Journal of Experimental Psychology:
Learning, Memory, and Cognition
2016, Vol. 42, No. 8, 1171–1192

0278-7393/16/$12.00 http://dx.doi.org/10.1037/xlm0000243

A General Valence Asymmetry in Similarity: Good Is More Alike Than Bad

Alex Koch and Hans Alves
University of Cologne

Tobias Krüger
University of Heidelberg

Christian Unkelbach
University of Cologne

The density hypothesis (Unkelbach, Fiedler, Bayer, Stegmüller, & Danner, 2008) claims a general higher similarity of positive information to other positive information compared with 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 1 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.

Keywords: valence, similarity, processing asymmetry, sampling, representative design

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, Fiedler, Bayer, Stegmüller, and Danner (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 interstimulus 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, Fiedler, Bayer, Stegmüller, and Danner, 2008). Then, we review existing evidence for the proposed generality of higher similarity of positive compared with 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 with 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 &

This article was published Online First February 11, 2016.

Alex Koch and Hans Alves, Social Cognition Center Cologne, University of Cologne; Tobias Krüger, Social Psychology, University of Heidelberg; Christian Unkelbach, Social Cognition Center Cologne, University of Cologne.

Correspondence concerning this article should be addressed to Alex S. Koch, Social Cognition Center Cologne, University of Cologne, Richard-Strauss-Straβe 2, 50931 Köln, Germany. E-mail: alex.koch@uni-koeln.de

Son, 2005; Hahn, Chater, & Richardson, 2003; Krumhansl, 1978; Medin, Goldstone, & Gentner, 1993; Tversky, 1977), interstimulus similarity undoubtedly impacts learning, memory, and cognition in profound ways ("sameness is [. . .] the backbone of our thinking", James, 1890; 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 (Nosofsky, 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, interstimulus 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, Sanbonmatsu, 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, 2013; 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 with 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, 1974; Cannon, 1926).

The *frequency of stimulus co-occurrence* argument is that positive compared with 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 (1978) 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 interstimulus 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 with 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; see also Alves, Koch, &

Unkelbach, 2016b). 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 with 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 with 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 with 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 interstimulus 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 et al.'s (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 with 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 with 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 et al. (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, Kenny, and Judd (2014), 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 interstimulus 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; Goldstone, 1994; Hout, Goldinger, & Ferguson, 2013; 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, Goldinger, and Ferguson (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 (latent semantic analysis [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, and Brysbaert (2013; WKB; ~14,000 words), and all positive and negative pictures in the international affective picture system (IAPS; Lang, Bradley, & Cuthbert, 2005; ~1,000 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 et al. (2008).

Study 1

Participants spatially arranged the 20 positive and 20 negative words investigated by Unkelbach et al. (2008) and then divided these words into between two and seven 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' interstimulus 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 U.S. college students (LSA; see Landauer & Dumais, 1997).

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. Fifty-five 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 et al. (2008; see Appendix A). These 40 words were first used by Fazio, Sanbonmatsu, Powell, and Kardes (1986), and were translated into German by Klauer and 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 (1,920 × 1,080 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 × 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 et al. (2008). The screen diagonal serves as a fixed calibration divisor.

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 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 min, and sorting them into between two and seven categories took less than 5 min.

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 et al. (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 with 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

[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 et al. (2008).

[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.

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.

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]

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 U.S. college students (LSA similarity; Landauer & Dumais, 1997). As frequency of co-occurrence in space and time is a widely accepted proxy for interstimulus 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 web pages 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 10 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, 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 et al. (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, 1997) online tool provided by the University of Colorado at Boulder, (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 104,852 words in all 942,425 passages in a collection of 738 books that is representative of the literature read by U.S. 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 U.S. 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.,

[3] The difference between our spatial arrangement version and the version used by Hout et al. (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 intracategory similarity of the 40 words was high when intracategory 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 interrater agreement for Pairwise and SpAM intracategory similarity is presumably due the number of trials across which it is measured. In the article by Unkelbach et al. (2008), participants on average made approximately 10 pairwise intracategory similarity judgments per word, whereas in Study 1, participants made only one spatial intracategory 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 intracategory similarity. This advantage comes at the cost of low interrater agreement about intracategory similarity. 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 with 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.

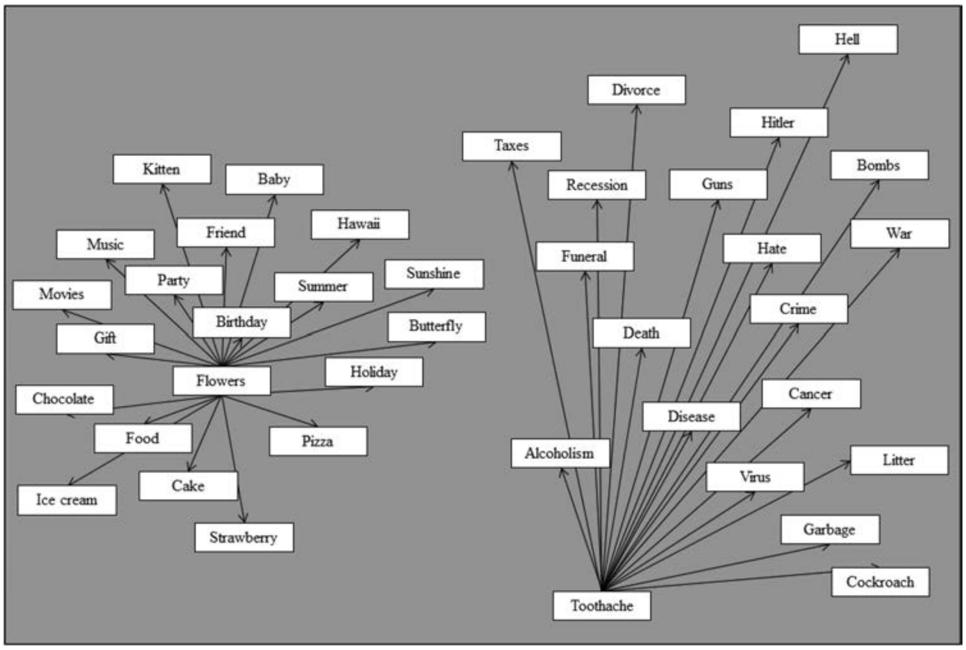

Figure 1. Example of distances in a possible SpAM solution. The 20 positive target words are clustered on the left side and the 20 negative target words are clustered on the right side. Here, the proximity, and thus similarity, between "flowers" and the other 19 positive words is greater than the proximity/similarity between "toothache" and the other 19 negative concepts.

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.

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 coactivate 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 coactivated 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.

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 with the standard measure of similarity.

Table 1
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	β	t	p	r	pr^2	VIF
Valence	.71	4.02	<.001	.70	.57	3.04
Valence intensity	.37	3.26	<.01	.25	.49	1.27
Frequency	.01	.04	.97	.24	.01	1.37
Familiarity	.12	.78	.44	.65	.13	2.62
Concreteness	−.10	−.81	.42	.17	−.14	1.37

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

Table 2
Correlations Between the SpAM, Pairwise, Google, and LSA Measures of Similarity

Similarity measure	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 web pages accessible through Google Search (Google) and in passages in a collection of books that is representative of the literature read by U.S. college students (LSA).
* $p < .05$. ** $p < .01$. *** $p < .001$.

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 reexamined the similarity of the 40 words examined by Unkelbach et al. (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, Judd, & Kenny, 2015; Westfall, Kenny, & Judd, 2014). Yet, the cost- and time-effective SpAM method allows testing the generality of the proposed higher similarity of positive compared with 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

Table 3
Predictive Strength of the SpAM, Pairwise, Google, and LSA Measures of Similarity

Similarity measure	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 web pages accessible through Google Search (Google) and in passages in a collection of books that is representative of the literature read by U.S. college students (LSA).
* $p < .05$. ** $p < .01$. *** $p < .001$.

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 min.

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 min; $M = 8.08$, $SD = 5.52$). This exclusion of participants did not affect any statistical inferences.

Study 2a. Participants generated 1,044 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; Koch, Imhoff, Dotsch, Unkelbach, & Alves, in press) to measure the similarity of a large sample of words (i.e., 1,044); 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 with 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 with 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 (Denrell, 2005; Fazio et al., 2004). 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. One-hundred and 10 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 $r = .70$, $p < .001$, 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 min.

Results

Participants in the idiosyncrasy condition took an equal amount of time to generate the 20 positive and 20 negative words ($M = 490$ s; $SD = 188$ s) as those in the consensus condition ($M = 499$ s; $SD = 197$ s), $F(1, 108) = 0.05$, $p = .83$, $\eta_p^2 = .00$, 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 (1,139 unique stimuli out of the 55 participants * 40 stimuli = 2,200 generated stimuli) than participants in the consensus condition (995 unique out of 2,200 generated stimuli). This difference was significant, $\chi^2(1) = 9.71$, $p < .01$.

Frequency of unique words. Independent of the idiosyncrasy versus consensus manipulation, participants generated less unique stimuli for the category "positive" (946 out of 2,200) compared with the category "negative" (1,180 out of 2,200), $\chi^2(1) = 27.44$, $p < .001$. This smaller diversity was apparent in both the idiosyncratic valence condition (511 unique positive words vs. 628 unique negative words), $\chi^2(1) = 12.01$, $p < .001$, and in the consensual valence condition (435 unique positive words vs. 560 unique negative words), $\chi^2(1) = 15.70$, $p < .001$.

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) $\times$ 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.

Discussion

Participants adhered to the instructions and generated more diverse stimuli in the idiosyncratic compared with 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 with dislike (Alves, Koch, & Unkelbach, 2016a; Smallman, Becker, & Roese, 2014), they spatially arranged positive idiosyncratic stimuli more densely to one another than negative idiosyncratic stimuli. This valence asymmetry was as pronounced ($\eta_p^2 = .28$) as in the consensual valence condition ($\eta_p^2 = .24$).

The effect sizes ($M\eta_p^2 = .26$) are close to Study 1 ($\eta_p^2 = .32$), 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 et al. (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*—

Table 4
Similarity Means and F-Tests for Positive and Negative Stimuli in Study 3's Idiosyncratic and Consensual Conditions (Standard Deviations in Parentheses)

Valence condition	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. CI = confidence interval; LB = lower bound; UB = upper bound.

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 with 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, Eiser, and Shook (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 nonredundant positive words generated in this way were combined to form a multisource (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 nonredundant negative words were combined in the same way. Out of these two multisource 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. Forty 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 subselection of 40 of the words retrieved in this way (20 out of 29 nonredundant positive words, e.g., “courage,” “happy,” “awesome,” etc.; and 20 out of 35 nonredundant 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 nonredundant positive words, and 20 negative words, randomly selected from the 35 nonredundant 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 min.

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 with 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, fre-

[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 & 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.

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	β	*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	.28	.78	.13	.04	1.10
Familiarity	−.18	−.86	.39	.35	−.11	3.88
Concreteness	.05	.42	.68	−.12	.06	1.35

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

quency, familiarity, and concreteness.[7] As in Study 1, the only other significant predictor of similarity was valence intensity.

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 et al. (2008; Experiment 2) they could not contribute, which shows that the asymmetry persists independent of these possible contributions.

The effect ($\eta_p^2 = .12$) 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 (Denrell, 2005; Fazio et al., 2004), 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 7 consecutive days, participants named one positive and one negative "event of the day" and then—on Day 8, 9, or 10—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 min). The links in these text messages redirected participants to the survey web site on which the study was hosted. On this web site, 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 e-mail with instructions on how to complete the final task (see below) within 3 days. One-hundred and 24 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/

[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 et al. (1995) and Mason et al. (1989), multicollinearity is a concern with VIFs greater than 10. None of our predictors had a VIF greater than 5.

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

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 five to seven 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) × 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, $SD = 9.67$; $M_{neg} = 21.44\%$, $SD = 10.21$). 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 ($\eta_p^2 = .03$–$.06$) 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, multiword 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 et al. (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 reanalyzed 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 reanalyzed 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.

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	.59 (.28)	.72 (.33)	620.32	<.001	.04	.04	.05
Valence	.54 (.37)	.74 (.46)	848.04	<.001	.06	.05	.06
Arousal	.69 (.53)	.73 (.54)	18.68	<.001	.00	.00	.00
Dominance	.55 (.42)	.70 (.50)	340.30	<.001	.02	.02	.03
Valence residuals	−.08 (.37)	.08 (.44)	603.74	<.001	.04	.04	.05
Arousal residuals	0 (.53)	0 (.53)	.14	.71	.00	.00	.00
Dominance residuals	−.04 (.42)	.04 (.48)	121.09	<.001	.01	.01	.01
Images							
Overall	.88 (.19)	1.30 (.29)	702.29	<.001	.42	.39	.46
Valence	.87 (.23)	1.28 (.27)	610.23	<.001	.39	.35	.43
Arousal	1.19 (.38)	1.46 (.43)	106.30	<.001	.10	.07	.13
Dominance	.59 (.26)	1.17 (.33)	910.11	<.001	.49	.45	.52
Valence residuals	−.06 (.26)	.06 (.23)	49.41	<.001	.05	.03	.07
Arousal residuals	.03 (.39)	−.03 (.32)	5.96	<.05	.01	.00	.02
Dominance residuals	−.11 (.30)	.11 (.21)	159.82	<.001	.14	.11	.18

Note. Values reflect the average absolute rating difference (on a 9-point scale) between all 6,958/478 positive WKB words/IAPS pictures and all 6,958/478 negative WKB words/IAPS pictures. Lower values indicate higher interstimulus similarity. WKB = Warringer, Kuperman and Brysbaert; IAPS = international affective picture system; CI = confidence interval; LB = lower bound; UB = upper bound.

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 with negative words/pictures was not higher (see Table 6).

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 ($\eta_p^2 = .04$), and separately for the valence ratings/residuals ($\eta_p^2 = .06/.04$) and dominance ratings/residuals ($\eta_p^2 = .02/.01$), but not for the arousal ratings/residuals ($\eta_p^2 = .00/.00$).

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 ($\eta_p^2 = .34$), and separately for the valence ratings/residuals ($\eta_p^2 = .39/.05$), dominance ratings/residuals ($\eta_p^2 = .49/.15$), and arousal ratings ($\eta_p^2 = .10$), but not for the arousal residuals ($\eta_p^2 = .01$).

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 ($M = 1.03$, $SD = 0.76$), $F(1, 14869) = 394.16$, $p < .001$, $\eta_p^2 = .03$, 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 with 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) 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 article 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 with 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.

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.

[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.

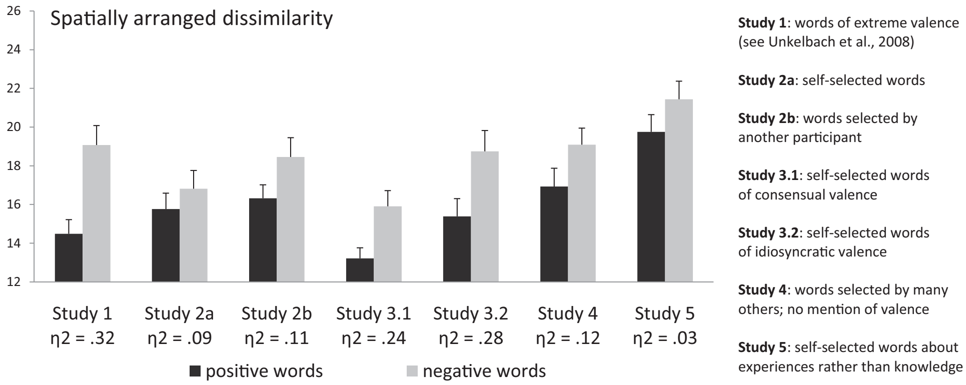

Figure 2. Spatially arranged distance of positive to other positive words and negative to other negative words in percentage of the screen diagonal, and effect sizes in Studies 1–5. Participants freely sampled positive and negative stimuli and spatially arranged them on a blank screen. Participants arranged positive words more densely (i.e., more similar) to one another than the negative words.

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 toward 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 artifact 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)

Study	Positive valence	Negative valence	F	p	η_p^2	90% CI LB	90% CI UB
Overall	24.00 (7.62)	24.09 (7.57)	.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)	.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)	.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)	.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. CI = confidence interval; LB = lower bound; UB = upper bound.

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 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 (a) 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; and (b) 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 (Nosofsky, 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

[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]).

similarity may explain downstream valence asymmetries on many levels of cognitive processing.

References

Alves, H., Koch, A., & Unkelbach, C. (2016a). My friends are all alike—On the relation between liking, knowledge and perceived similarity. *Journal of Experimental Social Psychology, 62,* 103–117.

Alves, H., Koch, A., & Unkelbach, C. (2016b). *The "Common Good" in Person Perception: An amplification/attenuation model of social comparison*. Manuscript in preparation.

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. http://dx.doi.org/10.3758/s13421-015-0515-5

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. http://dx.doi.org/10.1037/0022-3514.87.5.573

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 & Personality Science, 2,* 508–515. http://dx.doi.org/10.1177/1948550611399154

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

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5,* 323–370. http://dx.doi.org/10.1037/1089-2680.5.4.323

Berman, M. G., Hout, M. C., Kardan, O., Hunter, M., Yourganov, G., Henderson, J. M., . . . Jonides, J. (2014). The perception of naturalness converges with objective measures of naturalness. *PLoS One, 9,* e114572. http://dx.doi.org/10.1371/journal.pone.0114572

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

Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology, 3,* 339–344. http://dx.doi.org/10.1002/ejsp.2420030312

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, NY: Psychology Press.

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

Bruckmüller, S., & Abele, A. E. (2013). The density of the Big Two: How are agency and communion structurally represented? *Social Psychology, 44,* 63–74. http://dx.doi.org/10.1027/1864-9335/a000145

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62,* 193–217. http://dx.doi.org/10.1037/h0047470

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

Cannon, W. B. (1926). Physiological regulation of normal states. Some tentative postulates concerning biological homeostasis. In W. B. Cannon & A. Pettit (Eds.), *À Charles Richet, ses amis, ses collègues, ses élèves* [To Charles Richet and his friends, colleagues, and scholars]. (pp. 91–93). Paris, France: Les Editions Médicales.

Cilibrasi, R. L., & Vitanyi, P. M. B. (2007). The Google similarity distance. *IEEE Transactions on Knowledge and Data Engineering, 19,* 370–383. http://dx.doi.org/10.1109/TKDE.2007.48

Davies, M. (2011). *Word frequency data from the Corpus of Contemporary American English*. Provo, UT: COCA.

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* (pp. 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. http://dx.doi.org/10.1080/10463283.2014.892320

Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review, 112,* 951–978. http://dx.doi.org/10.1037/0033-295X.112.4.951

DeSoto, K. A., & Roediger, H. L., III. (2014). Positive and negative correlations between confidence and accuracy for the same events in recognition of categorized lists. *Psychological Science, 25,* 781–788. http://dx.doi.org/10.1177/0956797613516149

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. http://dx.doi.org/10.1037/h0033731

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17,* 124–129. http://dx.doi.org/10.1037/h0030377

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. http://dx.doi.org/10.3758/BRM.41.4.1149

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. http://dx.doi.org/10.1037/0022-3514.50.2.229

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

Fiedler, K. (2011). Voodoo correlations are everywhere—Not only in social neurosciences. *Perspectives on Psychological Science, 6,* 163–171. http://dx.doi.org/10.1177/1745691611400237

Fiedler, K., Kutzner, F., & Vogel, T. (2013). Pseudocontingencies: Logically unwarranted but smart inferences. *Current Directions in Psychological Science, 22,* 324–329. http://dx.doi.org/10.1177/0963721413480171

Fiedler, K., & Wänke, M. (2009). The cognitive-ecological approach to rationality in social psychology. *Social Cognition, 27,* 699–732. http://dx.doi.org/10.1521/soco.2009.27.5.699

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. http://dx.doi.org/10.1037/0022-3514.34.2.199

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. http://dx.doi.org/10.1177/0963721412474458

Goldstone, R. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments & Computers, 26,* 381–386. http://dx.doi.org/10.3758/BF03204653

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

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

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. http://dx.doi.org/10.1037/0033-295X.114.2.211

Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition, 87,* 1–32. http://dx.doi.org/10.1016/S0010-0277(02)00184-1

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

Henley, N. M. (1969). A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior, 8,* 176–184. http://dx.doi.org/10.1016/S0022-5371(69)80058-7

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551. http://dx.doi.org/10.1037/0033-295X.95.4.528

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. http://dx.doi.org/10.1016/j.neuropsychologia.2009.07.006

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. http://dx.doi.org/10.3758/s13414-014-0764-6

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. http://dx.doi.org/10.1371/journal.pone.0112644

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. http://dx.doi.org/10.1037/a0028860

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

James, W. (1890). *The principles of psychology*. Cambridge, MA: Harvard University Press. http://dx.doi.org/10.1037/11059-000

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.

Jones, M. N., & Mewhort, D. J. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review, 114,* 1–37. http://dx.doi.org/10.1037/0033-295X.114.1.1

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. http://dx.doi.org/10.1016/j.jml.2005.05.005

Klauer, K. C., & Musch, J. (1999). Eine Normierung unterschiedlicher Aspekte der evaluativen Bewertung von 92 Substantiven [Norms for different aspects of the evaluation of 92 nouns]. *Zeitschrift für Sozialpsychologie, 30,* 1–11. http://dx.doi.org/10.1024//0044-3514.30.1.1

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis, 63,* 190–194. http://dx.doi.org/10.1093/analys/63.3.190

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

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*.

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology, 3,* 245. http://dx.doi.org/10.3389/fpsyg.2012.00245

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review, 85,* 445–463. http://dx.doi.org/10.1037/0033-295X.85.5.445

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. http://dx.doi.org/10.1037/0033-295X.104.2.211

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. Gainesville, FL: 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. http://dx.doi.org/10.1111/j.1467-9280.1990.tb00079.x

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.

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. http://dx.doi.org/10.1016/j.jrp.2012.03.006

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.

Mason, R. L., Gunst, R. F., & Hess, J. L. (1989). *statistical design and analysis of experiments: Applications to engineering and science*. New York, NY: Wiley.

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

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 558–572. http://dx.doi.org/10.1037/0278-7393.24.3.558

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100,* 254–278. http://dx.doi.org/10.1037/0033-295X.100.2.254

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

Noordewier, M. K., & Breugelmans, S. M. (2013). On the valence of surprise. *Cognition and Emotion, 27,* 1326–1334. http://dx.doi.org/10.1080/02699931.2013.777660

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–61. http://dx.doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 54–65. http://dx.doi.org/10.1037/0278-7393.14.1.54

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review, 97,* 315–331. http://dx.doi.org/10.1037/0033-295X.97.3.315

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

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. http://dx.doi.org/10.1080/14792779108401856

Perea, M., Duñabeitia, J. A., & Carreiras, M. (2008). R34d1ng w0rd5 w1th numb3r5. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 237–241. http://dx.doi.org/10.1037/0096-1523.34.1.237

Plutchik, R. (2001). The nature of emotions. *American Scientist, 89,* 344–350. http://dx.doi.org/10.1511/2001.4.344

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. http://dx.doi.org/10.3758/BF03194079

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. http://dx.doi.org/10.1037/0022-3514.61.3.380

Quine, W. V. (1969). *Ontological relativity and other essays*. New York, NY: 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.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology, 57,* 199–226. http://dx.doi.org/10.1146/annurev.psych.57.102904.190208

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. http://dx.doi.org/10.1037/0278-7393.21.4.803

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. http://dx.doi.org/10.1037/h0026086

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. http://dx.doi.org/10.1080/02699930902793462

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5,* 296–320. http://dx.doi.org/10.1207/S15327957PSPR0504_2

Sauter, D. (2010). More than happy: The need for disentangling positive emotions. *Current Directions in Psychological Science, 19,* 36–40. http://dx.doi.org/10.1177/0963721409359290

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. http://dx.doi.org/10.1080/01434630408666532

Schwartz, R. M., & Humphreys, M. S. (1973). Similarity judgments and free recall of unrelated words. *Journal of Experimental Psychology, 101,* 10–15. http://dx.doi.org/10.1037/h0035511

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

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,* 1061–1086. http://dx.doi.org/10.1037/0022-3514.52.6.1061

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323. http://dx.doi.org/10.1126/science.3629243

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

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. http://dx.doi.org/10.1016/j.jesp.2013.12.004

Smallman, R., & Roese, N. J. (2008). Preference invites categorization. *Psychological Science, 19,* 1228–1232. http://dx.doi.org/10.1111/j.1467-9280.2008.02229.x

Smith, E. E., & Sloman, S. A. (1994). Similarity-versus rule-based categorization. *Memory & Cognition, 22,* 377–386. http://dx.doi.org/10.3758/BF03200864

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin, 110,* 67–85. http://dx.doi.org/10.1037/0033-2909.110.1.67

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629–640. http://dx.doi.org/10.1017/S0140525X01000061

Topolinski, S., & Deutsch, R. (2012). Phasic affective modulation of creativity. *Experimental Psychology, 59,* 302–310. http://dx.doi.org/10.1027/1618-3169/a000159

Topolinski, S., & Deutsch, R. (2013). Phasic affective modulation of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 414–436. http://dx.doi.org/10.1037/a0028879

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352. http://dx.doi.org/10.1037/0033-295X.84.4.327

Unkelbach, C. (2012). Positivity advantages in social information processing. *Social and Personality Psychology Compass, 6,* 83–94. http://dx.doi.org/10.1111/j.1751-9004.2011.00407.x

Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis. *Journal of Personality and Social Psychology, 95,* 36–49. http://dx.doi.org/10.1037/0022-3514.95.1.36

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.

Verhaeghen, P., Aikman, S. N., & Van Gulick, A. E. (2011). Prime and prejudice: Co-occurrence in the culture as a source of automatic stereotype priming. *British Journal of Social Psychology, 50,* 501–518. http://dx.doi.org/10.1348/014466610X524254

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. http://dx.doi.org/10.3758/s13428-012-0314-x

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25,* 1115–1125. http://dx.doi.org/10.1177/01461672992512005

Westfall, J., Judd, C. M., & Kenny, D. A. (2015). Replicating studies in which samples of participants respond to samples of stimuli. *Perspectives on Psychological Science, 10,* 390–399.

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. http://dx.doi.org/10.1037/xge0000014

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

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.

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 et al. (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. Groβartig	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. Groβzügig	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

(*Appendices continue*)

Appendix B (*continued*)

Study 4 original stimuli	Study 4 translations	Study 4 original stimuli	Study 4 translations
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.

Received May 19, 2015
Revision received December 8, 2015
Accepted December 14, 2015 ■