy', 'preservation-change', and 'uniformity-diversity') correlated strongly with B[14] (mean r = .68, all ps < .001) – in fact as strongly as the correlations between the four stereotype dimensions that form B (mean r = .70, all ps < .001).

[13] We paid 166 participants (67 females, 99 males; M = 42.80 years, SD = 7.91) $0.75 to rate the 80 Study 1-4 U.S. groups on one of seven stereotype dimensions that map onto exploitation-exploration: 'familiarity-novelty', 'safety-freedom', 'comfort-stimulation', 'loyalty-autonomy', 'preservation-change', 'uniformity-diversity', and 'prevention promotion'. Participants read: "To what extent do these 80 groups stand for … [e.g., safety vs. freedom] We are not interested in your personal view, but in how you think these 80 groups are viewed by today's society". Then, as in all studies reported here, they used 0-100 slider scales to rate the groups one atop the other in random order. There were between 21 and 27 raters per stereotype dimension, and as in the previous studies, raters' agreement about the groups was very high, all *ICC(2,k)*s >= .79, (McGraw & Wong, 1996).

[14] The exploitation-exploration stereotype dimensions 'familiarity-novelty', 'safety-freedom', 'comfort-stimulation', 'loyalty-autonomy', 'preservation-change', and 'uniformity-diversity' correlated strongly with B (mean r = .68, all ps < .001), but not with A (mean r = .36, four out of six ps < .001) and C (mean r = .24, none of the ps < .001).

Further, we combined the four stereotype dimensions that form B with the six exploitation-exploration stereotype dimensions. *Exploitation-exploration* and B were equally suitable for disambiguating the U.S. groups' 2D stereotype spaces reported in this paper, mean *R(2D axis)*s = .89 and .90 (*SD*s = .04 and .04), respectively, *p*s < .001. However, because the participant-generated labels for the horizontal and vertical axis of the 2D arrangement board in Study 4 and the participant-generated labels for the nine pairs of reversed axes of the 2D stereotype spaces scaled in Studies 5 and 6 mainly reflected B ('religious', 'traditional', 'conservative', 'non-religious', 'non-traditional', and 'liberal', see Figures 4-7) and not exploitation-exploration, it seems that on the manifest level individuals spontaneously use B to distinguish between groups. However, striking a balance between exploitation and exploration might be the latent regulatory function that distinguishing between conservative and progressive groups tries to serve.

Why not communion? Lastly, our data-driven model deviates from existing theoretical approaches in the role of *communion* or warmth. Classical models construe C as an orthogonal stereotype dimension (Fiske et al., 2002; 2006; Cuddy et al., 2007) that has processing priority over all other information (Abele & Bruckmüller, 2011; Brambilla, Rusconi, Sacchi, & Cherubini, 2011; Wojciszke, 1994; Wojciszke, Bazinska, & Jaworski, 1998). Following the functional logic developed above, one could of course ask why individuals should pay attention to whether the intentions of a group are communal or not, if that group does not have the A to implement its intentions (e.g., children, homeless, drug users, and agnostics; for a previously posed similar question, see Fiske et al., 2002). Consistent with the order of priority suggested by this question, our data showed that C is an emergent quality that is not independent from other stereotype dimensions but follows from A. Groups that are seen as particularly unagentic (e.g., homeless, welfare recipients) or overly agentic (e.g., rich, managers) are also seen as least trustworthy, sincere, likeable, warm, benevolent, and altruistic. Perhaps those groups are seen as contributing too little to society and profiting too much from society, respectively. As communion emerges at the center of the A dimension, it can be reconciled with the 2D AB model of stereotype content.

Importantly, this finding is not an artefact of asking for spatially arranged dissimilarity judgments and ratings on two self-labeled stereotype dimensions. Even if we completely ignore the multidimensionally scaled and property fitted dissimilarity data presented in Studies 1-6, and consider only on the ratings of A and C, it becomes apparent that these dimensions are not independent. Groups' C is the higher the more average their A is: $r = .40$, $p < .001$ (U.S. groups in Studies 1-4), $r = .44$, $p < .001$ (German groups in Study 3), $r = .51$, $p < .001$ (minimalist

U.S. groups in Studies 5 and 6; $r = .35$, $p < .01$ (naturalistic U.S. groups in Studies 5 / 6)[15]. This new look on communion as average agency is entirely consistent with the abundant literature that people trust and like typical, average faces and trait scores more than atypical, extreme faces and trait scores (Langlois & Roggman, 1990; Peabody, 1967; Potter, Corneille, Ruys, & Rhodes, 2007; Rhodes, Halberstadt, & Brajkovich, 2001; Sofer et al., 2015).

Theoretical implications of the ABC model of stereotypes about groups

The 2D ABC model allows for a new perspective on the well-established effects of compensation between warmth and competence (Kervyn, Yzerbyt, Judd, & Nunes, 2009; Yzerbyt, Kervyn, & Judd, 2008). Although warmth and competence are conceptualized as orthogonal dimensions in Fiske and colleagues' (2002) stereotype content model, individuals who are described as particularly competent are systematically inferred to be relatively cold (Kervyn, Bergsieker, & Fiske, 2012). Although none of the dimensions in the 2D ABC model was best described as competence, we observed a similar relation between C (high overlap with warmth) and A (related to competence) with one important qualification. In the 2D ABC model, the compensation between A and C should only hold for the upper half of the A dimension: moderately agentic groups are more communal than highly agentic groups because C is inferred from centrality on the A dimension. Importantly, our model makes further predictions that are in contradiction to general compensation effects. Groups less agentic than average will also be less communal. Starting from a very low position on A (e.g., drug users, homeless), an increase in a group's A towards the average will also lead to more favorable C impressions.

Given that stereotypic C (but not so much A and B) can be taken as a proxy for stereotypic valence[16], this new perspective on stereotypic C also allows further delineations. If

[15] Consistent with our finding that C emerges from A but not B, the social groups' C is not the higher the more average their B is, $r = -.17$, $p = .13$ (U.S. groups in Studies 1-4), $r = .06$, $p = .60$ (German groups in Study 3), $r = -.10$, $p = .55$ (minimalist U.S. groups in Studies 5 / 6), and $r = -.28$, and $p < .05$ (naturalistic U.S. groups in Studies 5 / 6; this correlation is the only exception, and it is rather weak).

[16] We paid 25 MTurkers (16 females, 9 males; $M = 42.80$ years, $SD = 7.91$) \$1 to rate the valence ('worse-better') of the 80 U.S. groups examined in Studies 1-4. Valence correlated with C, $r = .78$, $p < .001$, but neither with A, $r = -.01$, $p = .90$, nor with B, $r = .07$, $p = .55$. Note that this does not contradict the linear relation between valence and A as found by Abele and Wojciszke (2007) or Suitner and Maas (2008). These and other studies lack extremely agentic stimuli (e.g., 'aggressive', 'reckless', and 'conceited' rather than just 'assertive', 'brave', and

C emerges as average A, then after a certain point (i.e., being exactly average on A), social groups cannot be stereotyped as more communal, while they can always be stereotyped as less communal, because there is no limit to being more extreme in terms of A. This is consistent with the notion that negative stimuli are stronger (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001), more dominant / contagious (Rozin & Royzman, 2001), and more mobilizing (Taylor, 1991) than positive stimuli. Finally, maximal C and thus the highest positive valence at average agency is also consistent with the *density hypothesis* (Unkelbach, 2012; Unkelbach, Fiedler, Bayer, Stegmüller, & Danner, 2008), that is, the notion that positive stimuli are more similar to one another than negative stimuli (see also Alves et al., 2015; Koch et al, 2016a; Koch et al., 2016b). More precisely, if increasingly communal groups are increasingly close to the center of the A dimension, then they must be increasingly similar to one another, just because by necessity they are also increasingly close to one another. If so, then the groups' C should be related to their average similarity to all other social groups. This was indeed the case: $r = .38$, $p < .001$ (Study 1), $r = .59$, $p < .001$ (Study 2), $r = .79$, $p < .001$ (Study 3), $r = .53$, $p < .001$ (Study 4), $r = .64$ and $r = .67$, both *p*s < .001 (Study 5; minimalist and naturalistic groups, respectively), and $r = .53$ and $r = .42$, both *p*s < .001 (Study 5; minimalist and naturalistic groups, respectively. In Tolstoy's (1873-1877/2001) terms: communal social groups are all alike (i.e., they are all average on A); but every non-communal social group is non-communal in its own way (being either higher or lower than average on A).

Limitations and future directions

The studies described here leave open whether there are spontaneous / fundamental stereotype content dimensions other than agency / socio-economic success and conservative-progressive beliefs. With the exception of Study 1, in all studies reported in this paper, the 2D spatial arrangement board (Hout et al., 2013) that people used prompted them to spontaneously select no more than two unrelated stereotype content dimensions. Thus, it is possible that there is consensus about a third, fourth, fifth etc. spontaneous / fundamental stereotype content dimension that our research designs did not reveal. The third dimension might actually be communion, as communion was (not highly, but) to some extent suitable as a third independent dimension in Studies 1, 4, and, in part, 5. Although the question of whether

---

'confident'), and thus they found a linear relation. In our stimulus sample, there are extremely agentic groups (e.g., 'rich'; 'celebrities', 'elites', 'upper class', 'managers', 'politicians', 'lawyers'), and thus we find a quadratic relation between valence and A (more precisely, a linear relation between valence and averageness on A), $r = .31$, $p < .01$.

there are more than the two spontaneous / fundamental stereotype content dimensions is informative, insights about additional dimensions would not speak against our assertion that the two most spontaneous / fundamental ones are A and B. C as the third, fourth, fifth etc. fundamental stereotype content dimension would also be compatible with C as average A, as was found in Studies 1-6.

Despite the highly consistent results, our studies speak to the relatively abstract question of how individuals distinguish between all societal groups. In motivating our research we have labeled this approach to stereotyping the “lay sociologist” perspective and related it to previous research that (at least at the core of its empirical contribution) has followed a similar aim (e.g., Fiske et al., 2002, Cuddy et al., 2007). Stereotypes might, however, not only guide how people distinguish between all societal groups, but may also serve as knowledge structures that individuals recruit in social interactions with members of proximal groups (the “relational” perspective; e.g., Cambon, Yzerbyt, & Yakimova, 2014). It may be that stereotypical knowledge about the communion/warmth of such proximal groups receives relatively greater processing priority in social interactions compared to people’s perception of more remote groups.

More precisely, encountering members of proximal groups may elicit an affective, evaluative response that leads people to spontaneously construe these groups in terms of their perceived C rather than A and / or B. The results of Studies 5 and 6 speak against this idea, as participants spontaneously employed A and B rather than C to distinguish between encounters with members of all societal groups. However, maybe participants in Study 5 and 6 did not identify strongly and / or did not strongly oppose identifying with many of these groups, reducing the salience of C as a dimension. To test if C is most important for distinguishing between encounters with members of predominantly proximal groups, in future studies people should spatially arrange mostly proximal groups that they strongly identify with and / or strongly oppose identifying with. Nevertheless, we argue that distinguishing between all societal groups, and possibly also between proximal groups, in terms of their perceived A and B is functional and important, too. As stated above, groups’ A and B are informative about opportunities for reaching goals and climbing up the social ladder (A) and opportunities for exploitation versus exploration (B).

Even without the relational aspect, individuals may think differently about individuals than about social groups. Future studies might thus consider prompting people to name representatives of groups (e.g., Natalie Portman for actors and Pope Francis for Catholics), and to then spatially arrange these representatives rather than the groups they represent. Alternatively, people might spatially arrange individuals that do not markedly represent any particular group(s). Such individualized processing may bring communal information to the forefront so that participants spontaneously judge the group representatives / individuals primarily along the line of how trustworthy and friendly they perceive them to be. This would be one possibility to reconcile our group-based 2D ABC model with the finding that communion enjoys a privileged position in processing of information about individuals (e.g., Abele & Bruckmüller, 2011; Willis & Todorov, 2006; Wojciszke & Abele, 2008).

Given that data-driven approaches to modeling face perception show that dominance (or agency / socio-economic success), youthfulness versus agedness (one could argue that people with a youthful and aged face are likely to hold progressive and conservative beliefs, respectively), and trustworthiness (or communion) are fundamental dimensions (Oosterhof & Todorov, 2008; Sutherland et al., 2013), it seems promising to explore the extent to which the space of facial stereotypes also follows the 2D ABC pattern developed in this paper. There is already initial evidence that faces with more average features are perceived as more communal (Sofer et al., 2015; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). However, it is also conceivable that faces prompt a more individualized social information processing than abstract group labels, so that C is an independent dimension that is given more weight than A and B (Sutherland et al., 2013). In any case, the 2D ABC model suggested here based on a bottom-up, data-driven approach must now be tested in a top-down, theory-driven research program.

Our results consistently show very high overlap (up to $R >= .90$) between the axis rotated around the origin of the social groups' 1D-3D spaces and the independently gathered ratings of the groups on candidate stereotype content dimensions. Although this overlap is almost suspiciously high (cf. Vul, Harris, Winkielman, & Pashler, 2009), it should be stressed that we correlated data on a very high level of aggregation. Specifically, we correlated dissimilarity averaged across individuals with stereotypic A, B, and C averaged across individuals, which removed all variance due to inter-individual differences in judging the dissimilarities between the social groups and their A, B, and C. Thus, our data reflect correlations of group-level averages (social groups as cases), and not averaged individual-level correlations (participants as cases). Thus, our group-level effect sizes do not allow conclusions about individual-level effect sizes (Brand & Bradley, 2012; Brand, Bradley, Best, & Stoica, 2011; Monin and Oppenheimer, 2005). This does *not* threaten the validity of our 2D ABC model of stereotype content about groups, as stereotypes are defined as group-level effects (i.e., groups "as viewed by society"; Fiske et al., 2002, p. 884).

Finally, just like the SCM (e.g., Fiske et al., 2002), the 2D ABC model does not address how people's group identities influence stereotypes about groups (Smith, 1993; Mackie, Smith, & Ray, 2008). Obviously, in-group versus out-group memberships must influence stereotype content. That is, individual (or intergroup) differences in group stereotypes are lost in averaging across raters. The model therefore addresses stereotypes as shared knowledge structures. Nevertheless, Individual (or intergroup) differences in group stereotypes are a fascinating topic for future research. For example, it could be that communion remains a centrally located polar dimension also at the individual level (existing data suggest that this is so; Imhoff & Koch, 2016). It may be that even raters who are extreme on A or B still see groups average on these dimensions as most trustworthy. Alternatively, it might be that for these raters C transforms into an axial dimension that is more or less identical with the axial dimension on which they are extreme, with high C being located where the raters' groups are extreme. If the latter scenario holds true, then, for example, artists should perceive progressives and conservatives as high and low on C, whereas groups that differ in A should not differ in C for artists. These and other empirical questions are interesting and important avenues for future research on the 2D ABC model of stereotype content about groups.

Conclusion

We presented a data- rather than theory-driven answer to the nature and order of the stereotype content dimensions that people spontaneously employ to distinguish between social groups. Our analyses indicate that people mentally organize groups primarily based on their stereotypic *agency / socio-economic success* (A), and secondarily based on their stereotypic *conservative-progressive beliefs* (B). Further, social groups that are thought to be average on A are inferred to be communal, whereas social groups that are thought to be extreme (high or low) on A are inferred to be as non-communal (C), resulting in a two-dimensional *ABC model* of stereotype content.

Appendix A

Detailed Property Fitting Results for Study 1

Candidate stereotype content dimension	*M*	*SD*	*R(1D axis)*	*R(2D axis)*	*R(3D axis)*	*r(1D pole)*	*r(2D pole)*	*r(3D pole)*
Conventional-Alternative	44.00	15.96	**.745**	**.874**	**.914**	**-.356**	-.332	-.294
Traditional-Modern	53.91	17.03	**.693**	**.886**	**.908**	.002	.067	.086
Conservative-Liberal	54.46	20.55	**.702**	**.853**	**.903**	-.178	-.084	-.058
Poor-Wealthy	45.80	17.02	**.625**	**.796**	**.885**	.226	.271	.245
Religious-Science-oriented	50.17	17.81	.301	**.695**	**.848**	.112	.118	.116
Low status-High status	53.11	15.54	**.627**	**.737**	**.820**	**.441**	**.538**	**.511**
Powerless-Powerful	46.95	17.56	**.606**	**.718**	**.812**	**.414**	**.514**	**.502**
Dominated-Dominating	50.70	18.34	**.555**	**.649**	**.796**	**.374**	**.420**	**.399**
Unintelligent-Smart	62.24	10.24	**.485**	**.672**	**.780**	.230	**.353**	.339
Unconfident-Confident	59.95	14.22	**.615**	**.660**	**.719**	**.364**	**.351**	.320
Unassertive-Competitive	57.35	14.48	.296	.369	**.701**	**.483**	**.385**	.334
Intolerant-Tolerant	57.45	10.79	.246	**.638**	**.686**	.128	.295	.317
Unable-Skillful	65.43	11.36	**.454**	**.586**	**.656**	**.481**	**.498**	**.474**
Incompetent-Competent	66.17	12.16	.327	**.503**	**.612**	**.461**	**.530**	**.511**
Masculine-Feminine	46.90	12.09	.096	**.491**	**.602**	-.214	.053	.042
Egoistic-Altruistic	46.47	12.00	.019	.121	**.593**	.013	.054	.073
Communal-Individualistic	48.66	11.09	**.374**	**.435**	**.559**	-.214	-.221	-.189
Untrustworthy-Trustworthy	55.10	12.26	.258	.308	**.491**	**.388**	**.385**	**.381**
Dishonest-Sincere	59.65	11.48	.177	.243	**.481**	.263	.250	.252
Threatening-Benevolent	61.98	11.20	.218	.369	**.451**	**.368**	**.438**	**.429**
Typical (U.S.)-Unusual (U.S.)	53.05	14.19	.181	.321	.397	**-.650**	**-.791**	**-.798**
Cold-Warm	57.75	11.84	.013	.146	.390	**.384**	**.455**	**.480**
Repellent-Likable	61.50	14.16	.243	.255	.386	**.493**	**.581**	**.594**
Unfriendly-Friendly	56.29	8.41	.173	.121	.327	.303	.341	.338

Note. Study 1: the 80 U.S. social groups' means (*M*) and standard deviations (*SD*) on the 24 candidate stereotype content dimensions. The higher *R(1D axis)*, *R(2D axis)*, *R(3D axis)*, *r(1D pole)*, *r(2D pole),* and *r(3D pole),* the more suitable the respective candidate stereotype content dimension for interpreting the social groups' 1D, 2D and 3D dissimilarity configurations in an axial and polar way, respectively. Bold correlation coefficients are significant at *p* =< .001.

Appendix B

Detailed Property Fitting Results for Study 2

Candidate stereotype content dimension	*R(1D axis)*	*R(2D axis)*	*R(3D axis)*	*r(1D pole)*	*r(2D pole)*	*r(3D pole)*
Traditional-Modern	**.633**	**.765**	**.927**	.046	.192	.142
Conservative-Liberal	**.807**	**.819**	**.916**	.079	.100	.038
Conventional-Alternative	**.820**	**.849**	**.905**	-.206	-.181	-.250
Low status-High status	**.758**	**.803**	**.878**	.164	.291	**.366**
Powerless-Powerful	**.745**	**.786**	**.877**	.067	.237	.298
Poor-Wealthy	**.656**	**.709**	**.865**	-.027	.036	.069
Dominated-Dominating	**.748**	**.776**	**.833**	-.026	.122	.178
Unconfident-Confident	**.734**	**.750**	**.825**	.168	.134	.160
Religious-Science-oriented	.268	**.618**	**.825**	-.073	.131	.105
Unable-Skillful	**.478**	**.507**	**.699**	**.352**	**.422**	**.472**
Unassertive-Competitive	**.508**	**.620**	**.655**	.059	.188	.214
Intolerant-Tolerant	.343	.352	**.643**	.343	**.441**	**.436**
Unintelligent-Smart	.341	.397	**.605**	.318	.315	.348
Communal-Individualistic	.331	**.542**	**.533**	**-.395**	-.252	-.248
Incompetent-Competent	.279	.297	**.521**	**.356**	**.451**	**.477**
Masculine-Feminine	.191	.195	**.445**	-.009	.037	.047
Typical (U.S.)-Untypical (U.S.)	**.389**	.354	.438	-.318	**-.676**	**-.724**
Egoistic-Altruistic	.158	.388	.373	.292	.284	.263
Threatening-Benevolent	.177	.169	.332	**.499**	**.545**	**.556**
Dishonest-Sincere	.070	.256	.310	**.520**	**.453**	**.433**
Repellent-Likable	.212	.224	.268	**.574**	**.655**	**.694**
Untrustworthy-Trustworthy	.151	.200	.261	**.584**	**.555**	**.553**
Unfriendly-Friendly	.159	.246	.235	**.407**	**.428**	**.423**
Cold-Warm	.088	.184	.211	**.563**	**.626**	**.618**

Note. Study 2: the higher *R(1D axis)*, *R(2D axis)*, *R(3D axis)*, *r(1D pole)*, *r(2D pole)*, and *r(3D pole)*, the more suitable the respective candidate stereotype content dimension for interpreting the 80 U.S. social groups' 1D, 2D and 3D dissimilarity configurations in an axial and polar way, respectively. Bold correlation coefficients are significant at p =< .001.

Appendix C

Detailed Property Fitting Results for Study 4

Candidate stereotype content dimension	*R(1D axis)*	*R(2D axis)*	*R(3D axis)*	*r(1D pole)*	*r(2D pole)*	*r(3D pole)*
Low status-High status	**.887**	**.924**	**.935**	.262	.263	.269
Poor-Wealthy	**.882**	**.891**	**.914**	.139	.090	.119
Powerless-Powerful	**.847**	**.887**	**.889**	.231	.230	.242
Conservative-Liberal	**.535**	**.841**	**.884**	.095	.096	.074
Traditional-Modern	.257	**.841**	**.876**	.088	.092	.092
Conventional-Alternative	**.562**	**.811**	**.874**	-.113	-.145	-.130
Dominated-Dominating	**.751**	**.808**	**.811**	.133	.127	.144
Unable-Skillful	**.721**	**.804**	**.826**	.314	.325	.337
Unconfident-Confident	**.717**	**.772**	**.771**	.263	.240	.241
Religious-Science-oriented	.169	**.734**	**.770**	-.040	-.038	-.022
Unintelligent-Smart	**.629**	**.727**	**.743**	.317	.303	.296
Incompetent-Competent	**.584**	**.720**	**.758**	.335	.341	.339
Intolerant-Tolerant	.047	**.676**	**.773**	**.367**	**.383**	**.369**
Unassertive-Competitive	**.549**	**.578**	**.579**	.192	.201	.209
Threatening-Benevolent	.352	**.534**	**.675**	**.509**	**.503**	**.473**
Repellent-Likable	.346	**.481**	**.716**	**.479**	**.525**	**.498**
Untrustworthy-Trustworthy	.304	.418	**.590**	**.510**	**.529**	**.507**
Individualistic-Communal	.117	.389	**.523**	**-.421**	**-.423**	**-.373**
Dishonest-Sincere	.208	.340	**.541**	**.368**	**.387**	**.363**
Masculine-Feminine	.004	.332	.375	.168	.126	.102
Typical (U.S.)-Untypical (U.S.)	.243	.313	**.533**	**-.362**	**-.447**	**-.427**
Cold-Warm	.016	.303	**.609**	**.509**	**.547**	**.519**
Egoistic-Altruistic	.107	.152	.395	.231	.241	.212
Unfriendly-Friendly	.004	.118	**.573**	**.532**	**.546**	**.502**

Note. Study 4: the higher *R(1D axis)*, *R(2D axis)*, *R(3D axis)*, *r(1D pole)*, *r(2D pole)*, and *r(3D pole)*, the more suitable the respective candidate stereotype content dimension for interpreting the 80 U.S. social groups' 1D, 2D and 3D dissimilarity configurations in an axial and polar way, respectively. Bold correlation coefficients are significant at p =< .001.

Appendix D

Property Fitting Results for the Additional Study Mentioned in Footnote 5

Candidate stereotype content dimension	*R(1D axis)*	*R(2D axis)*	*R(3D axis)*	*r(1D pole)*	*r(2D pole)*	*r(3D pole)*
Agency	**.885**	**.877**	**.889**	.151	.183	.189
Beliefs	.458	**.910**	**.932**	-.024	-.113	-.124
Communion	.125	.529	**.600**	.278	.328	.316

Note. *R(1D-3D axis)* indicate the maximal correlations between the 40 U.S. social groups' agency, beliefs, and communion ratings and their projections on an axis rotated around the origin of their 1D, 2D, and 3D dissimilarity spaces; *r(1D-3D pol)* indicate correlations between the social groups' agency, beliefs, and communion ratings and their proximity to the origin of these three spaces. Bold correlation coefficients are significant at p =< .001.

Appendix E

Detailed Property Fitting Results for the Additional Study Mentioned in Footnote 5

Candidate stereotype content dimension	*R(1D axis)*	*R(2D axis)*	*R(3D axis)*	*r(1D pole)*	*r(2D pole)*	*r(3D pole)*
Traditional-Modern	.456	**.909**	**.931**	.088	.092	.092
Poor-Wealthy	**.880**	**.905**	**.909**	.139	.090	.119
Conservative-Liberal	**.627**	**.899**	**.918**	.095	.096	.074
Powerless-Powerful	**.853**	**.851**	**.866**	.231	.230	.242
Low status-High status	**.870**	**.847**	**.857**	.262	.263	.269
Dominated-Dominating	.836	**.843**	**.856**	.133	.127	.144
Unintelligent-Smart	.480	**.830**	**.855**	.317	.303	.296
Religious-Science-oriented	.068	**.822**	**.901**	-.040	-.038	-.022
Conventional-Alternative	**.639**	**.817**	**.846**	-.113	-.145	-.130
Intolerant-Tolerant	.315	**.806**	**.825**	**.367**	**.383**	**.369**
Unable-Skillful	**.646**	**.776**	**.791**	.314	.325	.337
Incompetent-Competent	.412	**.733**	**.755**	.335	.341	.339
Unconfident-Confident	.671	**.658**	**.697**	.263	.240	.241
Threatening-Benevolent	.006	**.616**	**.613**	**.509**	**.503**	**.473**
Unassertive-Competitive	**.547**	**.579**	**.668**	.192	.201	.209
Dishonest-Sincere	.135	.535	**.613**	**.368**	**.387**	**.363**
Cold-Warm	.372	.513	.584	**.509**	**.547**	**.519**
Untrustworthy-Trustworthy	.045	.477	.521	**.510**	**.529**	**.507**
Masculine-Feminine	.244	.462	.486	.168	.126	.102
Egoistic-Altruistic	.206	.458	**.618**	.231	.241	.212
Repellent-Likable	.031	.423	.468	**.479**	**.525**	**.498**
Unfriendly-Friendly	.275	.339	.536	**.532**	**.546**	**.502**
Communal-Individualistic	.022	.254	**.677**	**-.421**	**-.423**	**-.373**
Typical (U.S.)-Untypical (U.S.)	.137	.190	.286	**-.362**	**-.447**	**-.427**

Note. The higher *R(1D axis)*, *R(2D axis)*, *R(3D axis)*, *r(1D pole)*, *r(2D pole)*, and *r(3D pole)*, the more suitable the respective candidate stereotype content dimension for interpreting the 40 U.S. social groups' 1D, 2D and 3D dissimilarity configurations in an axial and polar way, respectively. Bold correlation coefficients are significant at p =< .001.

Appendix F

Detailed Property Fitting Results for the 42 Minimalist U.S. Groups in Study 5

Candidate stereotype content dimension	*R(1D axis)*	*R(2D axis)*	*R(3D axis)*	*r(1D pole)*	*r(2D pole)*	*r(3D pole)*
Spatial arrangement based on global (dis)similarity						
Agency (A)	**.827**	**.928**	**.938**	-.095	-.076	-.073
Beliefs (B)	**.721**	**.895**	**.922**	.223	.187	.216
Communion (C; standard)	.021	.049	.291	**.516**	**.576**	**.550**
Communion (C; alternative)	.112	.182	.230	.439	**.487**	**.503**
Spatial arrangement based on character (dis)similarity						
Agency (A)	**.558**	**.794**	**.725**	-.414	-.118	-.114
Beliefs (B)	**.799**	**.889**	**.915**	-.059	.035	.041
Communion (C; standard)	.038	.394	.477	**.529**	**.564**	**.608**
Communion (C; alternative)	.135	.474	.543	**.523**	**.475**	**.545**
Spatial arrangement based on personal encounter (dis)similarity						
Agency (A)	**.742**	**.921**	**.894**	-.093	-.115	-.135
Beliefs (B)	**.786**	**.926**	**.930**	.232	.178	.200
Communion (C; standard)	.053	.088	.297	.470	**.563**	**.575**
Communion (C; alternative)	.187	.208	.188	.452	**.516**	**.560**

Note. *R(1D-3D axis)* indicate the maximal correlations between the 42 minimalist social groups' agency / socio-economic success, conservative-progressive beliefs, and communion (standard and alternative operationalization) ratings and their projections on an axis rotated around the origin of their 1D, 2D, and 3D space; *r(1D-3D pol)* indicate correlations between the 42 minimalist social groups' A, B, and C (standard and alternative operationalization) ratings and their proximity to the origin of these three spaces. Bold correlation coefficients are significant at p =< .001.

Appendix G

Detailed Property Fitting Results for the 61 Naturalistic U.S. Groups in Study 5

Candidate stereotype content dimension	*R(1D axis)*	*R(2D axis)*	*R(3D axis)*	*r(1D pole)*	*r(2D pole)*	*r(3D pole)*
Spatial arrangement based on global (dis)similarity						
Agency (A)	**.665**	**.757**	**.800**	-.024	.042	.078
Beliefs (B)	.334	**.875**	**.888**	**.532**	.183	.214
Communion (C; standard)	.195	.166	.234	**.405**	**.588**	**.555**
Communion (C; alternative)	.203	.167	.236	**.594**	**.705**	**.654**
Spatial arrangement based on character (dis)similarity						
Agency (A)	**.653**	**.763**	**.762**	-.061	.041	.024
Beliefs (B)	.294	**.794**	**.789**	**.398**	.126	.199
Communion (C; standard)	.126	.115	.411	**.405**	**.553**	**.483**
Communion (C; alternative)	.133	.170	**.529**	**.567**	**.665**	**.586**
Spatial arrangement based on personal encounter (dis)similarity						
Agency (A)	**.816**	**.849**	**.867**	.162	.180	.217
Beliefs (B)	.372	**.889**	**.858**	**.580**	.240	.272
Communion (C; standard)	.001	.095	**.612**	.353	**.517**	**.404**
Communion (C; alternative)	.019	.201	**.629**	**.557**	**.634**	**.527**

Note. *R(1D-3D axis)* indicate the maximal correlations between the 61 naturalistic social groups' agency / socio-economic success, conservative-progressive beliefs, and communion (standard and alternative operationalization) ratings and their projections on an axis rotated around the origin of their 1D, 2D, and 3D space; *r(1D-3D pol)* indicate correlations between the 61 naturalistic social groups' A, B, and C (standard and alternative operationalization) ratings and their proximity to the origin of these three spaces. Bold correlation coefficients are significant at p =< .001.

Appendix H

Study 6: Minimalist and Naturalistic U.S. Groups Most Frequently Selected for Spatial Arrangement

1st - 21th most frequently SELECTED minimalist groups	22st - 42th most frequently SELECTED minimalist groups	1st - 31th most frequently SELECTED naturalistic groups	32st - 61th Most frequently SELECTED naturalistic groups
Working class (70%)	Blue collar (51%)	Women (54%)	Hispanics (35%)
Women (69%)	Religious (51%)	Middle class (50%)	Asians (34%)
Middle class (67%)	Elderly (50%)	Educated (49%)	Politicians (34%)
Rich (66%)	Gays (49%)	Whites (48%)	Military (34%)
Men (65%)	Lesbians (49%)	Men (48%)	Lesbians (34%)
Poor (64%)	Parents (48%)	Students (46%)	Homeless (34%)
Democrats (64%)	White collar (48%)	Rich (46%)	Musicians (34%)
Whites (64%)	Homosexuals (47%)	Americans (45%)	Mexicans (33%)
Students (63%)	Asians (47%)	Democrats (45%)	Artists (32%)
Upper class (63%)	Hipsters (46%)	Religious (45%)	Catholics (32%)
Liberals (62%)	Hippies (44%)	Christians (44%)	Athletes (32%)
Lower class (60%)	Athletes (43%)	Young (43%)	Immigrants (32%)
Republicans (56%)	Politicians (43%)	Liberals (42%)	Elites (31%)
Teenagers (56%)	Immigrants (40%)	Poor (42%)	Entrepreneurs (30%)
Nerds (56%)	Drug addicts (40%)	Adults (41%)	Actors (30%)
Blacks (53%)	Hispanics (40%)	Blacks (41%)	Europeans (29%)
Christians (53%)	Transgender (40%)	Old (41%)	Activists (29%)
Conservatives (53%)	Muslims (38%)	Republicans (40%)	Farmers (28%)
Celebrities (52%)	Goths (36%)	Straight (40%)	Writers (27%)
Homeless (52%)	Jews (35%)	Conservatives (39%)	Muslims (27%)
Atheists (52%)	Jocks (33%)	Homosexuals (39%)	Law enforcement (27%)
		Employed (39%)	Jewish (26%)
		Children (39%)	Chinese (25%)
		Professionals (39%)	Ethnic (24%)
		Scientists (38%)	Rebels (23%)
		Minorities (38%)	Tall (23%)
		Friends (37%)	Short (23%)
		Family (37%)	Indians (22%)
		Teens (37%)	Clubs (14%)
		Parents (37%)	Alcoholics Anonymous
		Gays (37%)	

Note. Percentage in parentheses is proportion of participants who spontaneously selected this group for either global, character-, or personal encounter-based (dis)similarity arrangement.

Appendix I

Detailed Property Fitting Results for the 42 Minimalist U.S. Groups in Study 6

Candidate stereotype content dimension	*R(1D axis)*	*R(2D axis)*	*R(3D axis)*	*r(1D pole)*	*r(2D pole)*	*r(3D pole)*
Spatial arrangement based on global (dis)similarity						
Agency (A)	**.867**	**.864**	**.893**	-.131	-.017	-.004
Beliefs (B)	**.587**	**.805**	**.880**	.224	.211	.231
Communion (C; standard)	.043	.165	.414	**.580**	**.659**	**.631**
Communion (C; alternative)	.189	.246	.454	**.564**	**.610**	**.615**
Spatial arrangement based on character (dis)similarity						
Agency (A)	**.819**	**.915**	**.930**	-.139	-.084	-.049
Beliefs (B)	**.720**	**.944**	**.961**	.230	.185	.158
Communion (C; standard)	.026	.043	.054	**.602**	**.575**	**.593**
Communion (C; alternative)	.148	.170	.264	**.520**	**.522**	**.575**
Spatial arrangement based on personal encounter (dis)similarity						
Agency (A)	**.784**	**.864**	**.904**	-.236	-.168	-.177
Beliefs (B)	**.762**	**.897**	**.922**	.132	.230	.253
Communion (C; standard)	.074	.076	.325	**.600**	**.652**	**.639**
Communion (C; alternative)	.189	.237	.296	**.536**	**.566**	**.557**

Note. *R(1D-3D axis)* indicate the maximal correlations between the 42 minimalist social groups' agency / socio-economic success, conservative-progressive beliefs, and communion (standard and alternative operationalization) ratings and their projections on an axis rotated around the origin of their 1D, 2D, and 3D space; *r(1D-3D pol)* indicate correlations between the 42 minimalist social groups' A, B, and C (standard and alternative operationalization) ratings and their proximity to the origin of these three spaces. Bold correlation coefficients are significant at p =< .001.

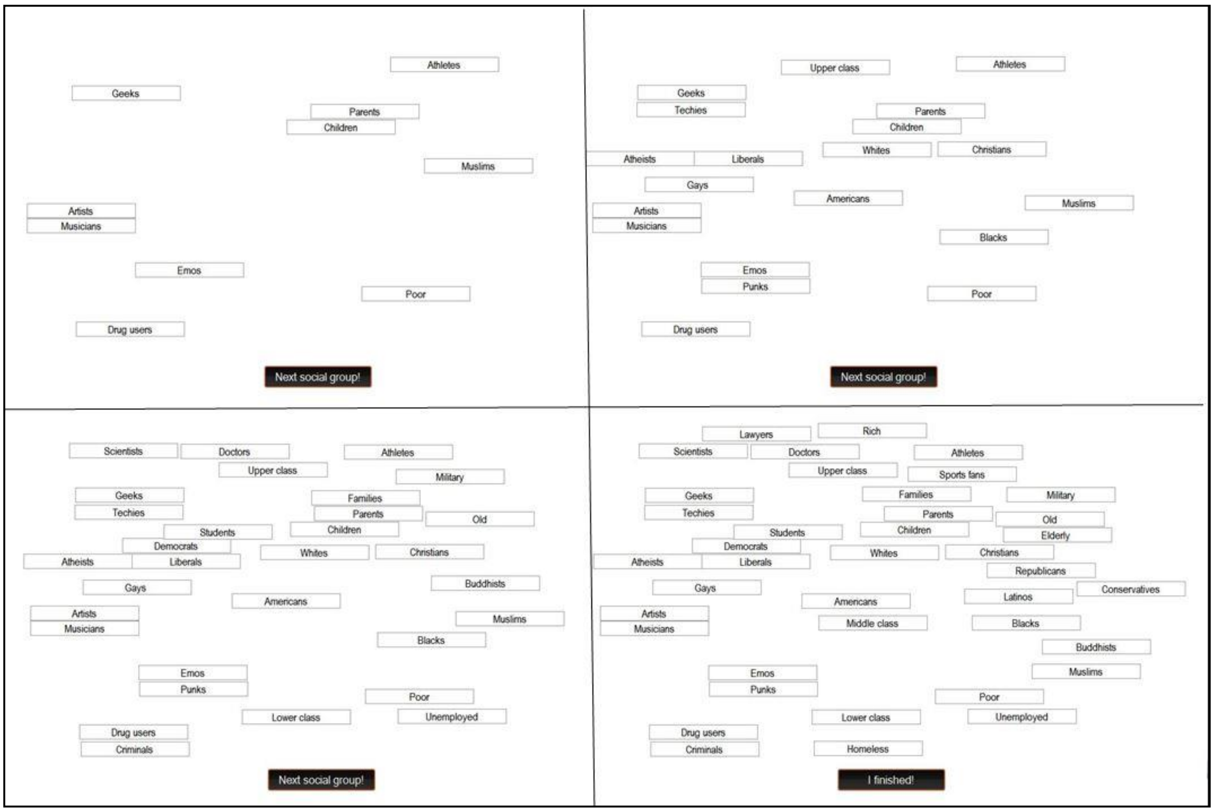
Appendix J

Detailed Property Fitting Results for the 61 Naturalistic U.S. Groups in Study 6

Candidate stereotype content dimension	*R(1D axis)*	*R(2D axis)*	*R(3D axis)*	*r(1D pole)*	*r(2D pole)*	*r(3D pole)*
Spatial arrangement based on global (dis)similarity						
Agency (A)	**.713**	**.733**	**.873**	-.002	.141	.146
Beliefs (B)	**.546**	**.864**	**.920**	.376	.220	.193
Communion (C; standard)	.030	.094	.264	**.422**	**.510**	**.451**
Communion (C; alternative)	.058	.221	.333	**.534**	**.585**	**.526**
Spatial arrangement based on character (dis)similarity						
Agency (A)	**.654**	**.700**	**.841**	-.046	.152	.100
Beliefs (B)	**.617**	**.877**	**.923**	.405	.343	.313
Communion (C; standard)	.045	.074	.335	.393	**.445**	**.444**
Communion (C; alternative)	.094	.258	.412	**.552**	**.516**	**.532**
Spatial arrangement based on personal encounter (dis)similarity						
Agency (A)	**.664**	**.754**	**.854**	-.069	-.036	-.013
Beliefs (B)	**.658**	**.699**	**.908**	.393	.365	.329
Communion (C; standard)	.074	**.461**	.419	**.414**	.358	**.362**
Communion (C; alternative)	.002	**.455**	.470	**.580**	**.447**	**.454**

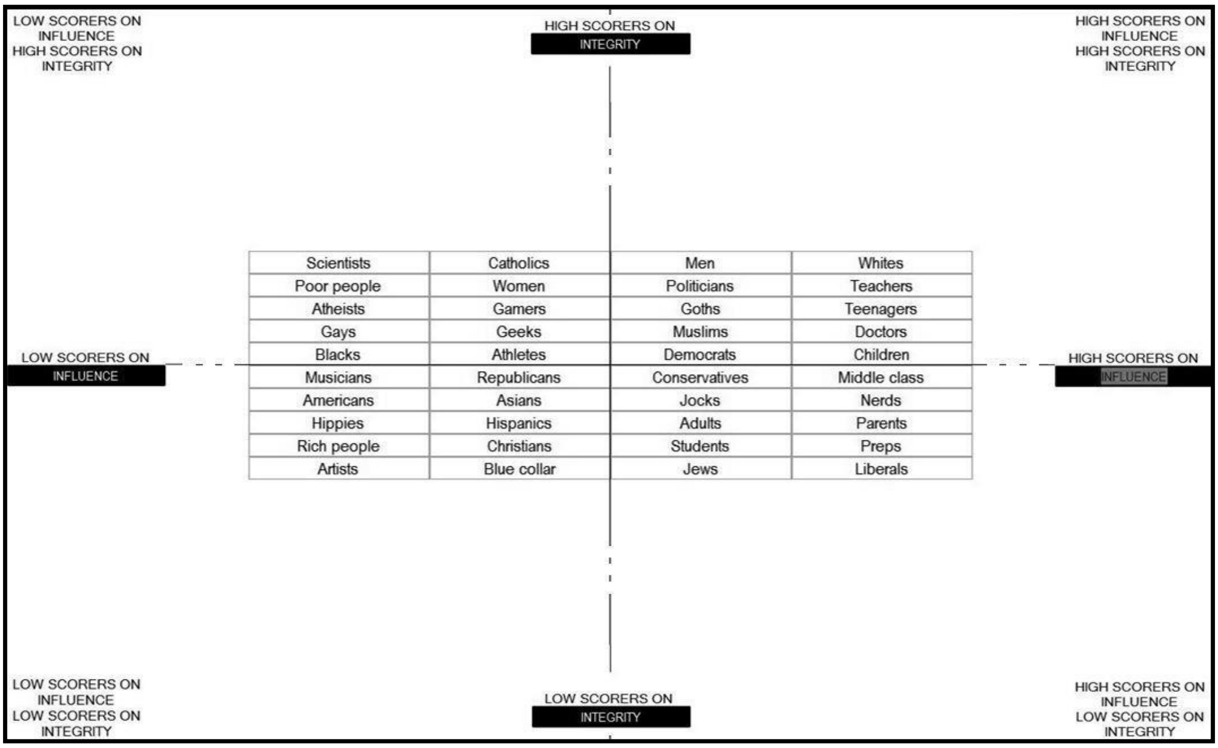
Note. *R(1D-3D axis)* indicate the maximal correlations between the 61 naturalistic social groups' agency / socio-economic success, conservative-progressive beliefs, and communion (standard and alternative operationalization) ratings and their projections on an axis rotated around the origin of their 1D, 2D, and 3D space; *r(1D-3D pol)* indicate correlations between the 61 naturalistic social groups' A, B, and C (standard and alternative operationalization) ratings and their proximity to the origin of these three spaces. Bold correlation coefficients are significant at *p* =< .001.

Appendix K



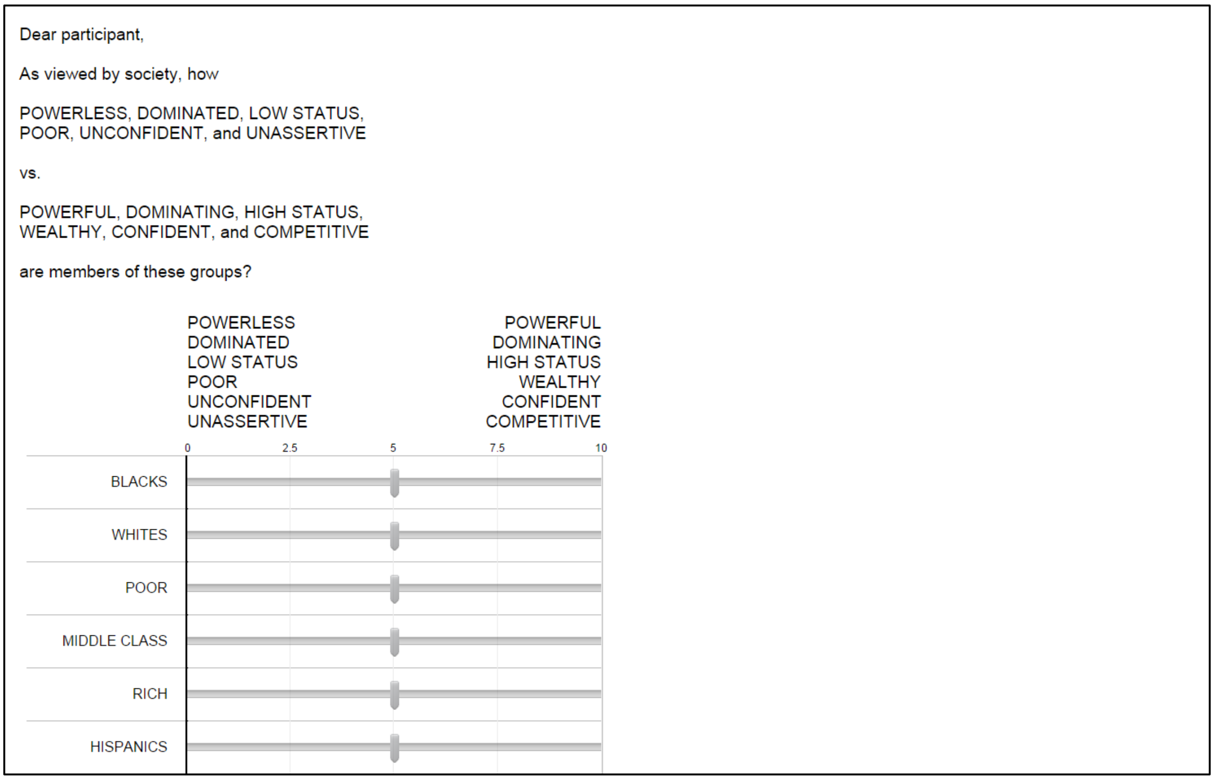
Note. One out of many ways to spatially arrange these 40 representative social groups (first 1, then 2, 3 … 10 … 20 … 30 … 40). Colors are inverted.

Appendix L



Note. In Studies 4 and 5, if a participant came to think that 'influence' and 'integrity' best describe the similarities and differences between these 40 representative social groups, then this is one of many possible setups to begin with the 2D rating task. Colors are inverted.

Appendix M


Dear participant,

As viewed by society, how

POWERLESS, DOMINATED, LOW STATUS, POOR, UNCONFIDENT, and UNASSERTIVE

vs.

POWERFUL, DOMINATING, HIGH STATUS, WEALTHY, CONFIDENT, and COMPETITIVE

are members of these groups?

Note. An example of the meaning clouds that participants in Studies 5 and 6 used to rate the groups' agency / socio-economic success, conservative-progressive beliefs, and communion. This particular meaning cloud was used to measure the groups' agency / socio-economic success.

Appendix N

Your task is to identify X.

X is the person characteristic based on which the social groups are ranked.

As viewed by society, groups at the top of the ranking are extremely X.

As viewed by society, groups above the center of the ranking are above-averagely X.

As viewed by society, groups at the center of the ranking are averagely X.

As viewed by society, groups below the center of the ranking are below-averagely X.

As viewed by society, groups at the bottom of the ranking are not at all X.

Republicans
Conservatives
Rich
Upper class
Politicians
Christians
Religious

⋮

Gays
Homosexuals
Transgender
Drug addicts

Please enter the person characteristic X in the textbox below.

If you have no idea about X, then enter "I don't know".

Note. An example of the social group rankings that participants labeled in Study 5. The label most often generated for this ranking was "conservative".

Chapter 4 – Discussion

In Chapter 1 I argued that valence asymmetry in stimulus similarity can be taken as a novel, independent, powerful, appealing, robust, and general explanation of valence asymmetry in cognitive processing. Chapter 2 and 3 elaborated on the empirical robustness and generality of the theoretically predicted higher similarity of positive compared to negative stimuli, showing the effect for thousands of words and pictures (Chapter 2), and for hundreds of groups (Chapter 3). In Chapter 4 I argue that our explanation of valence asymmetry in stimulus similarity – that is, the homeostatic distribution (i.e., too little-adequate-too much, see Chapter 1) of quantities on evaluatively relevant dimensions – predicts a social group stereotypes model that is different from, and possibly more complex than, the classical model developed by Fiske and colleagues (2002).

Groups' stereotypic communion and agency are not orthogonal, but related in a curvilinear manner such that communion peaks at average agency

Two dimensions have so often been postulated as the two most important dimensions of content in social perception and cognition that they have become known as the Big Two. These Big Two are communion and agency, and there are many synonyms such as femininity and masculinity, expressiveness and instrumentality, morality and competence, trustworthiness and autonomy etc. (for a review, see Abele & Wojciszke, 2014). Fiske and colleagues (2002) labeled the Big Two in the group perception and cognition literature, namely as warmth (tolerant, warm, good-natured, sincere) and competence (competent, confident, independent, competitive, intelligent). According to their model, it is crucial for to survive and thrive (i.e., overcoming challenges and taking chances) that people get a speedy and accurate impression of whether another person's intentions towards them are benign or malicious. Once another person's intentions are known, the next most important thing to know is whether that person is able or unable to achieve what he / she intends to do to the self. Fiske and colleagues (2002) argue that a person's group membership is to a great extent informative about his / her intentions towards the self, and about his / her ability to achieve what he / she intends to do to the self. Thus people stereotype groups on the Big Two. As groups with benign or malicious intentions (high vs. low warmth) can in principle be able or unable to achieve what they intend to do to the self (high vs. low competence), Fiske and colleagues (2002) claimed that warmth and competence are orthogonal dimensions of group stereotypes. There is some evidence in support of this claim (Cuddy et al., 2009; Kervyn, Fiske, & Yzerbyt, 2013), but the authors of the respective studies did not investigate whether groups' warmth / communion and competence / agency are related in a curvilinear manner.

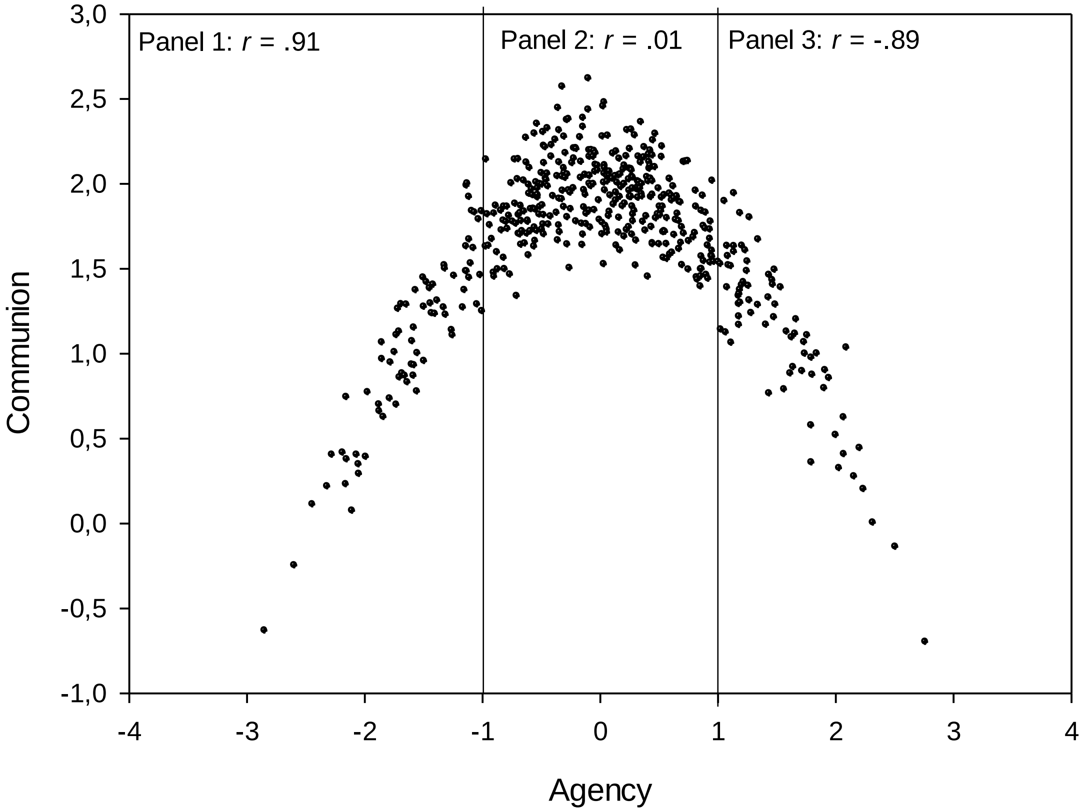
Consistent with the homeostatic valence model developed in Chapter 1 I propose that groups' stereotypic warmth / communion and competence / agency are not orthogonal, but related in a curvilinear manner such that communion peaks at average agency. Specifically, the homeostatic valence model distinguishes between evaluatively relevant dimensions and evaluation as a dimension that follows from one or more evaluatively relevant dimensions; positive and negative evaluations follow from intermediate (i.e., adequate) and extreme (i.e., insufficient and excessive) quantities on evaluatively relevant dimensions. I propose that groups' warmth / communion is identical to their evaluation. The extraordinarily high correlation between groups' warmth / communion and their evaluation observed in Study 1 in Chapter 3 (see Footnote 16 in the General Discussion in Chapter 3) supports this claim. Further, I propose that groups' competence / agency is not identical to their evaluation, but rather an evaluatively relevant dimension in the sense of the homeostatic valence model. If this is correct, then the correlations between groups' competence / agency and their evaluation, and between groups' competence / agency and their warmth / communion (i.e., the linear relation between groups' competence / agency and warmth / communion) should be low or zero, whereas the correlations between groups' squared centered competence / agency and their evaluation, and between groups' squared centered competence / agency and their warmth / communion (i.e., the inverted u-shaped relation between groups' competence / agency and warmth / communion) should be higher than low or zero. We found this pattern of correlations for the groups examined in Studies 1, 3, and 5 in Chapter 3, suggesting that groups' warmth / communion and competence / agency are indeed related in a curvilinear manner such that communion peaks at average agency.

Testing the robustness and generality of this curvilinear relation, we analyzed some more data that are not reported in Chapter 3. Specifically, we (Imhoff & Koch, 2016) compared the correlation between agency and communion with the correlation between squared centered agency and communion for groups generated by a large participant sample that is age- and gender-representative of the German population, for groups generated by 30+ large participant samples from 20+ countries (Durante et al., 2013), and for groups rated on agency and communion by single individuals rather than several participants (Imhoff & Bruder, 2014). For all these samples of groups, the linear relation between agency and communion was low to zero, whereas the inverted u-shaped curvilinear relation between the Big Two was higher than low to zero. These results suggest that our finding in Chapter 3 that groups' stereotypic agency and communion are related in a curvilinear manner (i.e., communion peaks at average agency) is robust and general. This conclusion speaks against the group stereotype model developed by Fiske and colleagues (2002; see also Cuddy et al., 2007) in which agency and communion are orthogonal dimensions, and at the same time speaks for the validity of the homeostatic valence model developed in Chapter 1. Further empirical support for the validity of the homeostatic valence model is provided by inverted u-shaped curvilinear relations between a variety of evaluatively relevant dimensions and evaluation, and by a lack of u-shaped curvilinear relations between evaluatively relevant dimensions and evaluation (see Chapter 1 or Grant & Schwartz, 2011).

If groups' stereotypic communion peaks at average agency, how come that the correlation between entities' agency and communion has been observed to be negative (e.g., Judd, James-Hawkins, Yzerbyt, & Kashima, 2005; Kervyn, Bergsieker, & Fiske, 2012; Wojciszke, 1994), null (Abele, 2003; Cislak & Wojciszke, 2008; Cuddy et al., 2009; Kervyn, Fiske, & Yzerbyt, 2013), and positive (Abele & Wojciszke, 2007; Suitner & Maas, 2008; Wojciszke & Abele, 2008; Durante et al., 2013)? Can the homeostatic valence model explain this inconsistent pattern of results? I propose that yes, but only yes if one additional assumption is correct, namely the assumption that previous studies on the nature of the relationship between entities' agency and communion have examined constrained samples of entities that are not representative of the entire agency spectrum.

As illustrated in Figure 1, given the inverted u-shaped curvilinear relation between agency and communion predicted by the homeostatic valence model (see Chapter 1), a positive (Figure 1, panel 1), null (panel 2), and negative (panel 3) linear relation is observed for entities low, average, and high on agency, respectively. In other words, if the relation between the Big Two is actually curvilinear in the sense of the homeostatic valence model, constrained stimulus sampling (Fiedler, 2000; 2011; 2014) obscures this curvilinear relation and might explain contradictory findings regarding the linear relation between entities' agency and communion.

Figure 1



Note. Exemplary inverted u-shaped curvilinear relation between agency and communion.

Finally, overlooking the inverted u-shaped curvilinear relation between entities' agency and communion predicted by the homeostatic valence model might happen even in studies that examine samples of entities that are representative of the entire agency spectrum. If authors of such studies fail to test for an inverted u-shaped curvilinear relation, they cannot find evidence in support of it. In fact, apart from our own studies to my knowledge there is no study that tested for an inverted u-shaped curvilinear relation between groups' agency and communion.

The ABC model of group stereotypes developed in Chapter 3 differs from the classical model by Fiske and colleagues (2002) in two ways. First, as discussed in the previous section, in the ABC model groups' stereotypic communion (C) and agency (A) are related in a curvilinear manner (see Figure 1) rather than orthogonal, a phenomenon that necessarily follows from the homeostatic valence model developed in Chapter 1. And second, a phenomenon that is well in line with the homeostatic valence model in which the evaluation of a class of stimuli can depend on more than just one evaluatively relevant dimension, the ABC model contains an additional dimension, namely groups' stereotypic conservative-progressive beliefs (B). How come that the classical model misses B? I propose that the reason is the same reason why it misses the curvilinear relation between C and A, namely the constrained sampling strategy (Fiedler, 2000; 2011; 2014) of its proponents. Specifically, the curvilinear relation between groups' C and A has been overlooked by proponents of the classical model because they did not test for an inverted u-shaped curvilinear relation between groups' agency and communion. Likewise, B has been overlooked because participants in the respective studies never rated groups on dimensions other than those relating to agency and communion. That is, as groups were never rated on B, surely B could not become part of the classical model of group stereotypes.

In the studies that reported in Chapter 3, participants could freely choose both the groups they wanted to stereotype, and the dimensions on which they wanted to stereotype the groups. Specifically, participants provided similarity ratings for freely chosen groups, which allowed them to freely choose stereotype dimensions. For example, the similarity of politicians and prostitutes can be rated on the stereotype dimensions A (politicians > prostitutes), C (politicians ~ prostitutes), attractiveness (prostitutes > politicians) etc. – it is totally up to participants which dimensions they spontaneously use. Multidimensional scaling, principal components, and property fitting, and axis labeling / categorization analyses (see Studies 1-6 in Chapter 3) revealed that participants consistently used not only A, but also B. Why do people stereotype groups based on their conservative-progressive beliefs?

Groups' are stereotyped based on their conservative-progressive beliefs to manage trade-offs between social exploitation and exploration

Managing the fundamental trade-off between exploiting known resources and exploring novel resources that are potentially better but risky is key to self-regulation (Inzlicht, Schmeichel, & Macrae, 2014). Adults, children, other primates and many other beings have always had to choose between available and alternative habitats, shelters, occupations, foods, mates, and these choices have always been crucial to survive and thrive. The ABC model argues that in today's society it is the conservative and progressive groups that provide access to known, available and novel, alternative resources, respectively.

Therefore, to successfully manage exploitation versus exploration trade-offs, today's citizens mentally represent groups along the stereotype dimension conservative-progressive beliefs.

Chapter 3 provides indirect correlational evidence for this claim. As reported in the General Discussion of Chapter 3, additional participants rated the extent to which the groups examined in Studies 1-4 stand for six exploitation versus exploration values, namely familiarity-novelty, safety-freedom, comfort-stimulation, loyalty-autonomy, preservation-change, and uniformity-diversity. If people stereotype groups on B to manage exploitation versus exploration trade-offs, these six exploitation versus exploration values should correlate with the groups' conservative-progressive beliefs, but not with their A or C. This was the case. In fact, the average correlation between the six exploitation versus exploration values and groups' B was as strong as the average correlation between the four facets of groups' B (traditional-modern, religious-science-oriented, conventional-alternative, and conservative-liberal, see Table 3 in Study 1 in Chapter 3).

A set of four unreported studies provides direct experimental evidence that people stereotype groups on B to manage exploitation versus exploration trade-offs. If so, then people should choose to cooperate with conservatives (progressives) in economic games in which they learn that exploitation (exploration) is the strategy that pays off best. We tested this prediction using two economic exploitation versus exploration games, namely the Balloon Analogue Risk Task (BART; Lejuez et al., 2002), and a modified version of the Iowa Gambling Task (IGT; Bechara, Damasio, Damasio, & Anderson, 1994).

Playing the BART (Lejuez et al., 2002), participants earned a point for each inflation of a digital balloon. Participants were told that the balloon can in principle be inflated up to 30 times. However, they were also told that the balloon bursts at a certain inflation step, and if it bursts, all points earned so far are lost. After each successful inflation, participants could decide to end the game and walk away with the points earned so far (i.e., exploiting resources), or to inflate some more (i.e., exploring potentially better but risky resources). There were two conditions. For half of the participants, the balloon burst early (inflation 11-13), and thus they learned that exploitation pays of better. For the other half, the balloon burst late (inflation 23-25), and thus they learned that exploration pays off better. Next, participants were presented with 8 groups high on A (e.g., rich people), 8 groups low on A (e.g., alcoholics), 8 conservative groups (e.g., religious people), and 8 progressive groups (e.g., environmentalists), and their task was to first rate the extent to which members of these groups would inflate the balloon, and to then indicate the extent to which they would be willing to delegate playing the

BART in their stead to members of the target groups. As predicted, participants indicated that members of conservative groups would inflate the balloon less than members of progressive groups. More importantly, as predicted, participants who learned that inflating the balloon less (i.e., exploitation) pays off better because the balloon bursts early preferred to delegate playing the BART to members of conservative groups, whereas participants who learned that inflating the balloon more (i.e., exploration) pays off better because the balloon bursts late preferred to delegate to members of progressive groups. The size of this interaction effect was large. We reasoned that its size was large because we forced participants to ponder about the extent to which members of the target groups would inflate the balloon, a thought process that they normally may not engage in. We repeated the experiment without the phase in which participants rated the extent to which members of the target groups would inflate the balloon, but the interaction effect remained significant and large (not as large as before, but still large). In sum, these two experiments suggest that people stereotype groups based on their conservative-progressive beliefs to manage exploitation versus exploration trade-offs.

Playing the modified version of the IGT (Bechara et al., 1994), participants first drew ten cards from a standard deck, and then drew 40 more cards from decks of their free choice, namely the standard deck or decks option 1-4. The cards in these decks were either wins or blanks. The five decks had different winning probability. In the exploitation condition, the standard deck had the highest winning probability ($p = .8$; the winning probabilities of decks option 1-4 were $p = .5$, .5, .2, and .2). Thus, as participants' task was to draw as many wins as possible, they learned that sticking with the standard deck (i.e., exploiting resources) pays of best. In the exploration condition, one of decks option 1-4 had the highest winning probability ($p = .8$; the winning probability of the standard deck was $p = .5$, and the winning probabilities of the remaining option decks were $p = .5$, .2, and .2). Therefore, participants learned that trying out decks option 1-4 (i.e., exploring potentially better but risky resources) pays off best. Next, participants rated the extent to which members of the same 32 target groups as in the BART (Lejuez et al., 2002) studies outlined above would stick with the standard deck versus try out decks option 1-4, and to then indicate the extent to which they would be willing to delegate playing the IGT in their stead to members of the target groups. As predicted, participants indicated that conservatives would stick with the standard deck more than progressives. More importantly, mirroring the main result of the two BART studies, in the exploitation condition in which sticking with the standard deck paid off better participants preferred to delegate to conservatives, whereas in the exploration condition in which trying out decks option 1-4 paid of better participants preferred to delegate to progressives. The size of this interaction effect was again large, and smaller but still medium when leaving out the phase in which participants rated the extent to which members of the 32 target groups would stick with the standard deck versus try out decks option 1-4. In sum, these two experiments increased our confidence that people stereotype groups based on their conservative-progressive beliefs to manage exploitation versus exploration trade-offs.

Additional experiments that we are about to conduct will deal with the research question whether there are actual differences in exploitation-exploration behavior between members of conservative and progressive groups – that is, is there a kernel of truth to people's stereotype that conservatives and progressives tend to exploit and explore, respectively? Also, we will try to replicate preference for conservatives and progressives in economic games that call for exploitation and exploration, respectively, when the stakes are high – that is, when people's delegation decisions earn them less or more real money.

I summarize Chapter 4 as follows: The homeostatic valence model developed in Chapter 1 predicts a group stereotype model that differs from the classical model developed by Fiske and colleagues (2002). According to the homeostatic valence model, the relation between the stereotype dimensions agency (A) and communion (C) should not be orthogonal as in the classical model, but inverted u-shaped curvilinear such that communion peaks at average agency. In the ABC model of group stereotypes developed in Chapter 3, agency and communion are related in this curvilinear manner supported by a variety of studies presented in Chapters 3 and 4. Also consistent with the homeostatic valence model, the ABC model contains a new stereotype dimensions that does not appear in the classical model, namely groups' conservative-progressive beliefs (B). I proposed that the classical model misses both B and the curvilinear relation between A and C because of the constrained sampling strategy (Fiedler, 2000; 2011; 2014) of its proponents. Chapters 3 and 4 provide evidence for this explanation. Finally, I proposed and presented both correlational and experimental evidence suggesting that people stereotype groups on B to manage trade-offs between exploiting known resources and exploring potentially better but risky novel resources.

Chapter 5 – Conclusion

In my thesis, I made the following points. Valence asymmetry in similarity – that is, higher similarity of positive compared to negative objects, people, and events (i.e., stimuli) – is an explanation of valence asymmetry in cognitive processing that is novel and independent of the classical explanation in terms of valence asymmetry in affective-motivational potential. Further, valence asymmetry in similarity is an ecological, more distal and thus more powerful explanation of valence asymmetry in cognitive processing than the intrapsychic phenomenon valence asymmetry in affective-motivational potential. Also, valence asymmetry in similarity is a general explanation of valence asymmetry in cognitive processing, as higher similarity of positive compared to negative stimuli holds true for a variety of emotions, faces, persons, groups, and, more generally, words and pictures. Valence asymmetry in similarity follows from what I refer to as the homeostatic valence model. The basic idea of this model is that on most if not all evaluatively relevant dimensions, positive, adequate quantities are flanked by two ranges of negative quantities, namely too little and too much. Given that social groups' agency and communion can be taken as an evaluatively relevant dimension and evaluation per se, respectively, this central property of the homeostatic valence model predicts a group stereotypes model in which agency (A) and communion (C) are not orthogonal, as postulated in the classical group stereotypes model, but related in an inverted u-shaped curvilinear manner. Correcting for sampling biases in the studies that support the classical group stereotypes model, we developed an updated ABC model in which A and C are related in this curvilinear model. Consistent with the homeostatic valence model, the ABC model contains a stereotype dimensions that does not appear in the classical model, namely groups' conservative-progressive beliefs (B). People might stereotype groups on B to manage trade-offs between exploiting known resources and exploring potentially better but risky novel resources.

References

Aaker, J., Garbinsky, E. N., & Vohs, K. D. (2012). Cultivating admiration in brands: Warmth, competence, and landing in the Golden quadrant. *Journal of Consumer Psychology, 22,* 191-194.

Abele, A. E. (2003). The Dynamics of Masculine-Agentic and Feminine-Communal Traits. Findings from a Prospective Study. *Journal of Personality and Social Psychology, 85,* 768-776.

Abele, A. E., & Bruckmüller, S. (2011). The bigger one of the “Big Two”? Preferential processing of communal information. *Journal of Experimental Social Psychology, 47,* 935-948.

Abele, A. E. & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology, 93*, 751-763.

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: a dual perspective model. *Advances in Experimental Social Psychology, 50*, 195-255.

Agerström, J., Björklund, F., Carlsson, R., Rooth, D. (2012). Warm and competent Hassan = Cold and incompetent Eric: The harsh equation of real-life hiring discrimination. *Basic and Applied Social Psychology, 34*, 359-366.

Akpinar, E., & Berger, J. (2015). Drivers of cultural success: The case of sensory metaphors. *Journal of Personality and Social Psychology, 109,* 20-34.

Alves, H., Koch, A., Krüger, T., & Unkelbach, C. (2016) My friends are all alike – on the relation between liking, knowledge and perceived similarity. *Journal of Experimental Social Psychology, 61,* 103-117.

Alves, H., Unkelbach, C., Burghardt, J., Koch, A. S., Krüger, T., & Becker, V. D. (2015). A density explanation of valence asymmetries in recognition memory. *Memory & Cognition*, *43*, 896-909.

Ames, D. R. (2004). Strategies for social inference: a similarity contingency model of projection and stereotyping in attribute prevalence estimates. *Journal of Personality and Social Psychology*, *87*, 573-585.

Ames, D. L., & Fiske, S. T. (2015). Perceived intent motivates people to magnify observed harm. *Proceedings of the National Academy of Sciences, 112,* 3599-3605.

Ames, D. L., & Fiske, S. T. (2013). Intentional harms are worse, even when they're not. *Psychological Science, 24*, 1755-1762.

Ames, D.R., & Flynn, F.J. (2007). What breaks a leader: The curvilinear relation between assertiveness and leadership. *Journal of Personality and Social Psychology, 92*, 307–324.

Anderson, N. H. (1965). Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Personality and Social Psychology, 2,* 1-9.

Anderson, C., Hildreth, J. A. D., & Howland, L. (2015). Is the desire for status a fundamental human motive? A review of the empirical literature. *Psychological Bulletin, 141*, 574-601

Ashton-James, C. E., Richardson, D. C., Williams, A. C. D., Bianchi-Berthouze, N., & Dekker, P. H. (2014). Impact of pain behaviors on evaluations of warmth and competence. *Pain, 155*, 2656-2661.

Augustine, A. A., Mehl, M. R., & Larsen, R. J. (2011). A positivity bias in written and spoken English and its moderation by personality and gender. *Social Psychological and Personality Science, 2*, 508-515.

Balliet, D., Mulder, L. B., & Van Lange, P. A. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin, 137*, 594–615.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*, 230-244.

Barry, B., & Stewart, G.L. (1997). Composition, process, and performance in self-managed groups: The role of personality. *Journal of Applied Psychology, 82*, 62–78.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5,* 323-370.

Baumeister, R.F., Campbell, J.D., Krueger, J.I., & Vohs, K.D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest, 4*, 1–44.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*, 7-15.

Becker, D. V., Anderson, U. S., Mortensen, C. R., Neufeld, S. L., & Neel, R. (2011). The face in the crowd effect unconfounded: happy faces, not angry faces, are more efficiently detected in single-and multiple-target visual search tasks. *Journal of Experimental Psychology: General, 140,* 637-659.

Becker, J.C., & Asbrock, F. (2012). What triggers helping versus harming of ambivalent groups? Effects of the relative salience of warmth versus competence. *Journal of Experimental Social Psychology, 48*, 19-27.

Berdahl, J. L., & Martorana, P. (2006). Effects of power on emotion and expression during a controversial group discussion. *European Journal of Social Psychology, 36*, 497-509.

Bergsieker, H. B., Leslie, L. M., Constantine, V. S., & Fiske, S. T. (2012). Stereotyping by omission: eliminate the negative, accentuate the positive. *Journal of Personality and Social Psychology*, *102*, 1214-1238.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political Analysis, 20*, 351-368.

Berman, S.L., Down, J., & Hill, C.W.L. (2002). Tacit knowledge as a source of competitive advantage in the National Basketball Association. *Academy of Management Journal*, *45*, 13–31.

Berman, M. G., Hout, M. C., Kardan, O., Hunter, M., Yourganov, G., Henderson, J. M., Hanayik, T., Karimi, H., & Jonides, J. (2014). The perception of naturalness converges with objective measures of naturalness. *PLoS One, 9,* e114572.

Bernard, C. (1974). *Lectures on the phenomena of life common to animals and plants* (Vol. 1, p. 84). *(Original work published 1865)* (E. Hebbel, R. Hoff, & L. Guillermin Trans.). Springfield, Illinois: Charles C Thomas.

Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology, 3,* 339-344.

Bless, H., & Fiedler, K. (2006). Mood and the regulation of information processing and behavior. In J. P. Forgas (Ed.), *Hearts and minds: affective influences on social cognition and behaviour* (pp. 65-84). New York: Psychology Press.

Borg, I., & Groenen, P. (2005). *Modern multidimensional scaling: Theory and applications.* New York, NY: Springer-Verlag.

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: the primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology, 41*, 135-143.

Brand, A., & Bradley, M. T. (2012). More voodoo correlations: when average-based measures inflate correlations. *The Journal of General Psychology, 139*, 260-272.

Brand, A., Bradley, M. T., Best, L. A., & Stoica, G. (2008). Accuracy of effect size estimates from published psychological research. *Perceptual and Motor Skills, 106*, 645-649.

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

Brewer, M. D. (2015). The rise of partisanship and the expansion of partisan conflict within the American electorate. *Political Research Quarterly, 58*, 219-229.

Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1972). Sex-Role Stereotypes: A Current Appraisal. *Journal of Social Issues, 28*, 59-78.

Brown, J.D., & Marshall, M.A. (2001). Great expectations: Optimism and pessimism in achievement settings. In E.C. Chang (Ed.), *Optimism & pessimism: Implications for theory, research, and practice* (pp. 239–255). Washington, DC: American Psychological Association.

Bruckmüller, S., & Abele, A. E. (2013). The density of the Big Two: How are agency and communion structurally represented? *Social Psychology, 44*, 63-74.

Bruner, J.S. (1957). Going beyond the information given. In J. S. Bruner, E, Brunswik, L. Festinger, F. Heider, K. F. Muenzinger, C. E. Osgood, & D. Rapaport, (Eds.), *Contemporary approaches to cogniti*on (pp. 41-69). Cambridge, MA: Harvard University Press.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193-217.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data?. *Perspectives on Psychological Science, 6*, 3-5.

Byrne, D. (1971). *The attraction paradigm.* New York: Academic Press.

Cambon, L., Yzerbyt, V., & Yakimova, S. (2015). Compensation in intergroup relations: An investigation of its structural and strategic foundations. *British Journal of Social Psychology, 54*, 140-158.

Cannon, W.B. (1926). Physiological regulation of normal states. Some tentative postulates concerning biological homeostasis. In W.B. Cannon & Pettit, A. (Eds.), *À Charles Richet, ses amis, ses collègues, ses élèves* (pp. 91-93). Paris: Les Editions Médicales.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior, 29*, 2156-2160.

Carrier, A., Louvet, E., Chauvin, B., & Rohmer, O. (2014). The primacy of agency over competence in status perception. *Social Psychology, 45*, 347-356.

Carroll, J. D., & Wish, M. (1974). Multidimensional perceptual models and measurement methods. *Handbook of Perception, 2,* 391-447.

Chang, J. J., & Carroll, J. D. (1969). *How to use PROFIT, a computer program for property fitting by optimizing nonlinear or linear correlation*. Murray Hill, N J: Bell Telephone Laboratories.

Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A., & Henrich, J. (2013). Two ways to the top: Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. Journal of Personality and Social Psychology, 104, 103-125.

Cikara, M., & Fiske, S. T. (2012). Stereotypes and Schadenfreude: Behavioral and physiological markers of pleasure at others' misfortunes. *Social Psychological and Personality Science, 3*, 63-71.

Cilibrasi, R. L., & Vitanyi, P. M. B. (2007). The Google similarity distance. *IEEE Transactions on Knowledge and Data Engineering, 19,* 370-383.

Cislak, A. & Wojciszke, B. (2008). Agency and communion are inferred from actions serving interests of self or others. *European Journal of Social Psychology, 38,* 1103-1110.

Clow, K. A., & Leach, A. M. (2015). Stigma and wrongful conviction: all exonerees are not perceived equal. *Psychology, Crime & Law, 21*, 172-185.

Cohen, J. D., McClure, S. M., & Angela, J. Y. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*, 933-942.

Collisson, B., & Howell, J. L. (2014). The liking-similarity effect: perceptions of similarity as a function of liking. *The Journal of Social Psychology, 154,* 384-400.

Csikszentmihalyi, M. (1990). Flow: *The psychology of optimal performance*. NY: Cambridge University Press.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2007). The BIAS map: behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology, 92*, 631-648.

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

Davies, M. (2011). *Google Books (American English) Corpus (155 billion words, 1810-2009).* URL: http://googlebooks.byu.edu.

Deaux, K., & Lewis, L. L. (1984). Structure of gender stereotypes: Interrelationships among components and gender label. *Journal of Personality and Social Psychology, 46*, 991-1004.

De Deyne, S., Peirsman, Y., & Storms, G. (2009). Sources of semantic similarity. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society* (p. 1834-1839). Austin, TX: Cognitive Science Society.

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology, 24*, 252-287.

Denrell, J. (2005). Why most people disapprove of me: experience sampling in impression formation. *Psychological Review, 112,* 951-978.

DePaulo, B. M., & Friedman, H. S. (1998). Nonverbal communication. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 3-40). Boston, MA: McGraw-Hill.

DeSoto, K. A., & Roediger, H. L. (2014). Positive and Negative Correlations Between Confidence and Accuracy for the Same Events in Recognition of Categorized Lists. *Psychological Science*, *25*, 781-788.

Dhami, M. K., Hertwig, R., Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, *130*, 959-988.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*, 285-290.

Durante, F., Fiske, S., Kervyn, N., Cuddy, A. J. C., Akande, A., Adetoun, B. E., …, Storaric, C. C.. (2013). Nations' income inequality predicts ambivalence in stereotype content: How societies mind the gap. *British Journal of Social Psychology, 52*, 726–746.

Durante, F., Volpato, C., & Fiske, S. T. (2010). Using the Stereotype Content Model to examine group depictions in Fascism: An archival approach. *European Journal of Social Psychology, 40,* 465-483.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*, 124–129.

Falbo, T. (1977). Multidimensional scaling of power strategies. *Journal of Personality and Social Psychology, 35,* 537-547.

Farias, A. R., Garrido, M. V., & Semin, G. R. (2013). Converging modalities ground abstract categories: the case of politics. *PloS ONE, 8,* e60971.

Farrell, D. (1983). Exit, voice, loyalty, and neglect as responses to job dissatisfaction: A multidimensional scaling study. *The Academy of Management Journal, 26,* 596-607.

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,* 1149-1160.

Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: valence asymmetries. *Journal of Personality and Social Psychology*, *87*, 293-311.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229–238.

Feinberg, M., Willer, R., & Schultz, M. (2014). Gossip and ostracism promote cooperation in groups. *Psychological Science, 25*, 656–664.

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review, 107*, 659-676.

Fiedler, K. (2011). Voodoo correlations are everywhere – Not only in social neurosciences. *Perspectives on Psychological Science, 6*, 163-171.

Fiedler, K. (2014). From intrapsychic to ecological theories in social psychology: Outlines of a functional theory approach. *European Journal of Social Psychology, 44*, 657-670.

Fiedler, K., Kutzner, F., & Vogel, T. (2013). Pseudocontingencies: logically unwarranted but smart inferences. *Current Directions in Psychological Science, 22*, 324-329.

Fiedler, K., & Wänke, M. (2009). The cognitive-ecological approach to rationality in social psychology. *Social Cognition, 27*, 699-732.

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

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

Fiske, S. T., & Dupree, C. (2014). Gaining trust as well as respect in communicating to motivated audiences about science topics. *Proceedings of the National Academy of Sciences, 111*, 13593–13597.

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.

Fiske, S. T., & Pavelchak, M. A. (1986). Category-based versus piecemeal-based affective responses: Developments in schema-triggered affect. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 167-203). New York: Guilford Press.

Flynn, F.J. (2003). How much should I give and how often? The effects of generosity and frequency of favor exchange on social status and productivity. *Academy of Management Journal, 46*, 539–553.

Ford, T. E., & Stangor, C. (1992). The role of diagnosticity in stereotype formation: perceiving group means and variances. *Journal of Personality and Social Psychology, 63,* 356-367.

Forgas, J. P. (1976). The perception of social episodes: Categorical and dimensional representations in two different social milieus. *Journal of Personality and Social Psychology, 34*, 199-209.

Forgas, J. P. (2013). Don't worry, be sad! On the cognitive, motivational, and interpersonal benefits of negative mood. *Current Directions in Psychological Science, 22*, 225-232.

Frable, D. E. (1993). Dimensions of marginality: Distinctions among those who are different. *Personality and Social Psychology Bulletin, 19*, 370-380.

Galinsky, A.D., Maddux, W.W., Gilin, D., &White, J.B. (2008). Why it pays to get inside the head of your opponent: The differential effects of perspective taking and empathy in negotiations. *Psychological Science, 19*, 378–384.

Galinsky, A. D., Magee, J. C., Inesi, M. E., & Gruenfeld, D. H. (2006). Power and perspectives not taken. *Psychological Science*, *17*, 1068-1074.

Ghavami, N., & Peplau, L. A. (2012). An intersectional analysis of gender and ethnic stereotypes testing three hypotheses. *Psychology of Women Quarterly, 37*, 113-127.

Giessner, S. R., & Schubert, T. W. (2007). High in the hierarchy: How vertical location and judgments of leaders' power are interrelated. *Organizational Behavior and Human Decision Processes, 104*, 30-44.

Giguère, G. (2006). Collecting and analyzing data in multidimensional scaling experiments: A guide for psychologists using SPSS. *Tutorial in Quantitative Methods for Psychology, 2,* 27-38.

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

Gino, F., & Pierce, L. (2009). Dishonesty in the name of equity. *Psychological Science, 20,* 1153–1160.

Goldstone, R. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers*, *26*, 381–386.

Goldstone, R. L., Medin, D. L., & Halberstadt, J. (1997). Similarity in context. *Memory & Cognition*, *25*, 237-255.

Goldstone, R. L., & Son, J. Y. (2005). *Similarity*. Cambridge University Press.

Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and Projects* (pp. 22-32). New York: Bobbs-Merrill.

Grammer, K., & Thornhill, R. (1994). Human (Homo sapiens) facial attractiveness and sexual selection: the role of symmetry and averageness. *Journal of Comparative Psychology, 108,* 233-242.

Grant, A. M., & Schwartz, B. (2011). Too much of a good thing the challenge and opportunity of the inverted U. *Perspectives on Psychological Science, 6*, 61-76.

Gräf, M., & Unkelbach, C. (2015). *Halo effects in trait assessment depend on information valence:Why being honest makes you industrious, but lying does not make you lazy*. Manuscript under review.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*, 211-244.

Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition, 87*, 1-32.

Haidt, J. (2012). *The righteous mind: why good people are divided by politics and religion.* New York: Pantheon Books.

Hair, J. F. Jr., Anderson, R. E., Tatham, R. L. & Black, W. C. (1995). *Multivariate Data Analysis, 3rd edn*. New York: Macmillan.

Hamilton, D. L., Sherman, S. J., & Ruvolo, C. M. (1990). Stereotype-based expectancies: Effects on information processing and social behavior. *Journal of Social Issues, 46*, 35-60.

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuro-imaging responses to extreme outgroups. *Psychological Science, 17,* 847-853.

Henley, N. M. (1969). A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior, 8*, 176-184.

Hills, T. T., Todd, P. M., & Goldstone, R. L. (2010). The central executive as a search process: priming exploration and exploitation across domains. *Journal of Experimental Psychology: General*, 139, 590-609.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528–551.

Hornberger, M., Bell, B., Graham, K. S., & Rogers, T. T. (2009). Are judgments of semantic relatedness systematically impaired in Alzheimer's disease? *Neuropsychologia*, *47*, 3084 –3094.

Hout, M. C., & Goldinger, S. D. (2015). Target templates: The precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception & Psychophysics, 77*, 128-149.

Hout, M. C., & Goldinger, S. D. (in press). SpAM is convenient but also satisfying: Reply to Verheyen et al.(2016). *Journal of Experimental Psychology: General*.

Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A multidimensional scaling database with similarity ratings for 240 object categories from the Massive Memory picture database. *PLoS ONE, 9*: e112644.

Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: a fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, *142*, 256-281.

Imhoff, R., & Bruder, M. (2014). Speaking (Un-)Truth to Power: Conspiracy Mentality as a Generalized Political Attitude. *European Journal of Personality, 28*, 25-43.

Imhoff, R., & Koch, A., (2016). How orthogonal are the "Big Two"? On the curvilinear ecological correlation between agency and communion. *Manuscript in preparation.*

Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in psychology, 4.386*.

Inzlicht, M., Schmeichel, B. J., & Macrae, C. N. (2014). Why self-control seems (but may not be) limited. *Trends in Cognitive Sciences, 18*, 127-133.

Izard, C. E. (1971). *The face of emotion*. New York: Appleton-Century-Crofts.

James, W. (1890). *The principles of psychology*. Cambridge, MA: Harvard University Press.

Janssen, O. (2001). Fairness perceptions as a moderator in the curvilinear relationships between job demands, and job performance and job satisfaction. *Academy of Management Journal*, *44*, 1039–1050.

Jaworska, N., & Chupetlovska-Anastasova, A. (2009). A review of multidimensional scaling (MDS) and its utility in various psychological domains. *Tutorials in Quantitative Methods for Psychology, 5*, 1-10.

Jolliffe, I. (2002). *Principal component analysis*. New York: Springer.

Jones, R. A., & Ashmore, R. D. (1973). The structure of intergroup perception: categories and dimensions in views of ethnic groups and adjectives used in stereotype research. *Journal of Personality and Social Psychology, 25*, 428-438.

Jones, M. N., & Mewhort, D. J. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*, 1-37.

Jost, J. T., Federico, C. M., & Napier, J. L. (2009). Political ideology: its structure, functions, and elective affinities. *Annual Review of Psychology*, *60*, 307-337.

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. Psychological Bulletin, 129, 339-375.

Jost, J. T., & van der Toorn, J. (2012). System justification theory. In P. A. M. van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 313–343). London, UK: Sage.

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology, 89*, 899-913.

Kail, R. V., & Cavanaugh J. C. (2010). *Human Development: a Life-Span View* (5th ed.). Australia: Wadsworth Cengage Learning.

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2006). Memory for specific visual details can be enhanced by negative arousing content. *Journal of Memory and Language*, *54*, 99-112.

Kervyn, N., Bergsieker, H. B., & Fiske, S. T. (2012). The innuendo effect: hearing the positive but inferring the negative. *Journal of Experimental Social Psychology, 48*, 77-85.

Kervyn, N., Chan, E., Malone, C., Korpusik, A., & Ybarra, O. (2014). Not all disasters are equal in the public eye: The negativity effect on warmth in brand perception. *Social Cognition, 32*, 256 – 275.

Kervyn, N., Dolderer, M., Mahieu, T., & Yzerbyt, V. Y. (2010). Atypicality and the two fundamental dimensions: Applying the negativity effect on warmth to group perception. *European Journal of Social Psychology*, *40,* 484-489.

Kervyn, N., Fiske, S. T., & Yzerbyt, V. Y. (2013). Integrating the stereotype content model (warmth and competence) and the Osgood semantic differential (evaluation, potency, and activity). *European Journal of Social Psychology, 43,* 673-681.

Kervyn, N., Fiske, S. T., & Yzerbyt, Y. (2015). Foretelling the primary dimension of social cognition: Symbolic and realistic threats together predict warmth in the stereotype content model. *Social Psychology, 46,* 36-45.

Kervyn, N., Yzerbyt, V. Y., Judd, C. M., & Nunes, A. (2009). A question of compensation: the social life of the fundamental dimensions of social perception. *Journal of Personality and Social Psychology, 96*, 828-842.

Klauer, K. C., & Musch, J. (1999). Eine Normierung unterschiedlicher Aspekte der evaluativen Bewertung von 92 Substantiven. *Zeitschrift für Sozialpsychologie*, *30*, 1–11.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, *63*, 190-194.

Koch, A., Alves, H., & Unkelbach, C. (2016) Why good is more alike than bad: an ecological explanation of similarity asymmetries in evaluative information. *Manuscript in preparation.*

Koch, A., Alves, H., Krüger, T., & Unkelbach, C. (in press). A general valence asymmetry in similarity: Good is more alike than bad. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (in press). The ABC of stereotypes about groups: Agency / socio-economic success, conservative-progressive beliefs, and communion. *Journal of Personality and Social Psychology*.

Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. *Journal of Personality and Social Psychology, 107*, 371-392.

Kriegeskorte, N., Mur, M. (2012). Inverse MDS: inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology, 3,* Article 245.

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review, 5,* 445-463.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Newbury Park, CA: Sage.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211-240.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. NIMH, Center for the Study of Emotion & Attention.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*, 390-423.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, *1*, 115-121.

Langston, C. A. (1994). Capitalizing on and coping with daily life events: Expressive responses to positive events. *Journal of Personality and Social Psychology, 67*, 1112–1125.

Layman, G. C., & Carsey, T. M. (2002). Party polarization and "conflict extension" in the American electorate. *American Journal of Political Science, 46,* 786-802.

Leising, D., Ostrovski, O., & Borkenau, P. (2012). Vocabulary for describing disliked persons is more differentiated than vocabulary for describing liked persons. *Journal of Research in Personality*, *46*, 393-396.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., ... & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the Balloon

Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied, 8*, 75-84.

Leonardelli, G. J., Pickett, C. L., & Brewer, M. B. (2010). Optimal distinctiveness theory: A framework for social identity, social cognition, and intergroup relations. *Advances in Experimental Social Psychology, 43,* 63-113.

Lyubomirsky, S., Sousa, L., & Dickerhoof, R. (2006). The costs and benefits of writing, talking, and thinking about life's triumphs and defeats. *Journal of Personality and Social Psychology, 90,* 692–708.

Mackie, D. M., Smith, E. R., & Ray, D. G. (2008). Intergroup emotions and intergroup relations. *Social and Personality Psychology Compass, 2,* 1866-1880.

Macrae, N. C., Milne, A. B., Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology, 66,* 37-47.

Magee, J. C., & Galinsky, A. D. (2008). Social hierarchy: the self-reinforcing nature of power and status. *The Academy of Management Annals*, *2,* 351-398.

Mason, R. L., Gunst, R. F. & Hess, J. L. (1989). *Statistical Design and Analysis of Experiments: Applications to Engineering and Science*. New York: Wiley.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44,* 1-23.

Matlin, M. W., & Stang, D. J. (1978). The Pollyanna principle: Selectivity in language, memory, and thought. Cambridge, MA: Schenkman Pub. Co.

McCarty, N., Poole, K. T., & Rosenthal, H. (2008). Polarized America: the dance of ideology and unequal riches. *Political Science Quarterly, 123*, 687-689.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. Journal of Personality and Social Psychology, 52, 81-90.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60,* 175-215.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30-46.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 558-572.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254-278.

Meier, B. P., Hauser, D. J., Robinson, M. D., Friesen, C. K., & Schjeldahl, K. (2007). What's "up" with God? Vertical space as a representation of the divine. *Journal of Personality and Social Psychology, 93*, 699-710.

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: associations between affect and vertical position. *Psychological Science, 15*, 243-247.

Menard, S. (1995). *Applied Logistic Regression Analysis: Sage University Series on Quantitative Applications in the Social Sciences*. Thousand Oaks, CA: Sage.

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*, 176-182.

Monin, B., & Oppenheimer, D. M. (2005). Correlated averages vs. averaged correlations: demonstrating the warm glow heuristic beyond aggregation. *Social Cognition, 23*, 257-278.

Nelson, L. J., & Miller, D. T. (1995). The distinctiveness effect in social categorization: you are what makes you unusual. *Psychological Science, 6*, 246-249c

Noordewier, M. K., & Breugelmans, S. M. (2013). On the valence of surprise. *Cognition & Emotion, 27*, 1326-1334.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. .*Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 54.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology, 43*, 25-53.

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

Oppenheimer, D. M., & Trail, T. E. (2010). Why leaning to the left makes you lean to the left: effect of spatial orientation on political attitudes. *Social Cognition, 28*, 651-661.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions?. *Psychological Review, 97*, 315-331.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5*, 411-419.

Parsons, T. & Bales, R. F. (1955). *Family socialization and interaction processes*. Glencoe, IL: Free Press.

Pattyn, S., Rosseel, Y., & Van Hiel, A. (2013). Finding our way in the social world: exploring the dimensions underlying social classification. *Social Psychology, 44*, 329-348.

Peeters, G. (1983). Relational and informational patterns in social cognition. In W. Doise & S. Moscovici (Eds.), *Current issues in European social psychology* (Vol. 1, pp. 201-237). Cambridge: Cambridge University Press.

Peeters, G. (1992). Evaluative meanings of adjectives invitro and in context: some theoretical implications and practical consequences of positive-negative asymmetry and behavioral-adaptive concepts of evaluation. *Psychologica Belgica, 32*, 211-231.

Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. *European Review of Social Psychology*, *1*, 33-60.

Perea, M., Duñabeitia, J. A., & Carreiras, M. (2008). R34d1ng w0rd5 w1th numb3r5. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 237-241.

Plutchik, R. (2001). The nature of emotions. *American Scientist*, *89*, 344-350.

Potter, T., Corneille, O., Ruys, K. I., & Rhodes, G. (2007). “Just another pretty face”: A multidimensional scaling approach to face attractiveness and variability. *Psychonomic Bulletin & Review*, *14*, 368-372.

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention grabbing power of negative social information. *Journal of Personality and Social Psychology, 61*, 380-391.

Quine, W. V. (1969) *Ontological Relativity and Other Essays*. New York, Columbia University Press.

Reis, H. T., Gable, S. L., & Maniaci, M. R. (2014). Methods for studying everyday experience in its natural context. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 373-402). Cambridge, MA: Harvard University Press.

Reis, H. T., Smith, S. M., Carmichael, C. L., Caprariello, P. A., Tsai, F. F., Rodrigues, A., & Maniaci, M. R. (2010). Are you happy for me? How sharing positive events with others provides personal and interpersonal benefits. *Journal of Personality and Social Psychology, 99,* 311–329.

Reyna, C., Dobria, O., & Wetherell, G. (2013). The complexity and ambivalence of immigration attitudes: Ambivalent stereotypes predict conflicting attitudes toward immigration policies. *Cultural Diversity and Ethnic Minority Psychology, 19,* 342.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology, 57***,** 199-226.

Rhodes, G., Halberstadt, J., & Brajkovich, G. (2001). Generalization of mere exposure effects to averaged composite faces. *Social Cognition, 19*, 57-70.

Rhodes, G., Proffitt, F., Grady, J. M., & Sumich, A. (1998). Facial symmetry and the perception of beauty. *Psychonomic Bulletin & Review, 5*, 659-669.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803-814.

Rom, S., & Hofmann, W. (2015). Labour-leisure tradeoffs in everyday life. *Manuscript in preparation.*

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9*, 283-294.

Rozin, P., Berman, L., & Royzman, E. (2010). Biases in use of positive and negative words across twenty natural languages. *Cognition and Emotion*, *24,* 536-548.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5,* 296-320.

Rusbult, C. E., & Zembrodt, I. M. (1983). Responses to dissatisfaction in romantic involvements: A multidimensional scaling analysis. *Journal of Experimental Social Psychology, 19*, 274-293.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161–1178.

Sauter, D. (2010). More than happy: The need for disentangling positive emotions. *Current Directions in Psychological Science, 19*, 36-40.

Schiffman, S. S., Reynolds, M. L., & Young, E. W. (1981). *Introduction to multidimensional scaling: Theory, methods and applications.* New York: Academic Press.

Schrauf, R. W., & Sanchez, J. (2004). The preponderance of negative emotion words in the emotion lexicon: A cross-generational and cross-linguistic study. *Journal of Multilingual and Multicultural Development*, *25,* 266-284.

Schubert, T. W. (2005). Your highness: vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology, 89*, 1-21.

Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values?. *Journal of Social Issues, 50*, 19-45.

Schwartz, S. H., & Bilsky, W. (1987). Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology, 53,* 550-562.

Schwartz, S. H., & Bilsky, W. (1990). Toward a theory of the universal content and structure of values: extensions and cross-cultural replications. *Journal of Personality and Social Psychology*, *58,* 878-891.

Schwartz, R. M., & Humphreys, M. S. (1973). Similarity judgments and free recall of unrelated words. *Journal of Experimental Psychology*, *101*, 10-15.

Seery, M. D. (2011). Resilience A Silver Lining to Experiencing Adverse Life Events?. *Current Directions in Psychological Science, 20*, 390-394.

Semin, G. R., & Fiedler, K. (1992). *Language, interaction and social cognition*. Thousand Oaks, CA US: Sage Publications, Inc.

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology, 52*, 1061-1068.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

Sherman, J. W., Lee, A. Y., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype efficiency reconsidered: encoding flexibility under cognitive load. *Journal of Personality and Social Psychology, 75*, 589-606.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin, 86, 420-428.

Skowronski, J., & Carlston, D. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*, 131–142.

Slepian, M. L., Masicampo, E. J., & Ambady, N. (2015). Cognition from on high and down low: Verticality and construal level. *Journal of Personality and Social Psychology, 108,* 1-17.

Smallman, R., Becker, B., & Roese, N. J. (2014). Preferences for expressing preferences: People prefer finer evaluative distinctions for liked than disliked objects. *Journal of Experimental Social Psychology*, *52*, 25-31.

Smallman, R., & Roese, N. J. (2008). Preference invites categorization. *Psychological Science*, *19*, 1228-1232.

Smith, E. R. (1993). Social identity and social emotions: Toward new conceptualizations of prejudice. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition and stereotyping: Interactive processes in group perception* (pp. 297-315). San Diego, CA: Academic Press.

Smith, E. E., & Sloman, S. A. (1994). Similarity-versus rule-based categorization. *Memory & Cognition*, *22*, 377-386.

Sofer, C., Dotsch, R., Wigboldus, D. H. J., & Todorov, A. (2015). What is typical is good: the influence of face typicality on perceived trustworthiness. *Psychological Science. 26*, 39-47.

Somers, M.J., & Casal, J.C. (1994). Organizational commitment and whistle-blowing: A test of the reformer and the organization man hypotheses. *Group & Organization Management, 19,* 270–284.

Sturman, M.C. (2003). Searching for the inverted U-shaped relationship between time and performance: Meta-analyses of the experience/performance, tenure/performance, and age/performance relationships. *Journal of Management, 29*, 609–640.

Suitner, C. & Maass, A. (2008). The role of valence in the perception of agency and communion. *European Journal of Social Psychology, 38*, 1073-1082.

Sutherland, C. A., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: ambient images generate a three-dimensional model. *Cognition, 127*, 105-118.

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Social Issues, 25*, 79–97.

Tajfel, H. & Turner, J.C. (1979). An integrative theory of intergroup conflict. In W.G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Monterey, CA: Brooks-Cole.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis. *Psychological Bulletin*, *110*, 67-85.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*, 629-640.

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

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*, 519-545.

Tolstoy, L. (1873-1877/2001). *Anna Karenina* (Original work published 1875-1877) (R. P. L. Volokhonsky, Trans.). New York, NY, USA: Viking Penguin.

Topolinski, S., & Deutsch, R. (2012). Phasic affective modulation of creativity. *Experimental Psychology*, *59*, 302.

Topolinski, S., & Deutsch, R. (2013). Phasic affective modulation of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 414.

Torgerson, W. S. (1965). Multidimensional scaling of similarity. *Psychometrika, 30,* 379-393.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327-352.

Unkelbach, C. (2012). Positivity advantages in social information processing. *Social and Personality Psychology Compass, 6*, 83-94.

Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis. *Journal of Personality and Social Psychology*, *95*, 36-49.

Unkelbach, C., Forgas, J. P., & Denson, T. F. (2008). The turban effect: the influence of Muslim headgear and induced affect on aggressive responses in the shooter bias paradigm. *Journal of Experimental Social Psychology, 44*, 1409-1413.

Unkelbach, C., von Hippel, W., Forgas, J. P., Robinson, M. D., Shakarchi, R. J., & Hawkins, C. (2010). Good things come easy: Subjective exposure frequency and the faster processing of positive information. *Social Cognition, 28*, 538–555.

van Elk, M., van Schie, H. T., & Bekkering, H. (2010). From left to right: processing acronyms referring to names of political parties activates spatial associations. *The Quarterly Journal of Experimental Psychology, 63*, 2202-2219.

van Kleef, G. A., De Dreu, C. K., Pietroni, D., & Manstead, A. S. (2006). Power and emotion in negotiation: power moderates the interpersonal effects of anger and happiness on concession making. *European Journal of Social Psychology, 36*, 557-581.

Verhaeghen, P., Aikman, S. N., & Van Gulick, A. E. (2011). Prime and prejudice: cooccurrence in the culture as a source of automatic stereotype priming. *British Journal of Social Psychology, 50*, 501-518.

Verheyen, S., Voorspoels, W., Vanpaemel, W., & Storms, G. (in press). Caveats for the spatial arrangement method: Comment on Hout, Goldinger, and Ferguson (2013). *Journal of Experimental Psychology: General.*

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science, 4,* 274-290.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*, 1191-1207.

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25,* 1115-1125.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General, 143*, 2020-2045.

Westfall, J., Judd, C. M., & Kenny, D. A. (2015). Replicating studies in which samples of participants respond to samples of stimuli. Replicating Studies in Which Samples of Participants Respond to Samples of Stimuli. *Perspectives on Psychological Science.*

Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin, 127*, 797-826.

Williams, K. D. (2007). Ostracism. *Annual Review of Psychology, 58*, 425–452.

Williams, J. E., & Best, D. L. (1990). *Measuring sex stereotypes: A multination study*. Newbury Park, CA: Sage.

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

Windsor, T.D., Anstey, K.J., & Rodgers, B. (2008). Volunteering and psychological well-being among young-old adults: How much is too much? *The Gerontologist, 48*, 59–70.

Wojciszke, B. (1994). Multiple meanings of behavior: Construing actions in terms of competence and morality. *Journal of Personality and Social Psychology, 67*, 222-232.

Wojciszke, B., & Abele, A. E. (2008). The primacy of communion over agency and its reversals in evaluations. *European Journal of Social Psychology, 38*, 1139-1147.

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin, 24*, 1251-1263.

Wrosch, C., Scheier, M.F., Carver, C.S., & Schulz, R. (2007). The importance of goal disengagement in adaptive self-regulation: When giving up is beneficial. *Self and Identity, 2,* 1–20.

Yerkes, R. M., & Dodson, J. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*, 459-482.

Young, F. W., Takane, Y., & Lewyckyj, R. (1978). ALSCAL: a nonmetric multidimensional scaling program with several individual-differences options. *Behavior Research Methods*, *10*, 451-453.

Yzerbyt, V. Y., Kervyn, N., & Judd, C. M. (2008). Compensation versus halo: the unique relations between the fundamental dimensions of social judgment. *Personality and Social Psychology Bulletin*, *34*, 1110-1123.