CHAPTER THREE

Negativity bias, positivity bias, and valence asymmetries: Explaining the differential processing of positive and negative information

Christian Unkelbach[a,*], **Hans Alves**[a], **Alex Koch**[b]
[a]Universität zu Köln, Köln, Germany
[b]University of Chicago, Chicago, IL, United States
[*]Corresponding author: e-mail address: christian.unkelbach@uni-koeln.de

Contents

Advances in Experimental Social Psychology, Volume 62
ISSN 0065-2601
https://doi.org/10.1016/bs.aesp.2020.04.005

Abstract

Distinguishing between "good" and "bad" is a fundamental task for all organisms. However, people seem to process positive and negative information differentially, described in the literature as instances of negativity bias, positivity bias, or valence asymmetries. We provide an overview of these processing differences and their explanations. First, we review negativity advantages: People attend more to negative information, recall it more, and weigh it more heavily, relative to positive information. Second, we review positivity advantages: People process positive information faster, have broader associations from it, and show stronger congruency effects, relative to negative information. We then discuss existing explanations for these differential effects in terms of phylogenetic pressures, correlates of valence, diagnosticity, mobilization-minimization, and top-down vs. bottom-up processing. Finally, we suggest the differential similarity of positive and negative information as a unifying explanation. We delineate why positive information should be more alike relative to negative information, and how differential similarity translates to the observed processing differences. Then we show how the similarity explanation leads to novel predictions and how it solves old puzzles. Similarity thereby provides an explanatory construct for both positivity and negativity advantages, allowing precise quantitative predictions for valence asymmetries beyond the mere classification of "good" and "bad."

The distinction between "good" and "bad" is a fundamental basis of social life. People constantly classify stimuli, events, or behaviors as good or bad; they *evaluate* their environment. These evaluations determine judgments (e.g., whether to accept/reject something) and behaviors (e.g., whether to approach/avoid something). While stimuli, events, or situations often simultaneously possess good or bad features, most people easily distinguish between positive and negative information. However, these two classes of good and bad information (i.e., stimuli, events, behaviors) are not created equal; they differ in the way they are processed. This phenomenon is known as *valence asymmetries* in processing (Kanouse & Hanson, 1972; Peeters, 1971;

Peeters & Czapinski, 1990); that is, people apparently process positive information and negative information differentially. A friendly smile, a kind comment, and a loving hug, all elicit different processes than an angry frown, a harsh remark, or an aggressive shove.

We provide a model that explains this differential processing of positive and negative information. To do so, we will first clarify our concepts and the distinction between "good" and "bad," or positive and negative information. Then, we will review some prominent differences in the way positive and negative information is processed. Next, we will present existing explanation for these differences, and building upon these explanations, we will present our own model. Finally, we present empirical evidence to support this model and end with a discussion of its advantages.

1. A definition of "good" and "bad"

Distinguishing between "good" and "bad" seems to be an easy task for people. However, defining what actually makes a stimulus, an event, or information in general "good" or "bad" is much less trivial. For example, Baumeister, Bratslavsky, Finkenauer, and Vohs (2001) stated in their review, that "By good, we understand desirable, beneficial, or pleasant outcomes including states or consequences. Bad is the opposite: undesirable, harmful, or unpleasant." (p. 325). As the authors openly state, this definition explains one concept in terms of other concepts.

For our definition, we take a Lewinian approach that is rooted in an interaction of organisms with their environment (Lewin, 1943): "Good" and "bad" are evaluation outcomes, and evaluations are based on the fit of informational input from the environment with the goals and needs of the organism. Good is therefore the organism's evaluation of input that serves its goals and needs, while bad is the evaluation of input that hurts its goals and needs. For example, a cold drink might be desirable, beneficial, or pleasant for a warm evening on the porch, but undesirable, unpleasant, or even harmful for a freezing night on a skiing cabin.

This definition has several implications. First, without an evaluating organism, there is no "good" or "bad." These categories necessitate people to evaluate the environment. The environment may offer a number of stimuli, events, or situations, but without the evaluating organism, these have no evaluative connotation. People may experience a warm day as pleasant, or a cold day as unpleasant, but without people, there are no pleasant or unpleasant days, but only days with an average temperature that might be 20 °F or 80 °F. The evaluations of pleasant and unpleasant, good or bad, reside on the

side of the organism. Second, similar to our example in the previous paragraph, when "good" and "bad" do depend on the organisms' goals and needs, the same environmental input might be good for one person, but bad for another person. Assuming that one person is on a skiing vacation, 20° might be perfect and very pleasant, while the same temperature would be less pleasant for a hiking vacation, or even a beach vacation. A piece of bread with butter might be wonderful after a day without food, but less appealing after an extensive Thanksgiving dinner. Third, the definition implies that there is a reality or substance that exists independently of the organism. This notion is uncontested in many scientific disciplines, (physics, biology, chemistry, or law), but less clear in psychological research (Leising, Scherbaum, Locke, & Zimmermann, 2015). We subscribe to the distinction between a factual reality and the organism's evaluation of this reality and we will come back to this aspect of our definition later.

This position is less relativistic or constructivistic than it might appear. First, there are goals and needs that are shared among all people. Maslow (1943) for example postulated biological needs, safety, belongingness, self-esteem, and self-actualization as fundamental needs that are shared among all people. By that token, food stimuli (biological needs), shelter (safety), spouses (belongingness), and praise (self-esteem) should represent positive information for all people. Conversely, foul smells or insults should represent negative information for all people, as they are at odds with biological needs or self-esteem needs (see also Kenrick, Griskevicius, Neuberg, & Schaller, 2010). Second, goals and needs do not have to be represented in consciousness or working memory. Most if not all theories of motivation allow for unconscious goals in the sense that people might not be able to verbalize them. Survival and pro-creation are probably the prime examples of goals that most people do not mention when they explain their behaviors, but which influence behavior to some extent (Dawkins, 1976; Kenrick et al., 2010). Most of the stimuli and the information that has been used to research differential processing of positive and negative information falls into the class of information that is positive/negative for everybody or at least positive/negative for most people.

2. Processing advantages and disadvantages of positive and negative information

A lot of research on valence asymmetries is guided by the idea that "bad is stronger than good," put forward by Baumeister et al. (2001), and

the general notion of a "negativity bias" (Kanouse & Hanson, 1972). Negative information seems to enjoy processing advantages. This idea is intuitively appealing and easy to communicate. At the same time, as we will show below, there are also substantial processing advantages for positive information, or a "positivity bias" (e.g., Matlin & Stang, 1978).

However, similar to our differentiated view of what constitutes "good" and "bad," or rather, positive and negative information, we believe the notion of "biases" or "advantages" are problematic starting points (see Corns, 2018). It is not a priori clear what constitutes an advantage or disadvantage in a given situation. For an example, Bohner, Bless, Schwarz, and Strack (1988) showed that negative information elicits more attributional thinking. The label "more attributional thinking" fits well with the notion of a "negativity bias" or "negativity advantage." However, whether it is factually a bias or an advantage or disadvantage depends fully on the judgmental contexts and the applied normative standard from which the "bias" might deviate. In some situations, when time is no issue and a full attributional analysis is possible, it might be advantageous to invest effort into attributional thinking, that is, a causal analysis. When time is short and information limited, it might be disadvantageous to spend time and effort on attributional thinking.

We do not contest that there might be unconditional advantages and disadvantages. For example, it would be difficult to argue that slower word comprehension is better than faster word comprehension, or that poor memory performance is better than good memory performance (but see Taylor & Brown, 1988; who argued that it is beneficial to forget negative events). However, for many valence asymmetries, such as the attributional thinking example, it is not unconditionally clear what may constitute and advantage and what may constitute a disadvantage; in the same way, it is not clear how the "unbiased" standard is construed. The situation is complicated by the relative nature of the effects (see Unkelbach, 2012). Most empirical work compares positive and negative, but rarely positive and neutral information, or negative and neutral information. Thus, it is often unclear if an observed difference is due to advantages for positive or negative information, or disadvantages for negative or positive information, or what the "unbiased" or correct standard would be.

We prefer the term "differential processing" of positive and negative information. It highlights that valence asymmetries mainly refer to relative effects; that is, they describe effect in the differential attending to, encoding of, storing of, and retrieval of positive compared to negative information.

Below, we nevertheless discuss "advantages" and sometimes "biases" to facilitate communication. Yet, it is important to keep in mind that these interpretations are not absolute, and only hold within the context of information of opposing valence; that is, good relative to bad, or vice versa.

3. What to expect?

We first review processing advantages for negative information with classic examples, and if necessary, we contextualize these examples with more recent evidence. These examples include attention, memory, person perception and impression formation, as well as attributional thinking. We then review examples of processing advantages for positive information. These include processing speed, associative potential, congruency, as well as attributional thinking.

Next, we review explanations for the differential processing and evaluate them against the background of the presented examples. In particular, we focus on whether the explanations also allow positivity advantages. Another distinction between explanations is if they focus on the "how" question (i.e., addressing cognitive mechanisms; e.g., negative stimuli are more unexpected; unexpected information draws attention and therefore, negative stimuli draw attention) or the "why" question (i.e., addressing the origins; e.g., negative stimuli may kill the organism; not attending to negatives therefore increases lethality risks and genetically fosters attention to negative stimuli).

Finally, we present the differential similarity of positive and negative information as a potential unifying explanation. Based on this explanation, we delineate answers to both the "why" and the "how" questions, we present empirical evidence that positive information is more alike compared to negative information, and we show how this explanation leads to novel hypotheses, allows quantitative predictions beyond mere good-bad main effects, and solves old puzzles present in the literature.

If not indicated otherwise, anything we report as a difference or a correlation was reported to be at least significant at a standard alpha error level of $P < 0.05$ in the original publication.

4. Advantages for negative information

We will first review advantages for negative information in terms of attention, memory, person perception and impression formation, and attribution.

4.1 Attention

People seem to attend more to negative information. This is one of the most prominent findings across many fields, and we will present three examples.

4.1.1 Attention I: The attention-grabbing power of negative information

Pratto and John (1991) hypothesized that people attend more to negative social information. They used a modified Stroop (1935) paradigm and presented personality traits in colors of blue, green, gold, pink, and red. Participants task was to name the color and to disregard the word (i.e., the personality trait). The typical Stroop paradigm uses words that present colors themselves (e.g., the word "red" written in blue color). As people cannot avoid reading the word, the secondary feature of the task (i.e., the word meaning "red") interferes with the primary task to name the color (i.e., "blue"). The interference of the task-irrelevant information with the task-relevant information is the Stroop effect. Pratto and John suggested that negative personality traits would more strongly interfere with the color-naming task, thereby supporting the attention-grabbing power of the negative information. Their first experiment found that participants took on average 650 ms to name the color of positive personality traits, while they took on average 679 ms to name the color of negative personality traits. Their second experiment tested whether attention is responsible for this latency difference. If so, then participants should also incidentally learn and remember negative traits better. The second experiment replicated the latency pattern: participants took 601 ms to name the color of positive, but 612 ms to name the color of positive personality traits. In addition, in a surprise free recall test, participants remembered on average 2.6 negative traits, but only 1.3 positive traits. Their third experiment precluded alternative explanations in terms of frequency. Negative traits might be less frequent, and thereby more unexpected. To test whether word frequency accounts for the latency difference, they used a set of frequent and infrequent traits, with frequency and valence being orthogonal. Similar to the first two experiments, participants took only 656 ms to name the color of positive, but 667 ms to name the color of the negative personality traits. Frequency of the personality traits showed no significant effects on the color naming latencies.

Across three experiments, participants' longer latencies to name the color of negative personality traits compared to positive personality traits suggests that negative information interferes more strongly with the primary color naming task. The recall data supports the attention advantage for negative information.

While this is one of the most frequently cited examples of negative information's influence on attention, the picture is more complex. For example, Harris and Pashler (2004) investigated in three experiments what they called the influence of "high-priority" information on attention; concretely, words with negative emotional meanings and participants' own name. Instead of a Stroop paradigm, they used a number parity judgment task. In this task, participants see two single digits on the left and the right side of the screen. Their primary task is to decide whether the digits match or not (Wolford & Morrison, 1980). Between the two digits, however, the stimulus of interest is presented; for example, a negative or positive word. The main DV is participants' latency to judge the digits. If the stimulus between the digits grabs attention, one would expect slower latencies on these trials.

In Experiment 2, Harris and Pashler (2004) compared response latencies between trials with negative words and trials with neutral words. Critically, the experiment featured two blocks of 50 trials each. In the first block, a randomly selected negative word appeared only twice, at position 30 and 40; in the second block, negative words appeared in half of the trials. At position 30 in the first block, participants responded much slower (1470 ms) compared to the preceding 10 trials ($M = 1258$) or the following 10 trials ($M = 1241$). However, at position 40, no such difference was observed, and in the second block, latencies for neutral ($M = 1058$) and negative words ($M = 1058$) were virtually identical. These authors attributed the latency difference at position 30 to an initial surprise reaction, and not the negative valence of the words. In addition, they observed the same effect when the high-priority stimulus was participants' own name.

This pattern was conceptually replicated by Aquino and Arnell (2007). They used the same task, but across two blocks with 100 trials each, they presented words related to threat, sex, school, as well as neutral words (25 words per category). They only observed a differential increase in latencies for the sex-related words, but not for the threat words, and this difference was mainly visible in the first block, but substantially reduced in the second block. The authors hypothesized that the underlying mechanism might be arousal, and not valence (see also Vogt, De Houwer, Koster, Van Damme, & Crombez, 2008; for a similar argument). Participants also rated their 100 stimuli on valence and arousal. As expected, on two scales from 1 (most negative/least arousing) to 7 (most positive/most arousing), participants rated threat words as most negative ($M = 2.04$) compared to all other categories. Sexual words ($M = 4.32$), school words ($M = 4.14$), and neutral words ($M = 4.37$) did not differ. However, participants rated sexual words more

arousing ($M = 5.18$) compared to threat words ($M = 4.44$), while school and neutral words were rated much lower on arousal ($M = 2.34$ and $M = 2.31$, respectively). In addition, on the stimulus level, arousal ratings significantly predicted latencies ($r(98) = 0.38$), while valence ratings did not ($r(98) = 0.15$).

From these studies, one may conclude that positive stimuli (i.e., people's own names or sex-related words) also grab attention, and it might not be valence per se, but the surprise value of the negative words, or the high arousal associated with these. In any case, one of the seemingly best-established negativity advantages is much less clear than it first appeared.

4.1.2 Attention II: The popping-out of negative information

Another frequently cited paradigm to support the attentional advantages of negative information is the face-in-the-crowd paradigm (Hansen & Hansen, 1988). In this visual search task participants have to detect faces (e.g., an emotional face, or a deviant face) within a configuration of many (sometimes few) other faces. Hansen and Hansen reported three experiments showing what they called an *anger-superiority-effect* and assumed that negative information "pops out" of the visual field. Their first experiment used a matrix of nine (3×3) faces (i.e., the "crowd") and participants' task was to decide in 108 trials if one of the faces (i.e., the potential target) showed a different expression. In half of the trials, one face showed a different expression, in the remaining trials, there was no target face. The design varied the crowd's expression ("neutral," "happy," and "angry") and the target face's expression ("neutral," "happy," and "angry"), as well as the position of the target face within the matrix. Participants showed overall faster latencies and fewer errors for "deviant emotion present" judgments for angry compared to happy faces.

Their second experiment involved a threshold detection task. In a reduced 2×2 matrix, participants either saw an angry face among three happy faces or a happy face among three angry faces. Participants' task was to verbally locate the deviant emotional expression. Again, participants showed lower thresholds for the angry faces within a happy crowd compared to the happy faces in an angry crowd. Experiment 3 then varied the crowd size (9 vs. 4 faces) and emotional expression of crowd and target. Again, in half of the trials a target with deviant emotional expression was present within the crowd (i.e., a happy face within an angry crowd or an angry face within a happy crowd). Across all factors, participants again responded faster to the presence of angry compared to the presence of happy targets. Thus,

the authors concluded that there is an anger-superiority-effect in face processing, and that negative information "pops-out" of the visual field.

This pattern was replicated in many studies (e.g., Öhman, Lundqvist, & Esteves, 2001), however, there are also a number of mixed results depending on the employed methodologies and stimuli; in particular the popping-out assumption has been criticized (see Frischen, Eastwood, & Smilek, 2008 for a review). A particular strong argument against the "popping-out" of negative information was provided by Becker, Anderson, Mortensen, Neufeld, and Neel (2011). They argued that most of the evidence had problematic designs and problematic stimuli. For example, presenting happy faces within angry crowds may slow down responses because of the angry crowd, which should consume attentional resources. Similarly, angry faces, either of real photos or schematic representation (Öhman et al., 2001), have visual features such as the "V" shape of the brows that have search advantages in visual search paradigms that lead to faster detections independent of emotional expressions. Across seven experiments that tried to avoid such confounds, Becker and colleagues found more efficient detection of positive emotional expressions, rather than evidence for an anger superiority effect. Thus, the evidence for this negativity advantage is less clear then it initially appeared.

4.1.3 Attention III: Lower thresholds for negative information

Another paradigm used to support attentional advantages for negative information is the presentation of stimuli close to or below the perceptual threshold; the prediction is that people should have lower thresholds for negative information. This effect is already present in Hansen and Hansen's (1988) second experiment, but stronger evidence comes from a study by Nasrallah, Carmel, and Lavie (2009). In three experiments, participants' task was to judge whether shortly presented words had emotional content or not. Their first experiment used a sample of 88 positive, 88 negative, and 176 neutral words. They manipulated presentation duration (22 ms vs. 33 ms) and word valence in blocks of 44 words. That is, a block either contained only positive or only negative words. Each participant thereby completed eight experimental blocks, two per valence times two per presentation duration. Independent of presentation duration, participants showed on average greater discrimination ability for negative compared to neutral words than for positive compared to neutral words. Their second experiment replicated this pattern for 22 ms presentation duration with reduced luminance, thereby decreasing overall performance. Participants again showed a higher

discrimination ability for negative words. Finally, their third experiment aimed to control for the influence of arousal on the detection ability and also presented positive and negative words within the same blocks. After controlling for participants' idiosyncratic arousal ratings for the word stimuli, the authors found higher accuracy rates for negative compared to positive words.

Similar results were also reported by Dijksterhuis and Aarts (2003), who showed preferential detection of negative words compared to positive words. Gaillard et al. (2006) also reported lower thresholds for the correct naming of words given they had negative valence. Zeelenberg, Wagenmakers, and Rotteveel (2006) also presented participants briefly with positive, negative, or neutral words. After presentation, participants had to choose between a new word and the presented word, guessing which they had just seen. Different from the presented findings, they reported a general advantage for both negative and positive words, but no difference between positive and negative words. Contradicting the lower threshold for negative information, Snodgrass and Harring (2004) found an advantage for positive words in terms of participants' discrimination ability when participants had to judge whether a stimulus was a word or not a word. Again, the initially clear pattern is more complex than expected.

4.2 Memory

People seem to recognize and remember negative information better. We already reviewed some examples of this negativity advantage within the attention section, as better memory should follow from higher attention. Here, we will present two further examples.

4.2.1 Memory I: Discrimination ability and response bias in recognition memory

Ortony, Turner, and Antos (1983) were among the first to report a differential effect of stimulus valence on recognition memory, in particular, on signal detection (SDT) measures. SDT measures allow estimating participants' discrimination ability d' (i.e., how well they are able to discriminate old items from new items) and their response threshold β (i.e., how often do they classify items as "old"). Theoretically, these two measures are independent (see Stanislaw & Todorov, 1999). They presented participants with 80 sentences; half of the sentences had an overall positive emotional tone, and half of the sentences had an overall negative emotional tone. Participants read the sentence and classified them as "positive" or

"negative." After the presentation phase, participants faced a surprise recognition test and classified 80 sentences using a scale from 1 to 6 as "old" (1–3) or "new" (4–6). Participants discriminated better between old and new statements that had a negative emotional tone ($d' = 3.37$) compared to statements with a positive emotional tone ($d' = 2.66$). Participants also had a lower response threshold for statements with a positive emotional tone ($\beta = 1.04$) compared to statements with a negative emotional tone ($\beta = 1.21$).

Ortony et al. (1983) labeled this differential pattern for discrimination ability and response threshold a puzzle: "This result is puzzling because there seems to be no reason to expect such a difference, and it is interesting because it suggests that affective aspects of stimuli interfere with what are usually considered to be relatively "cold" recognition mechanisms." (p. 725).

The pattern found by Ortony et al. (1983) has been replicated several times. Robinson-Riegler and Winton (1996) replicated the recognition advantage for 48 positive emotion terms and 48 negative emotion terms. However, instead of signal detection parameter, they estimated the contributions of "recollection" and "familiarity" to the recognition judgments, based on Jacoby's (1991) process-dissociation procedure. While not identical, the implications are similar for recognition judgments, as recollection is akin to discrimination ability and familiarity is akin to a lower threshold for calling items "old." Familiarity contributed stronger to positive emotion words' probability to be judged "old" (i.e., implying a lower threshold), while recollection contributed stronger to negative emotion words' probability to be judged "old" (i.e., implying better discrimination ability).

Ohira, Winton, and Oyama (1998) replicated this pattern for a sample of female Japanese college students with the 96 emotion words presented in Katakana, a Japanese syllabary. A slightly different pattern was found by Inaba, Nomura, and Ohira (2005), who used 38 positive and 38 negative emotion words, with 76 neutral words for baseline. They only replicated the discrimination advantage for negative words, but no response bias difference for positive words. In addition, they also reported a difference in EEG recordings. We are not aware of contradicting evidence, in particular for the differentiated pattern of discrimination ability and response bias/threshold for positive and negative information.

4.2.2 Memory II: Better free recall

Beyond recognition, people also seem to recall negative information better. For most examples, this hypothesis seems a priori true. It is difficult to

imagine that one would remember a hug better compared to a slap in the face. The asymmetry in these cases is so strong that many memory researchers are not even considering it as a variable. For example, in their paper on "flashbulb" memories, Brown and Kulik (1977) only used memories for the deaths or attempted assassinations of public figures or a "personal, unexpected shock, such as death…" (p. 79, table 1). They did not include potentially positive events that may lead to "flashbulb" memories.

To illustrate this strength, we will use Skowronski and Carlston's (1987) seminal paper on cue diagnosticity, which we will address in more detail later. This is a particularly interesting example, as the authors did not aim for a main effect of negative vs. positive information. They argued that in the ability domain, positive information is more diagnostic, and should have more impact. In the morality domain, negative information is more diagnostic and should have more impact. Their first study found exactly this predicted pattern. In their second study, they extended this prediction to free recall data. Participants read booklets that described target persons with two cues, and the cues either related to morality or ability. The cues also varied orthogonally on five behavior levels from extremely positive to extremely negative, resulting in 25 target descriptions. After each description, participants evaluated the targets on a scale of "extremely dishonest (stupid)" to "extremely honest (intelligent)," depending on the trait domain. After 4 min of filler, participants wrote down as many behaviors as possible within 5 min.

The target ratings replicated Study 1's interaction of ability and morality. In the morality domain, negative behaviors had more impact on trait ratings, but in ability domain, positive behaviors had more impact. However, the free recall data did not follow predictions. Although there was a significant cue valence (positive vs. negative) by domain (ability vs. morality) interaction, this interaction was comparably weak ($\eta_p^2 = 0.044$, 90% CI[0.010; 0.074]), and did not follow a clear patter, while the valence main effect was substantially larger ($\eta_p^2 = 0.191$, 90% CI[0.129; 0.242]). Thus, the valence asymmetry for memory is even apparent in studies that did not aim to find it.

There is substantial evidence that emotional connotated information, both positive and negative, enjoys memory advantages compared to neutral information (see Kensinger, 2009); and in most cases, this advantage is greater for negative compared to positive information. However, there are also counter-examples. Matlin and Stang (1978) reviewed in their book

a large number of studies, and reported a recall advantage for positive information. More recently, especially with older participants, there are studies showing disadvantages for negative information (Charles, Mather, & Carstensen, 2003). Hess, Popham, and Growney (2017) suggested that negative information's higher arousal might account for the memory differences between young and old people. With younger people, they found a recall advantage for negative information, which was moderated by arousal. The difference was substantially reduced for older people. Thus, albeit most prominent examples of free recall suggest advantages for negative information, such as disasters or catastrophes, this might not be a main effect and moderated by the higher arousal associated with negative information. Again, the full picture is more complex than a categorical main effect between positive and negative information.

4.3 Person perception and impression formation

People seem to assign more weight to negative information in person perception and impression formation. Within social psychology, these areas are the most prominent examples of negative information's greater impact. Impression formation research tries to specify rules by which people integrate single pieces of information into coherent impressions. In most investigations, negative information contributes more strongly to impression formation as expected from simple adding or averaging rules (Kanouse & Hanson, 1972; for an early review). There are several aspects of this greater impact, of which we discuss two in the following.

4.3.1 Forming impressions from traits

The classic example for negative information's stronger impact is Anderson's (1965) study on averaging and adding rules in impression formation. Anderson presented participants with two or four traits and participants rated how much they would like a person characterized by these traits. The traits were either highly positive, mildly positive, mildly negative, or highly negative. These classifications were based on pre-ratings on a list of 555 trait adjectives, and Anderson selected the traits to be equidistant from the scale's neutral point based on these pre-ratings. Participants always observed two or four traits from these classes (i.e., two highly positive, two mildly positive, and so forth) and rated the respective target person characterized by these traits on a scale from 0 to 100. Anderson observed that especially the negative traits led to more extreme judgments on likeability; positive traits led to smaller deviations from the mid-points compared to negative traits.

Similarly, in a study by Feldman (1966), cited after Kanouse and Hanson (1972), participants rated a target person characterized by a single trait, sampled from 12 positive and 13 negative traits, on a scale from "bad" to "good." Then, participants rated the target persons characterized by two traits; the original trait and all combinations with the 24 other traits. The dependent variable was the change in rating due to the presence of the second trait. For example, participants rated a "wise" person, and then a "wise and clean" person, a "wise and dirty" person and so forth. The ratings showed that negative traits changed the impression of "good" and "bad" much more than the positive trait. Thus, negative traits seem to have a stronger impact on impressions based on personality traits.

4.3.2 Forming impressions from behaviors

Another seminal data set for the negative information's higher impact study was reported by Fiske (1980). In her study, participants rated the likeability of targets engaged in various activities. Behaviors varied on three dimensions: extremity (mild or extreme behaviors), valence (positive or negative behaviors), and behavior dimension (sociability or activism). Participants rated the likeability of 16 targets that varied on these factors. Participants' target impressions were strongly influenced by valence; that is, negative behaviors had more impact on participants' likeability ratings. In addition, as one may expect, extreme behaviors (both positive and negative) had more impact on participants' likeability judgments. Fiske conclude that both negativity and extremity contribute to the "informativeness" of behaviors for likeability.

Probably the most important qualification of this negativity advantage in person perception and impression formation was introduced in the already mentioned study by Skowronski and Carlston (1987). As discussed above, the authors suggested that it is not valence per se, but the diagnosticity of this information for a given impression formation task. This notion is also present in the study by Fiske (1980). Diagnosticity predicts greater impact of positive traits and behaviors when these are diagnostic. To illustrate diagnosticity, Skowronski and Carlston used the dimensions of morality (e.g., honest-dishonest) or ability (e.g., intelligent-stupid). For example, an honest person (i.e., morality domain) should not lie while a dishonest person may also tell the truth from time to time. Conversely, an intelligent person may behave stupidly from time to time, but a stupid person should never behave in an intelligent way. Two studies supported this idea. In both studies, participants rated targets that varied on honesty (varying on five

levels from extremely honest to extremely dishonest) or targets that varied on intelligence (varying on five levels from extremely intelligent to extremely stupid). The task was how likely a target characterized either by a trait (e.g., "extremely intelligent") would show a certain behavior (e.g., "understood the equations presented in calculus class" or "can't remember to tie his own shoelaces"). For trait-inconsistent behaviors, the pattern was as predicted. Participants expected that stupid people will not solve calculus equations, but intelligent people may forget how to tie their shows. Conversely, honest people will not write bad checks, but dishonest people may report all their taxable income to the IRS.[a] For trait-consistent behaviors, there was a clear advantage for positive behaviors, both for honesty and intelligence. That is, participants predicted overall more positive compared to negative behaviors. The authors also computed a cue-diagnosticity score, which indicated that for the morality/honesty dimension, negative behaviors are more diagnostic, while for the ability/intelligence dimension, positive behaviors are more diagnostic. From this context, one may conclude there is no unqualified negativity advantage in impression formation (Skowronski & Carlston, 1989). Rather, people use trait diagnosticity for a given behavioral prediction, and conversely, make stronger predictions from diagnostic behaviors.

4.4 Attribution

People seem to show more causal reasoning based on negative information; that is, people search more for the cause of a negative outcome. Here, we present examples of the differential attribution processes resulting from personal successes and failures, and attributing intentionality for positive and negative outcomes of another person.

4.4.1 Attributions after success and failure

In a study by Bohner et al. (1988), participants performed a fake "professional skills test" by trying to solve 10 items of the Raven intelligence test. Success was defined as solving at least seven items. Participants were assigned to four conditions resulting from the combination of success and failure feedback with the expectation of success and failure. Half of the participants learned that only 23% of their peers passed the test, or that 77% passed the test, and orthogonally, half were told that they succeeded and half that they failed. The dependent variables were participants open-ended

[a] Examples are taken from table 1 in Skowronski and Carlston (1987, p. 692).

answers to a question about the testing situation, and their specific test result; in addition, participants directly rated the intensity of their causal reasoning. Participants' responses were content-analyzed. On all variables, the authors found more causal reasoning after failure feedback compared to success feedback. Importantly, they found no main effect of expectancy or an interaction of the failure-success main effect with expectancy. Thus, with regards to personal successes and failure, participants generated more reasons and thought more intensely about negative outcomes compared to positive outcomes.

4.4.2 Attributions of intent after positive and negative consequences

People seem to attribute more intention to actors when actions have negative rather than positive consequences. This phenomenon is also known as the "Knobe"-effect. Knobe (2003) presented his participants (visitors in a public park in Manhattan) with a "harm" or a "help" vignette. In the harm vignette, the chairman of a board approved a program that would increase profits, but also harm the environment. In the help vignette, the chairman approved a program that would increase profits, and also help the environment. Participants rated on a scale from 0 to 6 whether they thought the chairman harmed (respectively helped) the environment intentionally. Splitting the scale, 82% of participants attributed intentionality to the chairman in the harm condition, but only 23% in the help conditions. Knobe also reported a replication of this pattern with another scenario and concluded that there are asymmetries in assigning praise or blame based on positive or negative outcomes.

This pattern has been replicated in many ways (see Feltz, 2007). To the best of our knowledge, there is little counter-evidence for the assumption that negative information elicits more search for reasons compared to positive information (see reviews by Weiner, 1985, 1986). However, as we shall see below, the search for causality has also another side, namely that positive information elicits more attributions to different causes. We will discuss this case under "positivity advantages."

4.5 Summary on negativity advantages in information processing

We reported classic examples for negativity advantages in various stages of information processing; from attention, memory, person perception and impression formation, to attribution. Table 1 summarizes these examples. We also contextualized the classic evidence with newer data sets. The

Table 1 Summary of examples for negative information's advantages in information processing.

Observed advantage	Classic example	References
Attention		
Attention grabbing	Stronger interference of negative compared to positive traits for color naming tasks	Pratto and John (1991); but see Harris and Pashler (2004)
Popping-out	Faster detection of angry compared to happy faces in a crowd of faces	Hansen and Hansen (1988); but see Becker et al. (2011)
Detection thresholds	Higher detection rates and discrimination ability at very short presentation times for negative compared to positive stimuli	Nasrallah et al. (2009); but see Snodgrass and Harring (2004)
Memory		
Recognition	Better recognition of stimuli with negative compared to positive valence	Ortony et al. (1983) and Robinson-Riegler and Winton (1996)
Free recall	Better recall of stimuli with negative compared to positive valence	Brown and Kulik (1977) but see Matlin and Stang (1978)
Impression formation		
Impressions from traits	Stronger influence of negative compared to positive traits on likeability evaluations	Anderson (1965) and Feldman (1966)
Impressions from behaviors	Stronger influence of negative compared to positive behaviors on respective trait evaluations	Fiske (1980) but see Skowronski and Carlston (1987)
Attribution		
Achievements	Stronger causal inferences after failures compared to success	Bohner et al. (1988) and Weiner (1985)
Intentionality	Stronger intentionality inferences after negative outcomes compared to positive outcomes	Knobe (2003) and Feltz (2007)

Note: The respective sections on these effects also discuss evidence that is not in line with the typically observed advantages. These studies are given in the "references" column under the "but see" label.

present contexts illustrate that some of the classic examples are not well explained in terms of simple main effects of negative information. Results on the processing of negative information are quite diverse, and at some places, unstable.

This overview aimed to be illustrative, rather than exhaustive or comprehensive, though. For example, we omitted differential valence effects in behavioristic learning (see Öhman & Mineka, 2001), as these are not necessarily processing effects. To be sure, there are many other processing differences that one may conceptualize as negativity advantages that we did not address here. For example, people seem to believe statistical claims more when these are framed negatively (Hilbig, 2009; but see Unkelbach, Bayer, Alves, Koch, & Stahl, 2011; Unkelbach & Rom, 2017), negative examples of a stimulus class generalize more to similar stimuli compared to positive examples (Fazio, Eiser, & Shook, 2004; see also Fazio, Pietri, Rocklage, & Shook, 2015), and negative information is "stickier" than positive information in serial evaluations (Sparks & Ledgerwood, 2017). There is also EEG—evidence that in good-bad categorizations, negative pictures elicit stronger brain responses compared to categorizing positive pictures (Ito, Larsen, Smith, & Cacioppo, 1998).

In addition, summaries with broader foci beyond processing effects have been presented elsewhere (Baumeister et al., 2001; Rozin & Royzman, 2001). Our examples, though, illustrate valence asymmetries in information processing.

5. Advantages for positive information

Previous summaries by and large omitted systematic evidence for what one may construe as positivity advantages in differential processing of good and bad.[b] There are several potential reasons why positivity advantages are the step-child when it comes to valence asymmetries. First, one may argue that this absence reflects the state of the world. Positivity advantages may be weaker or may simply not exist. Second, given our review of the negativity advantages in attention, this may also reflect a meta-phenomenon in science. Negativity advantages draw and hold attention, while positivity advantages go unnoticed. Third, difference in processing of positive and negative information may be easier to construe as "negativity biases" compared to

[b] Taylor (1991), Peeters (1971), as well as Rozin and Royzman (2001, p. 313) briefly discuss positivity advantages.

potential "positivity biases." Fourth, and finally, there is a challenge that arises from presenting both advantages for positive and advantages for negative information. The search for an explanation becomes more difficult if one considers processing advantages of positive information as well. Explanation in terms of unconditional negativity advantages need additional assumptions and subclasses, taking away some of their appealing simplicity. In the following, we provide examples of such positivity advantages.

5.1 Processing speed

People seem to classify positive information faster as "good." This advantage is latently present in almost every data set in which participants classified information as "good" or "bad" and response latencies are measured, but also in word/non-word classification. We will present two examples of both cases.

5.1.1 Processing speed I: Faster "good"-"bad" classifications

There are many examples of faster "good"-"bad" classifications, probably the most extensive one is present in the English Lexicon Project (Balota et al., 2007). However, we will focus on one of the most widely used sets of evaluative words within social psychological research, namely the 92 stimulus words used by Fazio, Sanbonmatsu, Powell, and Kardes (1986, Experiment 2), which they used to investigate the automatic activation of attitudes. Bargh, Chaiken, Govender, and Pratto (1992) investigated the generality of the automatic activation effect hypothesized by Fazio and colleagues; to do so, they collected norm data on these 92 stimuli, including evaluative ratings and classification latencies. Participants rated the stimulus words on an 11-point scale from −5 (extremely bad) to +5 and 3–6 weeks later, they classified the stimuli using a response box as "good" or "bad." These two variables correlated negatively on the stimulus level, $r(92) = -0.38$, indicating that positive stimuli elicited faster responses. Bargh and colleagues also noted this correlation, but the fact was assigned little importance: "Finally, positive evaluations were made more quickly than negative evaluations, but this may have been due merely to a greater readiness to respond *good* rather than *bad*." (p. 897).

The faster classification of words as good compared to bad classifications was replicated for words translated into German by Klauer and Musch (1999), who provided norm data on these 92 stimulus words for a German sample. In their data, evaluations and latencies also correlated

negatively, $r(92) = -0.32$. Thus, both for an American sample and a German sample, positive stimuli elicited faster responses.

5.1.2 Processing speed II: Faster lexical decisions and word identifications

Unkelbach et al. (2010; Study 2) also used the stimulus set by Fazio et al. (1986); however, different from Bargh et al. (1992), participants made lexical decision; that is, they decided whether a string of symbols represented a word or not. They presented 90 word trials and 90 pronounceable non-word trials; the non-words matched the words in length. Across the 90 selected stimulus words, valence and lexical decision times again correlated negatively, $r(90) = -0.36$. In addition, they presented the words for 33 ms followed immediately by a mask of non-alphanumeric symbols. Participants were then asked to type in the word they believed to have seen. If the typed word matched the presented word, it was counted as a correct identification. Across the 90 words, correct identification and evaluation correlated positively, $r(90) = 0.43$. Thus, positive words were more frequently identified correctly at very short presentations.

A possible alternative explanation for these effects is higher frequency of positive words within written and spoken language (Boucher & Osgood, 1969; Matlin & Stang, 1978; Unkelbach, Koch, & Alves, 2019, for overviews). This argument is akin to the explanation of negativity effects in terms of extremity, rarity, or unexpectedness. Unkelbach et al. (2010; Study 3) addressed this by selecting 33 positive and 33 negative words from the Affective Norms for English Words stimulus set (Bradley & Lang, 1999). These words were matched in pairs with regards to frequency and equidistant with regards to the evaluation scale's neutral valence point. Thus, valence and frequency were now quasi-manipulated by stimulus-selection. Across this new set of words, valence still correlated with both measures of processing speed; positive words had higher correct identification rates, $r(66) = 0.25$, and led to faster lexical decisions, $r(66) = -0.28$.[c]

The relative faster processing of positive information is apparent in almost every data set that reports response latencies of positive and negative information. We are not aware of any counterexamples. The effect is open to alternative interpretations, though, in particular in terms of other stimulus

[c] These correlations are not in the published article, because they were deemed redundant by reviewers. The published article reports standardized regression coefficients. We report correlations here to increase comparability.

variables such as frequency of occurrence. We will address this point later when we discuss various explanations for processing asymmetries.

5.2 Associative potential

People seem to have more associations with positive information. One might say that positive information has more associative potential compared to negative information. This is a clear finding mainly in research on mood effects on creativity (for quantitative reviews, see Baas, De Dreu, & Nijstad, 2008; Davis, 2009). One would need to conceptualize mood as the input variable to fit these findings within the present framework of informational input. However, positive information's higher associative potential is also present outside of mood and creativity research.

5.2.1 Associative potential I: Learning associations

In a set of 10 experiments, Anisfeld and Lambert (1966) paired 12 pleasant and 12 unpleasant words with neutral stimuli, such as neutral words, numbers, or nonsense syllables. They matched pleasant and unpleasant words on relevant variables such as content and frequency. For example, participants saw the pairing of "bax-faith" and "bax-devil" (i.e., providing the same nonsense word for both positive and negative words) or "dyg-faith" and "gus-devil" (i.e., providing different nonsense words for both positive and negative words). They varied the dependent variable; in some experiments, participants saw the nonsense word and had to provide both associated pleasant and unpleasant words (i.e., two possible responses for "bax"). In other experiments, participants received all 24 word stimuli and had to match them with the neutral stimuli. For a last measure, participants received only one stimulus and had to generate the other.

Across these 10 experiments, 8 yielded significantly higher associative potential for positive compared to negative words. Anisfeld and Lambert (1966) also provided the number of participants who acquired more associations with pleasant compared to unpleasant words, and only a single experiment yielded more participants who acquired more associations with unpleasant words. In addition, Anisfeld and Lambert (1966) also report four experiments on free recall after exposure to the 24 stimulus words, and they found no significant differences in these experiments. Thus, the effect seems to be located in the associative potential of the pleasant words relative to the unpleasant words, and not in their memorability per se.

5.2.2 Associative potential II: Measuring associations using the IAT

Since Greenwald, McGhee, and Schwartz's (1998) publication of the Implicit Association Test (IAT) as a measure of "implicit" attitudes, no other measure within the experimental social psychologist's toolbox has received more attention. The IAT compares the classification latencies of attitude objects or concepts (e.g., in a race IAT: "white faces" vs. "black faces") with evaluative attributes (i.e., "good" vs. "bad") using two response keys. The typical effect for White American participants is a preference for white faces, evident in faster classifications when in an experimental block, "white faces" and "good" share a key, and "black faces" and "bad" share another key. The comparison is the stimulus classification latency in an experimental block when "white faces" and "bad" share a key, and "black faces" and "good" share another key.

Interestingly, the IAT also shows a substantial valence asymmetry, namely a "positive association primacy." Anselmi, Vianello, and Robusto (2011) used this term to describe a valence asymmetry in IAT scores (p. 376): "We argue that if positive (rather than negative) words are categorized more quickly in the condition White-Good/Black-Bad than in the condition Black-Good/White-Bad, then they are the stimuli that mostly contribute to the IAT effect and that a positive associations primacy can be observed."

Using a multi-faceted Rasch model, Anselmi et al. (2011) estimated the contributions of positive association and negative associations for a typical race IAT, using the categories "White People," "Black People," "Good," and "Bad," Stimuli were 12 morphed faces of black and white people as well as 16 words with positive or negative meaning. They found that positive words contributed significantly more to the IAT effect compared to the negative words (i.e., differential response latencies in the two experimental blocks). In addition, with one exception, the overall classification speed of faces did not change between blocks. The authors replicated this finding with a weight IAT, using the categories "Thin People," "Fat People," "Good," and "Bad," and found that the apparent preference for thin people based on the difference between the two experimental blocks was largely due to two single positive words.

From the impact of negative information in person perception and impression formation, one would expect that associations with negative information have a greater impact on attitudes measured with the IAT. The presented evidence suggests the opposite. This is not an outlier of the specific measure or the categories. Similar results are presented with

other associative measures and other stimulus categories. For example, Sriram and Greenwald (2009) introduced the Brief Implicit Association Test (BIAT). They reported as a surprising finding that the BIAT score only had good psychometric properties and only correlated with self-reports when the focal category was "good" rather than "bad." Similarly, Bar-Anan, Nosek, and Vianello (2009) found a similar asymmetry in the Sorting Paired Feature Task (SPF), construed as a measure of association strength between concepts. Across three experiments, they found that attitude objects' association with "good" showed stronger effects on the relevant dependent variables compared to the assumed associations with "bad." Across their experiments, they also report a higher internal consistency of associations with "good" compared to associations with "bad." Finally, Sherman, Calanchini, and Hehman (2017) presented three experiments showing that intergroup biases as measured with the IAT are stronger determined by pro-ingroup (i.e., positive) associations. Overall, there seems to be consistent evidence for a positivity advantage in indirect attitude measurement.

5.3 Congruency

People seem to perceive positive information as more compatible with other positive information. We illustrate this difference within two examples; the evaluative priming paradigm and a semantic version of this paradigm.

5.3.1 Congruency I: Evaluative priming

Initially introduced to show stimuli's potential to automatically activate an attitude (Fazio et al., 1986), evaluative priming has become a prominent indirect attitude measure (Fazio, Jackson, Dunton, & Williams, 1995), and a versatile tool to investigate the structure of the cognitive system (Klauer, Teige-Mocigemba, & Spruyt, 2009). The typical priming paradigm presents an evaluative prime, for example, a face, and measures the latency to respond to a following target, typically a positive or negative stimulus. Responses should be facilitated when prime and target are evaluatively congruent (i.e., both positive or both negative) compared to when they are evaluatively incongruent (i.e., positive-negative, or negative-positive; see Herring et al., 2013, for a meta-analysis).

Given the discussion above on negative information's stronger impact on person perception and impression formation, one might expect that negative primes lead to stronger facilitation effects. Accordingly, Dijksterhuis and Aarts (2003) suggested that negative primes may have a stronger influence

within the evaluative priming paradigm. Unkelbach, Fiedler, Bayer, Stegmüller, and Danner (2008; Study 3) did a full quantitative re-analysis of the seven studies reviewed by Dijksterhuis and Aarts, including 17 experiments in total. First, they replicated the classification speed advantage reported above. Independent of primes, participants classified positive stimuli faster than negative stimuli. To estimate differential congruency, they calculated the facilitating effect of positive primes on positive targets relative to positive targets following negative primes. Similarly, they calculated the facilitating effect of negative primes on negative targets relative to negative targets following positive primes. This difference score excludes the overall speed advantage of positive targets. Over and above the target main effects, positive-positive pairs still elicited stronger facilitation compared to negative-negative pairs.

5.3.2 Congruency II: Integrative priming

A similar pattern was found by Ihmels, Freytag, Fiedler, and Alexopoulos (2016). They hypothesized that another important factor in evaluative priming is how well prime and target form semantically meaningful compounds, a feature of the pairs they termed "integrativity." For example, they predicted that the prime-target pair "shower-pleasant" should lead to a facilitation effect in classifying "pleasant" as good, but the equally congruent but not matching prime-target pair "candle-funny" should not show such facilitation. Across three experiments, they showed the predicted influence of integrativity on latencies; facilitation effects were significantly larger for pairs of high integrativity compared to pairs of low integrativity. Importantly, in every experiment, positive primes facilitated the processing of positive targets, both for pairs high and low in integrativity.

These results mirror the reported advantages of positive information within IAT research, and the congruency advantages are apparent in almost all experiments that build on the matching of evaluative materials, in particular studies on mood and memory (e.g., Isen, Shalker, Clark, & Karp, 1978). These results are again rather surprising if one assumes the generally greater impact of negative information.

5.4 Attribution

People seem to draw broader inferences from positive information. This is mainly visible in inferences from good and bad outcomes for others. People typically infer more than one cause for a success. For example, winning a race requires both talent *and* effort. Failures on the other hand are sufficiently

explained by a single cause, for example, lack of talent *or* lack of effort. We present two related examples of people's broader inferences based on positive outcomes.

5.4.1 Deliberate inferences from described emotions

Liu, Karasawa, and Weiner (1992) investigated how people explain positive and negative emotions. In three experiments, they presented participants with vignettes and asked for reasons for the actors' emotional response in these vignettes. First, they varied the described emotions reaction (e.g., happy vs. unhappy). Second, they varied the described intensity (i.e., mildly vs. extremely) to account for the fact that extreme effects lead to inferences about multiple causes (see Kelley & Michela, 1980); people believe that strong effects need strong causes (e.g., Fiedler, Freytag, & Unkelbach, 2011) or multiple causes (Kun & Weiner, 1973). Third, they varied a positive or negative outcome, and fourth, whether the event was minor or major. They also included two kinds of outcomes, namely academic success or academic failure, and finding or losing money. Finally, they varied whether the vignette addressed someone else (Tom) or the participant (you). An example would be: "Tom is extremely happy. Tom did well on his math test".

Participants then assessed the additional causal contributions on three potential explanations besides the eliciting event: dispositional (i.e., something else in the person contributed), situational (i.e., something else in the situation contributed), and mood (i.e., something about his mood contributed). The DVs thus assessed multi-causality, or how likely an additional cause is for a given emotional reaction.

Experiment 1 investigated this for happiness (i.e., happy vs. unhappy), Experiment 2 for anger, gratitude, and pride, and finally, Experiment 3 for the positive emotions excitement, pleasure, and relaxation, and the negative emotions anxiety, fear, and guilt. Across all three experiments, a very stable data pattern emerged. The strongest effect was due to intensity. Participants rated the likelihood of additional causes higher for extreme compared to mild emotional reactions. Importantly, there was also a main effect of emotion valence. Participants rated the likelihood of an additional causes higher for positive emotions compared to the negative emotions.

Thus, across experiments and emotional reactions, participants believed that positive emotions have more than one cause, while negative emotions elicited less ascription of additional causes within the person, the situation, or the person's mood.

5.4.2 Spontaneous inferences from observed emotions

In a study by Krull and Dill (1998), participants observed happy or sad targets in 10 s long silent video clips. To test for spontaneous inferences, participants received either the instructions to focus on dispositional information (i.e., "figure out if the individual has a very happy personality, a not at all happy personality, or somewhere in between"), or to focus on situational information (i.e., "figure out if the individual discusses a very happy topic, a not at all happy topic, or somewhere in between"). As a dependent variable, participants responded to three questions about the behavior ("happy or sad behavior"?), the target's personality ("happy or sad personality"?), or the situation ("happy or sad topic"?). The variable of interest was participants' response latency. Responses to the behavior question served as a baseline. Spontaneous inferences are visible if participants respond quickly to information that was not in the instructed focus (e.g., fast responses to the situation question when the personality was in the focus). The latencies supported the hypothesis that participants who observed sad targets only made the instructed inferences. Participant who observed happy targets made also inferences about the alternative cause.

Krull and Dill's (1998) second experiment replicated this pattern without instructing a situational or dispositional focus. They only instructed participants to figure out if the individual behaves happily or sadly. Afterwards, they again answered questions about the target's behavior, the target's personality, and the situation. The DV was again participants' latency to respond to these questions. Participants who observed a happy target were faster both for the dispositional question as well as for the situational question compared to participants who observed a sad target, indicating that these participants spontaneously thought more about the potential causes of the happy behavior, indicating a strong influence of positive information on causal thinking.

The presented data on positivity advantages in attributions is on the surface at odds with the reported negativity advantages. This was already noted by Liu et al. (1992). However, one may construe the findings by Liu and colleagues as an example of the strength of negative information. A negative event is sufficient to explain a given emotional reaction; positive events do not suffice to explain a given emotional reaction alone; they require additional causes. The data by Krull and Dill (1998), however, do not allow this conclusion. Their second experiment suggests that positive information triggered more search for explanations. As stated in the beginning, considering both positivity and negativity advantages creates a

challenge for simple explanations of the differential processing of positive and negative information, and the case of attributional thinking illustrates this challenge.

5.5 Summary on positivity advantages in information processing

We reviewed several processing advantages of positive information across various stages of processing, from processing speed in evaluative and lexical decisions, to positive information's stronger associative potential visible in higher reliabilities and stronger effects within implicit measures, to more complex cognitive processes such as causal attribution. Table 2 summarizes these examples. We again do not claim this overview to be comprehensive, but to be illustrative.

We deliberately omitted the literature on positivity biases in self-perception (see Mezulis, Abramson, Hyde, & Hankin, 2004 for a review). Similarly, we omitted effects of the higher frequency of positive information (Boucher & Osgood, 1969; Matlin & Stang, 1978; Unkelbach et al., 2019 for a review). Rather, we focused on examples that have positive information as the input variable, and not as the outcome or intended outcome. However, there are a number of research examples that fall in this class that we omitted here. For example, Becker and Srinivasan (2014) describe "happy advantages" at early stages of face processing, leading to the more accurate classification of happy faces compared to angry, fearful, or neutral faces (Becker, Kenrick, Neuberg, Blackwell, & Smith, 2007), and they claim that features of happiness engages multiple cognitive processes early and spontaneously, which is in line with the reported research by Krull and Dill (1998), as well as the faster classification of positive information discussed above. More recently, Król and Król (2019) used eye-tracking to show that the presence of negative information had no influence on the processing of subsequent negative information, but the absence negative information (i.e., relatively positive information) facilitated processing of subsequent positive information, which is in line with the reported congruency advantages.

Different from negativity advantages, there is little counter-evidence for the presented effects. This is also apparent from Table 2. This may result from the lower interest in the literature on positivity advantages (Unkelbach, 2012); a smaller body of research may appear more stable and coherent. Nevertheless, on the surface level, results on the processing of positive information are quite similar and stable.

Table 2 Summary of examples for positive information's advantages in information processing.

Observed advantage	Classic example	References
Processing speed		
Classifications	Faster classification of stimuli as "good" compared to "bad"	Bargh et al. (1992) and Klauer and Musch (1999)
Lexical decisions	Faster classification of positive compared to negative stimuli as words or non-words	Unkelbach et al. (2010)
Associative potential		
Learning	Better learning for word pairs that contain a positive compared to a negative word	Anisfeld and Lambert (1966)
Association primacy	Stronger IAT effects for trials with positive stimuli compared to negative stimuli	Anselmi et al. (2011) and Bar-Anan et al. (2009)
Congruency		
Evaluative Priming	Stronger congruency effects of positive primes compared to negative primes	Unkelbach et al. (2008)
Integrative priming	Stronger formation of semantically meaningful compounds for positive compared to negative stimuli	Ihmels et al. (2016)
Attribution		
Deliberate inferences from emotions	More inferred causal factors for positive compared to negative emotions	Liu et al. (1992)
Spontaneous Inferences from Emotions	Attributions to both the person and the situations for positive compared to negative emotions	Krull and Dill (1998)

Note: Different from Table 1, the references do not contain other evidence.

This summary's goal was to provide examples for the kind of effects explanations of valence asymmetries must tackle. The common denominator of these findings is that they are by and large incompatible with a straightforward negativity bias or the assumed stronger impact of negative

information on cognitive processes. If negative information would simply be "stronger," than the faster processing, the higher association potential, and the broader elicited attributional processes for positive information would not follow without substantial additional assumptions.

6. Explanations

Our summaries above illustrate the challenge of simultaneously explaining positivity advantages and negativity advantages. Can lower thresholds for negative information be reconciled with faster categorization of positive information? Can negative information's greater impact in impression formation be reconciled with positive information's greater influence in attitude measures? Is there a common explanation for the differential processing of positive and negative information? Or do we have to confine ourselves to more specific explanations of specific phenomena, as pointedly addressed by Taylor (1991), who noted that it may be impossible to specify a general process that explains valence asymmetries?

Taylor (1991) might be right; yet, we will make the bold claim that at least for the area of information processing, there might be a single underlying and unifying explanation to account for the differential effects we have reviewed here. Before we present this explanation, we will first summarize existing explanations of valence asymmetries and then evaluate them against the present phenomena; in particular with regards to whether they explain *why* there should be valence asymmetries, and second, *how* these asymmetries come about.

6.1 Evolutionary pressures and phylogenetic leaning

There is a cluster of evolutionary explanations, which by and large follow from the same assumption: The observed asymmetries result from a factual asymmetry in the consequences of positive and negative events. Negative events are on average more harmful than positive events are on average beneficial (Baumeister et al., 2001; Rozin & Royzman, 2001).

In other words, overlooking the sabretooth tiger in the bushes would have resulted in death for one of our ancestors, preventing the passing of her or his genes to the next generation. Conversely, missing a potential source of nourishment or a chance of procreation would not prevented later offspring and passing of the genes. Thus, due to natural selection, mammals in general and humans in particular might have evolved a sensitivity for

negative information, in particular threat (Baumeister et al., 2001; Öhman & Mineka, 2001). Accordingly, the cognitive system has evolved to adapt to the affordances of the environment. To be precise, organisms who pay more attention to negative information, encode it more deeply, and retrieve it more easily should have a reproductive advantage compared to hypothetical organisms who pay more attention to positive information.

Phylogenetic learning explanations are intuitively plausible. Their strongest feature is the answer to the question *why* negative information might enjoy processing advantages. However, there are also drawbacks. First, phylogenetic learning does not explain apparent positivity advantages. If one attempts to explain both sets of advantages based on phylogenetic learning, predictions run the risk of becoming arbitrary; one may construe evolutionary pressure that work in one case for positivity advantages in one context, but for negativity advantages in another context. Second, the cluster does not provide an answer to the *how* question. Some models (e.g., fear-module theory by Öhman & Mineka, 2001) are highly elaborate, yet they are confined to specific cases of differential processing, and it is difficult to arrive at predictions of how the apparent negativity advantages in other areas come about.

6.2 Explanations based on correlates of valence

Another cluster of explanations focuses on correlates of negative information. These explanations assume that negative information is on average less frequent, more extreme, more intense, less expected, and more surprising compared to positive information (see Zajonc, 1968; Jones & Davis, 1965 for classic social psychological examples of this notion; and Matlin & Stang, 1978; Unkelbach et al., 2019 for reviews).

Peeters and Czapinski (1990) argued that one should control for these effects to show valence asymmetries: "If the greater impact of a negative stimulus is due to the greater intensity of that stimulus, we do not have a genuine negativity effect but simply a trivial intensity effect." (p. 34). Independent of whether one sees these as confounds or integral parts of negative information, these correlates do not have a phylogenetic basis, but follow from ontogenetic learning.

Importantly, well established cognitive principles of information processing for expected, frequent, or surprising information (Posner, 1980; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977) may account for the described differential processing. For example, differential attention,

Fig. 1 An example of the Von Restorff effect (Von Restorff, 1933). Stimuli that differ from the context enjoy processing advantages. The same way the "5" in the left panel differs from the surrounding characters, negative information "neg" may differ from the positive context of "pos."

encoding and retrieval could be all variants of the Von Restorff effect (Von Restorff, 1933) illustrated in Fig. 1. Thus, without assuming any phylogenetic learning, children may learn that negative information is rarer, and negative information thereby grabs more attention, is elaborated more, and remembered better.

This cluster of explanations provides clear answers to the *how* question. It builds on existing cognitive models and allows computational implementations. In addition, one may predict positivity advantages when positive information is less frequent, more extreme, more intense, less expected, and more surprising. A drawback is that there is no precise answer to the *why* question, which requires additional explanations. In addition, many studies that controlled for these correlates (e.g., Fiske, 1980; Pratto & John, 1991; Unkelbach et al., 2010) still found substantial processing differences.

6.3 Diagnosticity

A special case of the correlates explanations is the diagnosticity approach by Skowronski and Carlston (1989), which is to some extent already present in Fiske's (1980) concept of informativeness. As discussed above, Skowronski and Carlston attributed the differential advantages of positive and negative information in impression formation to the diagnostic value of this information. For example, Skowronski and Carlston (1987) delineated that in the morality domain, negative information is more diagnostic and predictive, while in the ability domain, the same is true for positive information. Liars will sometimes tell the truth, but an honest person should never lie. Conversely, an intelligent person will sometimes behave stupidly, but a stupid person cannot act smartly.

This explanation predicts both positivity and negativity advantages, depending on the information's diagnosticity. In addition, Skowronski and Carlston (1989) provide a computational approach for diagnosticity, which allows quantitative predictions. The drawback of the diagnosticity explanation is that it specifically explains valence asymmetries in impression formation, but fails to explain valence asymmetries outside the impression formation domain; for example, in terms of memory, congruency, or processing speed.

6.4 Mobilization and minimization

The mobilization and minimization hypothesis was presented by Taylor (1991), who started with the observation that "other things being equal, negative events appear to elicit more physiological, affective, cognitive, and behavioral activity and prompt more cognitive analysis than neutral or positive events. Negative events tax individual resources, a response that appears to be mirrored at every level of responding." (p. 67). The core of Taylor's explanation, however, is that organisms follow a homeostatic principle, and the initial strong reaction to negative stimuli is counter-acted by a minimization response towards the negative input. This explanation would for example account for the lower detection threshold of negative information, but the faster classification of positive information (i.e., because the initial strong response is counteracted).

Beyond the homeostatic argument, Taylor (1991) presents several reasons why this might be the case; for example, strategic behavior on the side of the organism such as mood repair tendencies, or protection of self-worth. The counter-reaction to negative input, up to the denial that something bad has happened, has several benefits. According to Taylor and Brown (1988), it leads to a positive view of the self, to perceived control, and optimism about the future. Despite the fact that these positive views are basically due to a motivational counter-reaction, they result in higher emotional well-being, the ability to form social bonds, and the ability to be creative and productive.

Thus, there are motivational forces within the organism that re-interpret the impact of negative information. As Taylor (p. 78) openly states, this is not a precise process-model, but rather an overall description that fits well with the minimization phase. Overall, she suggests a family of interlinked processes that both lead to strong and quick mobilization of organismic resources in response to negative events (or, more generally, to negative information), and the following downregulation by the suggested cognitive coping mechanisms.

Ultimately, the argument is again that both mobilization and minimization are adaptive; not on the phylogenetic level, but on the level of personal well-being: "Thus, a strong rapid response to negative events, coupled with a strong and rapid diminution of the impact of those events, may be most effective for the organism in both the short term and the long term." (p. 79). While Taylor's hypothesis thereby presents answers for the *why* question of differential processing, and also acknowledges and discusses several positivity advantages, it avoids the *how* questions and delegates this to potentially different mechanisms in different domains.

6.5 Top-down vs. bottom-up and positivity offset vs. negativity bias

The model by Taylor (1991) implies that internally-generated information should show a bias towards positivity (e.g., retrieval and maintenance in memory), while information from the environment shows the typical negativity biases. For example, people attend more to negative events, but balance them by remembering positive events. Peeters (1991) was among the first to notice and discuss the presence of both positivity and negativity effects, and related these effects back to the bottom-up (i.e., environment-provided) and top-down (i.e., organism-generated) types of information. He proposed that evaluations are partially controlled by the stimulus (i.e., bottom-up, or stimulus-driven), and partially controlled by subjective responses (i.e., top-down, or organism-driven; p. 135). Peeters denied both positivity and negativity biases in their simple form as unconditional principles; rather, he proposed that both biases exist as "the complementary sides of a more complex positive-negative asymmetry (PNA) of psychological functioning which has survival value for the subject." (p. 135). Simplified, one might assume that for bottom-up or stimulus information, organisms should show preferential processing of negative information. For top-down or organism-driven information, organisms should show preferential processing of positive information. Peeters provided the example of a fungus eater who lives in a world of overwhelmingly inedible or poisonous toadstools, while edible or nutritious mushrooms are scarce. The argument is that in such an environment, it is adaptive to approach all novel fungi as if they were edible (i.e., positivity effects), while it is simultaneously adaptive to avoid novel fungi at the slightest hint that they are poisonous (i.e., negativity effects).

This explanation is akin to Cacioppo and Berntson's (1994) assumption that there are two independent systems for positive and negative affect. They argue that at zero informational input, there is a tendency to approach stimuli and called this "positivity offset." Conversely, they argued for a negativity bias, namely that the motivational result will be stronger for negative stimuli per each piece of additional information. Thus, approach gradients are stronger at long distances from a positive stimulus compared to a negative stimulus, but the slope of the avoidance function is steeper than the approach function's slope. Rozin and Royzman (2001) also included this as the "gradient steepness" in their discussion of aspects of negativity bias. These explanations converge in predicting that organisms should slowly approach

all stimuli (i.e., "positivity offset"), but should show a strong avoidance reaction at the slightest hint of negative information (i.e., "negativity bias").

Evidently, Peeters' (1991) and Cacioppo and Berntson's (1994) explanations are evolutionary in nature, yet refined to allow positivity advantages in information processing. For example, assuming that congruency and association effects are organism-driven, the observed positivity advantages follow. However, the explanation does not provide answers to the *how* question of processing, and there are data points, such as the faster processing of positive information, which do not follow from the explanation suggested by Peeters (1991) without additional assumptions. Table 3 summarizes the presented explanations and their scope.

In the following, we will present the similarity explanation, which predicts both positivity and negativity advantages, and also provides answers for the *why* and *how* question of valence asymmetries in information processing.

7. The similarity explanation

The reviewed explanations laid substantial foundations for the similarity explanation. For example, many data points seem to follow the "bad is stronger than good" metaphor (Baumeister et al., 2001). However, as noted by Taylor (1991), there are clear contradictions to the underlying phylogenetic argument; for example, the apparent weak interconnections of negative information in memory: "The relative inaccessibility of negative events in memory would seem to create an evolutionary lacuna in the form of an inability to learn from past mistakes." (Taylor, 1991, p. 78). A similar observation was recently made by Eskreis-Winkler and Fishbach (2019). Most importantly, though, the strength metaphor cannot explain the apparent positivity advantages.

Further, the ontogenetic approaches in terms of correlates of negative information provide explanations for valence asymmetries that related them to well-established processing principles (see Fig. 1), and even allow for quantitative predictions. In the following, we aim to provide an explanation that borrows from both classes, and thereby aims to answer both the *why* and the *how* question. It relates the differential processing of positive and negative information to the differential intra-class similarity of evaluative information, and importantly, delineates why this differential similarity is not only a correlate or a confound, but an integral part of what distinguishes good from bad.

Table 3 Overview of classic explanations for the differential processing of positive and negative information.

Explanation	Basis	Example	Scope
Phylogenetic pressures	Negative events are on average more harmful than positive events are on average beneficial	Missing a lethal predator leads to instant death, while missing a potential mating partner does not preclude future procreation	All areas of information processing
Correlates of valence	Valence systematically covaries with variables that influence processing (e.g., frequency or extremity)	Frequent occurrences increase processing efficiency, extreme occurrences draw attention	All areas of information processing
Diagnosticity/ informativeness	Depending on context, positive or negative information is differentially diagnostic	A moral person should never lie, but a liar may behave morally; conversely, a stupid person cannot behave intelligently, but an intelligent person may behave stupidly	Person perception and impression formation
Mobilization-minimization	Strong responses to negative information are counteracted by motivational forces within the organism	Negative feedback elicits stronger emotional responses, which is then re-interpreted as not so negative after all.	Self-relevant and social information
Top-down vs. bottom-up negativity bias and positivity offset	Separate systems for positive and negative information with differential transfer functions	Any person might be a friend and can therefore be approached, but avoidance happens at the slightest hint of negative information	All areas of information processing

7.1 Differential similarity

Negative information is more diverse and consequently, negative pieces of information or negative stimuli are less alike than positive pieces of information or positive stimuli. We believe that the greater diversity of negative

information relative to positive information and the resulting similarity difference is the key explanation for valence asymmetries in processing. The greater diversity and lower coherence of negativity advantages matches the theoretical explanations (see above) as well as the empirical findings; due to its greater diversity, negative information delivers more apparent effects; yet, these are overall less consistent and stable, due to negative information's greater diversity.

The advances resulting from this explanation are threefold. First, differential diversity of positive and negative information might simultaneously explain both advantages for positive and negative information. Second, it allows quantitative predictions for these advantages within simple computational models. In addition, modeling differential similarity immediately leads to the greater "strength" of negative information, as we will show below. Finally, our approach both answers both the *why* question and the *how* question of differential processing.

We are not the first to notice that there are more varieties of negative information compared to positive information. For example, in categorical theories of emotions, there are more negative emotions compared to positive emotions. If emotions are reactions to incoming information, this implies the greater diversity of negative information. Rozin and Royzman (2001, p. 312) provided an overview that in categorical theories of emotions, there are more negative emotions. Similarity, Unkelbach et al. (2019) counted the number of specific negative and specific positive emotions in the emotion theories overview by Ortony and Turner (1990); from their list of emotion theories, the average number of positive emotions was 1.23, while the average number of negative emotions was 3.69. Liu et al. (1992, p. 603) noted that descriptions of positive emotions mainly differ in intensity (e.g., euphoria, elation, joy, satisfaction), while negative emotions differ in quality (e.g., anger, disgust, fear, guilt).

The greater variety of verbal descriptors for negative information is not restricted to emotions, but found in evaluative language in general. There are more distinct negative words compared to positive words. This asymmetry follows from the "markedness" principle described by Clark and Clark (1977). Accordingly, positive states reflect the unmarked, or normal state of the world, while negative states reflect deviations from this unmarked state. Clark and Clark provided the example of "milk." Milk is by default assumed to be "good" milk, while the marked state of "bad" milk will be reflected in communications. It is also possible to go from the unmarked state to the marked state by changing the descriptor of the unmarked state; which is

not possible the other way round. Thus, it is possible to change the evaluative descriptor "happy" to "unhappy," but it is not possible to change "sad" to "unsad." In addition, "good" milk might also become "putrid" milk, "sour" milk, and so forth. As a result, there will be more words and greater diversity in language on the negative side. This linguistic argument was empirically confirmed by Rozin, Berman, and Royzman (2010) for 20 languages.

7.2 Good is more alike than bad

Greater diversity of negative information implies higher relative similarity of positive information. To serve as a general basis for explaining valence asymmetries, one needs to ascertain the generality of a similarity difference for evaluative information. The notion that positive things are more similar is latently present both in science and arts (e.g., Tolstoi's opening to *Anna Karenina* that "All happy families are alike."). However, systematic investigations of this hypothesis less frequent and we report here our own research on the differential similarity of positive and negative information.

Unkelbach et al. (2008; Study 2) used multidimensional scaling (MDS) to assess the similarity of the 20 most positive and 20 most negative stimuli from the stimulus set used by by Fazio and colleagues (Bargh et al., 1992; Fazio et al., 1986; Klauer & Musch, 1999). In MDS studies, participants compare pairs of stimuli and rate their similarity. From the resulting similarity matrix, one can estimate an *n*-dimensional space that best fits the underlying comparison data. In this space, one can interpret distance as stimulus similarity (i.e., smaller distances imply higher similarity). Participants judged the similarity of these 40 stimuli. Fig. 2 shows the resulting three-dimensional solution. As can be seen, positive stimuli cluster more densely compared to the negative stimuli, and the average distances within the positive cluster were substantially smaller compared to the negative cluster, independent from the chosen dimensionality of the MDS solution. The authors hypothesized that this is a general effect and labeled it the "density hypothesis."

Obviously, this provides only tentative evidence, as the differential similarity might follow from potential idiosyncrasies of the stimulus set by Fazio et al. (1986). Koch, Alves, Krüger, and Unkelbach (2016) provided a more thorough investigation of the hypothesized similarity difference. First, they validated an alternative method to assess similarity, namely the spatial arrangement method proposed by Hout, Goldinger, and Ferguson (2013).

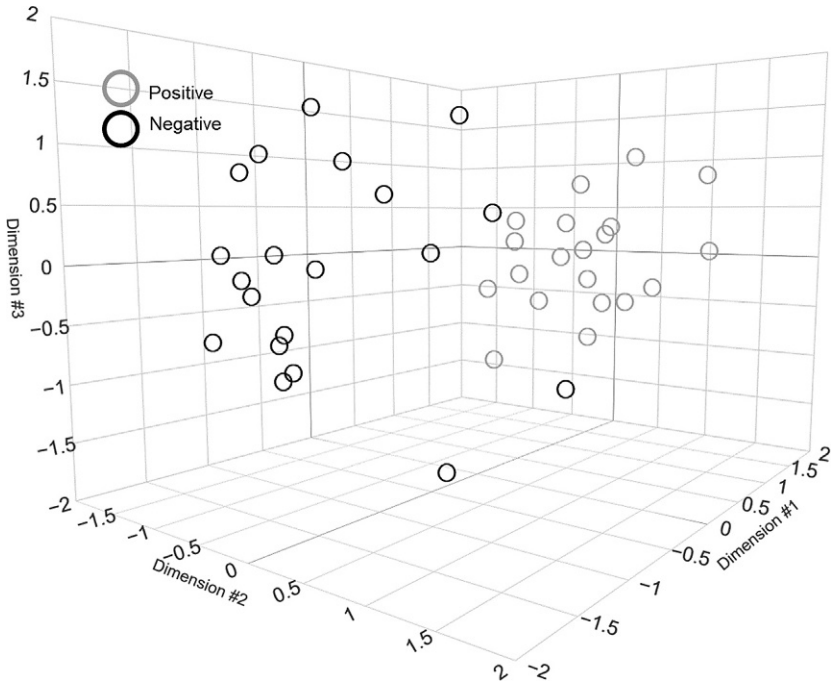

Fig. 2 A 3D plot of the 20 positive and 20 negative stimuli based on the multidimensional scaling solution reported by Unkelbach et al. (2008; Study 2).

MDS studies become time intensive for larger stimulus sets, due to the multiplicative nature of the comparisons; n stimuli require $n^*(n-1)/2$ comparisons; thus, for example, while 10 stimuli require only 45 comparisons, 100 stimuli require 4950 comparisons. Hout and colleagues therefore suggested to have participants arrange their stimuli directly on a computer screen, with the instruction that similar stimuli should be arranged together, and dissimilar stimuli apart. This provides parsimonious access to as spatial measure of similarity. Fig. 3 shows an example of this spatial arrangement method.

In Study 1, Koch and colleagues validated this method for the 40 stimuli used by Unkelbach et al. (2008). First, they replicated the general finding: Participants arranged the positive stimuli more densely together compared to the negative stimuli. Second, across the 40 stimuli, the similarity indices derived from pairwise comparisons and spatial arrangement correlated $r(38) = 0.84$, which is close to the test-retest-reliability of psychological measures. Third, in a multiple regression analysis, they predicted SpAM similarity from valence and other potential variables that might influence

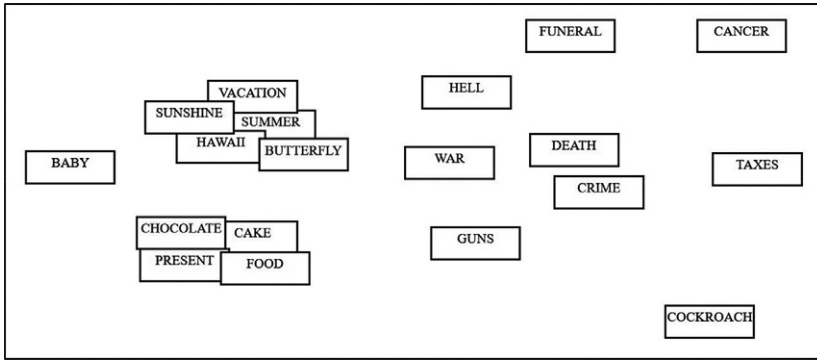

Fig. 3 An example screen from the spatial arrangement method suggested by Hout et al. (2013) and implemented by Koch, Speckmann, and Unkelbach (2020) in an easy-to-use online tool. Participants use the mouse pointer to click and move a given stimulus word on the screen with the instruction to move similar stimuli together, and dissimilar stimuli apart.

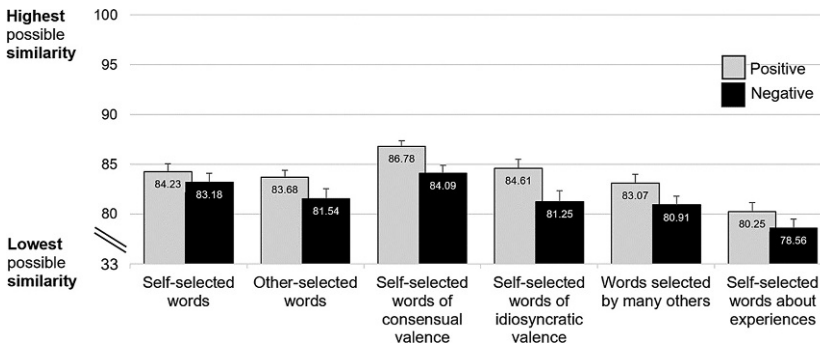

Fig. 4 An overview of Studies 2–6 reported in Koch, Alves, et al. (2016) and Koch, Imhoff, Dotsch, Unkelbach, and Alves (2016). In every study, positive information is more similar compared to negative information.

similarity, namely valence intensity, word frequency, subjective familiarity, and subjective concreteness. Only valence and valence intensity predicted similarity in the expected direction.

In five further studies, Koch, Alves, et al. (2016) investigated the generality of the finding that good is more alike than bad. Fig. 4 provides an overview of these studies. Study 2.1 (Fig. 4; "self-selected words") asked 46 online participants to generate 20 positive nouns and 20 negative nouns. Next, they spatially arranged these 40 stimuli. Overall, participants generated 1044 unique words. Across all words, participants arranged positive

words closer together compared to negative words. Study 2.2 (Fig. 4; "other-selected words") asked 46 yoked participants to arrange the words generated by people in Study 2. Again, participants arranged positive words closer together compared to negative words. Study 3 (Fig. 4; "consensual valence and idiosyncratic valence") replicated Study 2 with 110 participants. They were randomly assigned to an *idiosyncratic* and a *consensual* condition. In the idiosyncratic condition, participants generated words that they personally liked and disliked ("things that you personally find positive and negative"). In the consensual condition, participants generated words that everybody liked and disliked ("things that all people find positive and negative"). Participants then again spatially arranged the 40 words they generated. Overall, participants generated 2126 unique words, but participants in the idiosyncratic condition contributed significantly more unique words compared to the consensual condition. As Fig. 4 shows, both conditions yielded the predicted valence difference, and there was no interaction of this valence difference in similarity with instructions. In both conditions, participants arranged positive words closer together compared to negative words.

Another potential alternative is that the similarity asymmetry might not reside in the stimuli per se, but in the retrieval process. People might retrieve clusters of positive information, but single instance of negative information. As suggested by Fazio et al. (2004), positive stimuli may invite exploration, while negative stimuli are abandoned. For example, people might think of "baby," then "kitten," then "blanket," then "love," then "family," and so forth. In comparison, they might think of "death," then "garbage" (rather than "war"), then "taxes" (rather than "junk"), then "rain" (rather than "fraud"), and so forth. To address this alternative explanation and to dissociate retrieval processes from similarity estimations, Study 4 first asked 40 participants to generate a single positive and a single negative word. This yielded a pool of 29 unique positive words, and 35 unique negative words, which cannot be influenced by differential retrieval processes, as each participant generated only a single stimulus. From these word pools, 54 participants spatially arranged random samples of 20 positive and 20 negative words each. Again, as Fig. 4 (words selected by many others) shows, participants arranged positive words more closely together, indicating higher similarity.

The final study involved an event-sampling approach. Across 7 days, participants were asked in the evening (around 9 pm) to report one negative event and one positive event of the day. On the eighth day, 168 participants who responded at least on 5 out of 7 days were invited to spatially arrange the

described events. Of these, 124 completed the task. As Fig. 4 (self-selected words about experiences) shows, even for these daily life events, participants indicated higher similarity for the positive events compared to the negative events.

Across studies, Koch, Alves, et al. (2016) and Koch, Imhoff, et al. (2016) thereby generalized the valence asymmetry in similarity across several thousand word stimuli. Providing even more general data, they also analyzed available data sets from 13,915 words in the database by Warriner, Kuperman, and Brysbaert (2013) and 956 pictures from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2005). Both sets provide ratings of these stimuli regarding valence, arousal, and dominance/potency (Osgood, Suci, & Tannenbaum, 1957). Based on these ratings, one may compute the location of each stimulus in a three-dimensional space of valence, arousal, and potency, which is illustrated in Fig. 5. As the Figure already suggests, both for words and pictures, positive stimuli cluster closer together compared to negative stimuli. The overall difference is substantial for the IAPS pictures ($n_p^2 = 0.42$), while it is smaller for words ($n_p^2 = 0.04$). This difference is most likely due to the fact that many of the words in the set by Warriner and colleagues did not have a strong evaluative connotation. However, the differential similarity is also apparent on each dimension of valence, arousal, and potency alone (see table 6 in Koch, Alves, et al., 2016).

Given the present evidence, it seems a viable hypothesis that negative information is on average more diverse compared to positive information, and thus, good is more alike than bad.

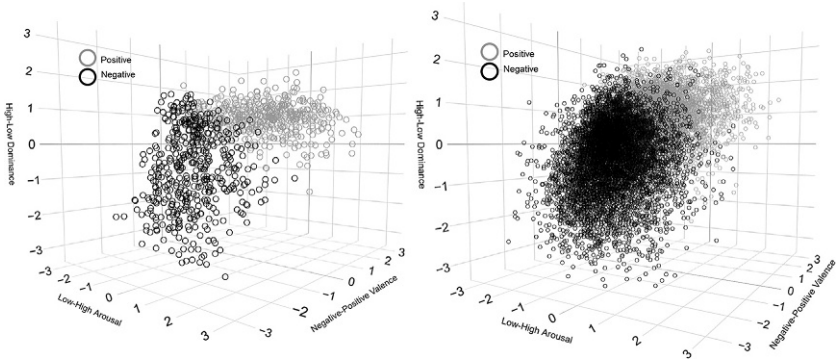

Fig. 5 3D plots of the 956 pictures in the IAPS database by Lang et al. (2005) in the left half, and the 13,915 words in the Warriner et al. (2013) database in the right half.

7.3 Why is good more alike?—The range principle

We explained the differential similarity with the range principle (Alves, Koch, & Unkelbach, 2017a).[d] It builds on our definition of good and bad: positive states are states that are beneficial for human life (i.e., the survival goal), and negative states are states that are harmful for human life. On the most basic level, human life is only possible in a very narrow range that is framed by "too much" and "too little" on any given dimension of physical input; for example, temperature, UV radiation, or oxygen concentration. States enabling human life are non-extreme, and thereby, good states should also be non-extreme states.

An illustrative case for the range principle is facial appearance and the beauty-in-averageness effect (Langlois & Roggman, 1990). For facial dimensions, noses may be too long or too short, foreheads might be too large or too small, and lips might be too thin or too thick. People prefer average values on facial dimensions. Thus, as shown by Langlois and Roggmann, when morphing two faces together, people prefer the more average morph over the two original faces. This phenomenon is not restricted to faces, but extends to other stimulus sets such as dogs, birds, or watches (see Halberstadt & Rhodes, 2000).

The range principle also translates from physical to psychological dimensions. Grant and Schwartz (2011) argued that on almost all psychological dimensions, positive states are non-extreme (see also Koch, Imhoff, et al., 2016). A person may be too talkative or too quiet, overly helpful or not helpful enough. People are likeable if they are non-extreme on their psychological variables. Even on psychological dimensions that seemingly have good or bad poles, people experience and evaluate the extreme ranges as negative. Thus, agreeableness may turn into conformity, conscientiousness may turn into perfectionism, and courage may turn into recklessness. This idea of a positive range framed by excess and deficiency of the given dimension goes back to Aristotle (1999; original 349 BC), who stated in the Nicomachean Ethics "…temperance and courage, then, are destroyed by excess and defect, and preserved by the mean."

Fig. 6 illustrates the range principle for the evaluation of a meal with the dimensions "Temperature" and "Spiciness." The figure immediately shows why the range principle leads to the greater diversity of negative states, as well as the higher similarity of positive states. There is a single "positive"

[d] The range principle should not be confused with the range-frequency principle by Parducci (1965) that Kanouse and Hanson (1972) recruited to explain valence asymmetries in evaluations.

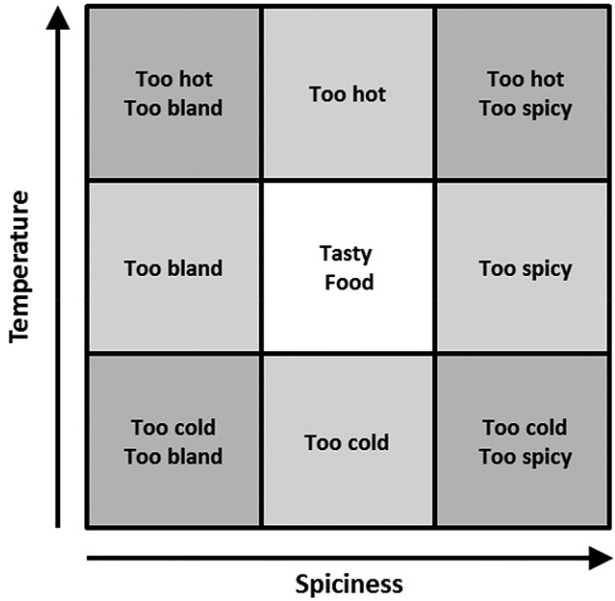

Fig. 6 A categorical representation of the range principle using the evaluation of food as an example.

area, which is framed by eight different areas of negativity. Thus, there is a greater diversity of negative states. It follows that in terms of spatial distance, positive points within this space will be closer together on average and therefore more similar to one another compared to negative points.

For the range principle to work, though, we need to refer to our definition of what is "good" and what is "bad." We assumed that there is a substance of reality and people's evaluation of this reality (see Leising et al., 2015). The range principle builds on reality's substance, not its evaluation (see also Leising, Scherbaum, Packmohr, & Zimmermann, 2018). Yet, substance and evaluation are psychologically not clearly separated, as attribute dimensions vary with regards to their evaluative implications. For example, the attribute "rudeness" would be highly evaluative; as most people would consider the absence of rudeness as positive, and the presence of rudeness as negative. Thus, the stronger the evaluative implication of an attribute dimension, the less likely it follows the range principle. However, one may also construe a given behavior as "rude" or "assertive"; assertiveness is less clear in its evaluative implication, and the assertiveness dimension again follows the range principle. By implication, direct organismic evaluations in terms of good and bad do not follow the range principle, as by

definition, more of "good" cannot be worse than less of "good." For substance dimensions, in particular those low on direct evaluative connotations, the range principle will hold (see also Grant & Schwartz, 2011).

While this might seem like a strong restriction, there are actually very few exceptions to the rule that most positive ranges are framed by two negative ranges, and we are unaware of any case where a negative range on an attribute is framed by two positive ranges. Even very prominent examples, such as Skowronski and Carlston's (1987) distinction between morality (i.e., being honest) and ability (i.e., being intelligent) follow the range principle. Being immoral and being stupid is evaluated by most people as negative; yet, being too honest and being too intelligent is also not necessarily positive. Imagine a student or a scientist who always succeeds at everything. Would this person be seen as likeable compared to the more average peer? Similar for morality, most people probably prefer to be friends with the fallible and sometimes not perfectly moral person, compared with an incredible moral and always infallible person. For most dimensions, most of the time, the range principle will hold.

The range principle then leads to higher similarity for positive information on average. If one samples any random combination of two coordinates of positive states on the two dimensions depicted in Fig. 6, as well as two coordinates of negative states, there is a substantially higher likelihood that the two positive coordinates are closer together and thereby more similar. In fact, if one neglects the distances within the nine depicted quadrants, the likelihood that the positive states will be more similar is eight to one.

The higher similarity also holds if one considers only one dimension. It also does not depend on the width of the ranges. However, the higher similarity of positive information is amplified if one assumes that dimensions are wider on the negatives of "too much" and "too little" (i.e., temperature ranges are far wider on the "too cold" and "too hot" sides), and it is attenuated if the positive range in the middle is very broad (e.g., for shoulder width, there is a broad positive range, and only at the very extremes people experience others at "too broad" or "too thin"). To be clear, though, the only necessary assumption is that the positive range is framed by two negative ranges; if this is given, the higher similarity of positive information follows.

We prefer the illustration with two dimensions, as it incorporates what has been termed the *Anna Karenina* principle (Diamond, 1997), as well as negativity dominance as described by Rozin and Royzman (2001). The Anna Karenina principle states that successes depend on multiple, conjunctive co-occurrences of positive factors, while failures depend on the

occurrence of a single negative factor. This principle is inherently present in our range explanation when more than one dimension is considered. Moving on any dimension in the area of too much or too little moves the respective evaluation coordinates outside the quadrant of positivity. Rozin and Royzman described negativity dominance as the most stable feature of "negativity bias," and it basically states that one drop of mineral oil may spoil one gallon a 1 of drinking water, but you cannot make oil drinkable by adding water. In the same vein, moving outside the good range on any dimension moves the whole feature combination into a negative quadrant, while a negative quadrant stays typically negative if only one attribute is shifted in the positive direction.

In addition, the two-dimensional illustration explains the attributional patterns observed for positive and negative outcomes. As people learn that positive states result from conjoint factors (i.e., a tasty meal needing to be well-seasoned and warm), they also spontaneously generate multiple factors for positive outcomes. Negative outcomes may be explained by a single factor, but it is important to determine which negative factor is responsible. Thereby, Fig. 6 explains both facets of attributional thinking reviewed above.

One may argue that Fig. 6 is flawed as it implies that negative states are also eight times more frequent, which would contradict most of the available literature on differential frequency of positive and negative information (e.g., Matlin & Stang, 1978; Unkelbach et al., 2019). To accommodate this empirical fact one may assume a normal distribution across the given dimensions. In other words, the combination of extremely hot and extremely dry days is rather infrequent, and most food is rather palatable.

Finally, the proposed range principle is also in line with the conception of an evaluative space (Cacioppo, Gardner, & Berntson, 1997). In such a space, positive and negative evaluations are independent unipolar axes, ranging from "not positive" to "very positive," and from "not negative" to "very negative." This creates a two-dimensional evaluative space in which a stimulus may be both positive and negative at the same time (i.e. ambivalent; see Schneider & Schwarz, 2017), rather than being located on a single bipolar dimension. The assumption that stimuli may have simultaneously negative and positive aspects follows from the range principle, if one allows that different evaluations following from different substance dimensions are summed up into separable evaluations (e.g., a meal being good because it is warm, but also bad because it is too bland).

Cacioppo et al. (1997) discussed mainly the case of attitudes; and an attitude object may be in a positive range in one attribute dimension, but in the

negative range in another attribute dimension. For a single dimension, though, a value on a given dimension cannot be simultaneously "good" or "bad." This is implied by Fig. 6. What may change is someone's goal or the evaluation context, making the same value on a given dimension "good" or "bad," the same way that a drink's evaluation may change depending on the weather circumstances (e.g., a hot drink on a hot vs. cold day). This flexibility follows from our Lewinian definition of what is "good" and "bad." However, without changes inside or outside the organism, a given value on an attribute dimension cannot be both "good" and "bad." The final location of a stimulus in the evaluative space is then determined by the organism's needs and goals, which assigns importance to the evaluation of a given stimulus' substance dimension.

7.4 How does differential similarity lead to differential processing?

Having delineated an answer to *why* negative information is more differentiated, we may now answer the *how* question of differential processing (Alves, Koch, & Unkelbach, 2016; Alves et al., 2015). To explain valence asymmetries in information processing, we refer to how information is represented in sub-symbolic, distributed memory models. In such models, information is presented as a vector of binary values that encodes the information (e.g., Fiedler, 1996; McClelland, McNaughton, & O'Reilly, 1995).

Let us illustrate the range principle within such an architecture. If we simplify the approach as mere categories of two states of negative information, but a single state of positive information, one needs two binary values to model negative information: "too much" may be present or absent, and "too little" may be present or absent. Fig. 7 illustrates this approach. Thus, because one needs more binary values to encode negative information, positive information becomes more similar (see Alves et al., 2016, for a person

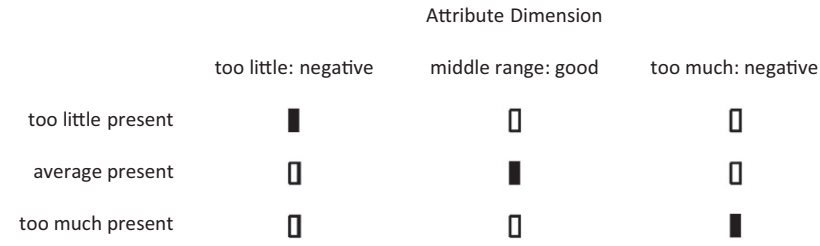

Fig. 7 Implementing the range principle in a sub-symbolic information vector for a single attribute dimension.

perception example). In other words, assuming a world with a single attribute dimension, all good information is identical (i.e., maximally similar), but negative information may differ. The important point for the *how* question is that if one models the greater similarity of positive information in a cognitive architecture, the greater impact, or the greater strength, of negative information follows.

First, this point may be illustrated with our food example. Again, a given dish might be too bland, well-seasoned, or too spicy. Or the meal might be too cold, warm, or too hot. As discussed above, independent of the ranges' breadth, as long as people evaluated the warm meal better as the cold meal, and the warm meal better as the hot meal, differential similarity follows.

Second, as already illustrated in Fig. 6, if we simplified the world again into categorical states, there are eight negative states, but a single positive state, for combinations of two attribute dimensions. To represent the greater diversity of the eight negative states, these need three binary units to code. The single positive state again needs a single unit. Fig. 8 illustrates this case for the example of a meal varying on the temperature and tastiness dimensions.

Finally, let us assume a simplified world that consists only of four positive states, based on two attribute dimensions each (i.e., 32 dimensions), which are then within their respective eight negative states. To fully code all states of this simplified world, we only need two binary units for the positive states, but eight units for the negative states. Fig. 9 compares the greater variety of negative information within our simplified world with the potential variety of negative information.

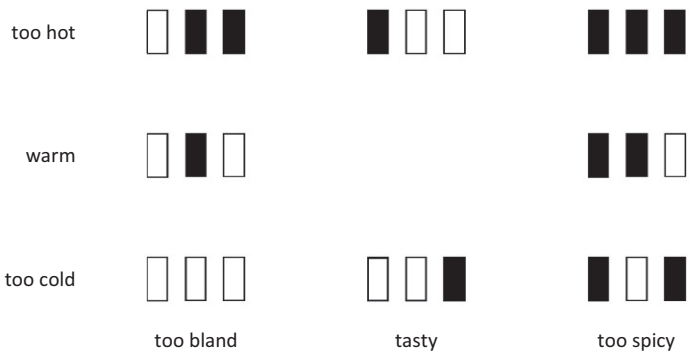

Fig. 8 An example of how one may code eight negative states with three binary variables.

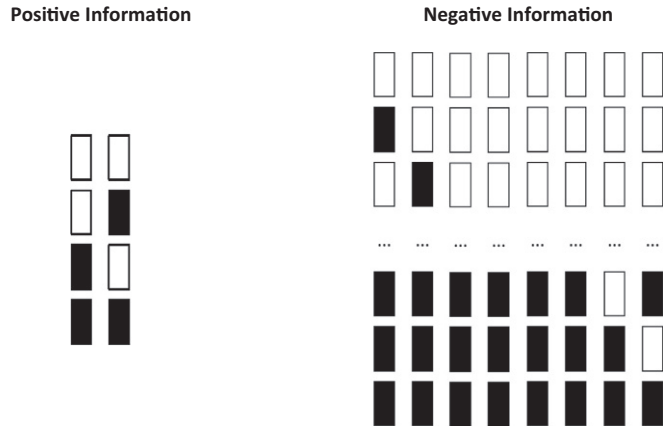

Fig. 9 Comparing the coding of four positive states with the coding of the corresponding 32 negative states, assuming the four states result from the combination of two attribute dimensions each.

As Fig. 9 visualizes, there is more information on the negative side, and the larger number of units necessary to code the information will lead to higher impact of this information. Typically, in sub-symbolic networks, an inactive unit carries as much meaning as an active unit. However, aiming to make this explanation more accessible, one can consider the black boxes as examples of "active" units, and the white boxes as examples of inactive units. If we compare Fig. 9 left and right panels, it is apparent that on average, negative information has many more active units. In other words, the greater diversity visualized in Fig. 6 will lead to both differential similarity and differential impact of negative information.

The presented architecture also accounts for the phenomenon of positivity offset and negativity bias as proposed by Cacioppo et al. (1997) and the positive-negative asymmetry described by Peeters (1991; see above). They defined positivity offset as "the tendency for there to be a weak positive (approach) motivational output at zero input" (Cacioppo et al., p. 12). As zero informational input should resemble the positive states on average more than negative states, people should interpret ambiguous information as positive at low levels of input. However, given that information is factually negative, the stronger match of the longer information vector will lead to a stronger response from the cognitive system, which is typically observed as a negativity bias.

This delineation is valid without considering the broader negative ranges on each dimension; this broader range is visible within Fig. 5, that is, the

stimulus words by Warriner et al. (2013) and the IAPS pictures by Lang et al. (2005). As we have discussed above, these are psychological evaluations, while the range principle builds merely on the physical substance, from which evaluations are derived. Arousal, dominance, and valence are psychological constructs, and one observes the differential similarity of positive and negative information within these constructs; however, we claim that these psychological differences emerge as a function of reality's substance.

The presented effects of greater diversity and lower similarity are amplified if one assumes that positive information is in addition more frequent (see Unkelbach et al., 2019). In addition, they are amplified if one assumes broader ranges for negative information. That is, if negative information is allowed to be more extreme as well, the similarity and intensity effects delineated here are stronger, as negative information will need even more units to code, for example, the difference between "too spicy" and "burning hot." As presented here, the differential similarity alone may suffice to explain the differential processing of positive and negative information.

In summary, the range principle and its cognitive implementation may operate without phylogenetic learning. Organisms only need to learn during their development that negative states are more differentiated (i.e., more ways to fail, more ways to get hurt, etc.) and code them accordingly. This simple principle incorporates and unifies almost all discussed explanations. It makes negative information more diverse and thereby less similar compared to positive information; and as a collateral, it also makes negative information on average more intense, more informative, and more distinct. Conversely, it makes positive information less diverse and thereby more similar, but it is also easier to process and more associative compared to negative information; and if one goes beyond a single dimension, the attributional patterns follow as well.

7.5 Testing the reversed causality: Does negative valence lead to greater differentiation?

So far, we have delineated why negative information should be more diverse, and accordingly, in spatial models of similarity, positive information should be more alike. Yet, one of the effects of negative information we have reviewed above is that negative information attracts more attention and that it is processed more deeply. It is thereby possible that the greater diversity and lower similarity for negative information is a

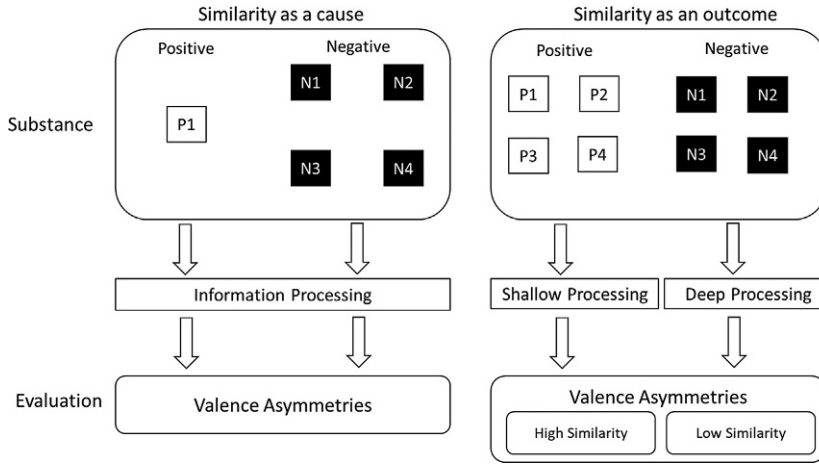

Fig. 10 Similarity as a cause (left side) and as an outcome (right side). The left side illustrates a simplified range principle, in which there are four states that people might evaluate negatively, but a single positive state. The right side assumes no differences in the substance, but due to the processing differences, positive information will appear more similar, and negative information more differentiated. *Based on Alves, H., Koch, A., & Unkelbach, C. (2019). The differential similarity of positive and negative information–an affect-induced processing outcome?* Cognition and Emotion*,* 33*, 1224–1238.*

consequence, and not the cause, of processing differences. Fig. 10 shows these two possibilities.

This alternative was suggested by Topolinski and Deutsch (2013), who argued that the negative affect elicited by negative information would lead to differential similarity (i.e., Fig. 10 right panel). Alves, Koch and Unkelbach (2019) tested this possibility in five experiments using a learning paradigm. In three experiments, they paired different Pokémon with monetary incentives or positive and negative words, holding the similarity of the incentives and word stimuli constant.

Fig. 11 shows the results across these experiments. Similarity was assessed by pairwise comparisons. Higher values indicate higher similarity. The Figure shows that the learning paradigm reliably changed stimulus valence, but it did not influence participants' similarity ratings. This differential effect was also not due to a lack of sensitivity. Participants' similarity ratings clearly separated "positive" from "negative" Pokémon; that is, participants rated "positive-negative" pairs as less similar compared to "negative-negative" and "positive-positive" pairs. However, "positive-positive" and "negative-negative" pairs' similarity did not differ.

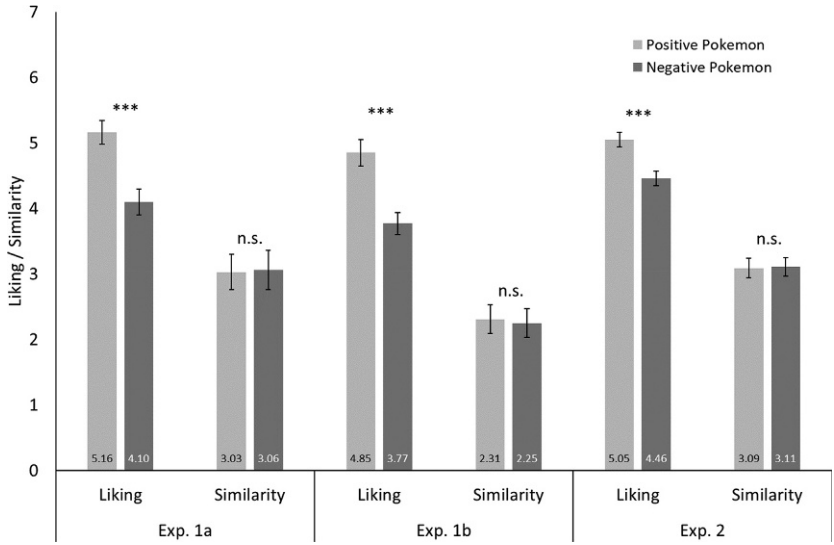

Fig. 11 Liking and similarity ratings for Pokémon. Higher values indicate higher likeability and higher similarity. Error bars represent standard errors of the means. *Data adapted from Alves, H., Koch, A., & Unkelbach, C. (2019). The differential similarity of positive and negative information–an affect-induced processing outcome?* Cognition and Emotion, 33, *1224–1238.*

In two further experiments, Alves et al. (2019) replicated the procedure by Topolinski and Deutsch (2013) and induced positive or negative affect by presenting pleasing or displeasing sounds during the similarity ratings of positive and negative stimuli. While they replicated the typical word valence asymmetry in similarity in Experiment 3 (see above; Fig. 4), sound valence had no effect on judged similarity. Finally, they presented fruit names as targets in Experiment 4, and again played the sounds used by Topolinski and Deutsch to induce positive and negative affect. Again, similarity ratings of the fruits did not differ as a function of externally induced positive or negative sounds.

In summary, these results do not support Fig. 10 right half. Neither experimentally induced valence (keeping similarity constant), nor externally induced valence changed similarity ratings. The differential similarity of positive and negative information is apparently not due to differential processing elicited by the micro-affect of the information's valence.

7.6 Cases of no valence asymmetries

Given the evidence so far, one may ask if there are cases when there should be no difference in how people process evaluative information. The

similarity explanation is clear in this respect. If for a given processing task positive and negative information is equally similar, then one should not find differential processing effects. The Pokémon experiments in the previous section lead to an illustrative thought experiment. In Pokémon card games, the animals' traits are given by arbitrary numbers on a set of dimensions. If the "good" Pokémon and "bad" Pokémon are equally diverse (and frequent; see Unkelbach et al., 2019), then one would not expect differential processing of these friendly or hostile Pokémon in terms of attention, memory, or speed. In fact, by manipulating similarity in such artificial ecologies, one may predict reversals of the standard effects, and we will present examples of such manipulations in the following section (e.g., Alves, Koch, & Unkelbach, 2018; Alves et al., 2015).

8. Novel insights, quantitative predictions, and old puzzles

The suggested similarity explanation provides substantial advances for research on valence asymmetries. Theoretically, it incorporates many of the previously discussed explanations such as informativeness, diagnosticity, or intensity. It also explains both positivity and negativity advantages as an outcome of ontogenetic learning. The proverbial saber tooth tiger has thus more impact on attention, perception, and memory, not because overlooking the tiger would prevent the passing of the genes, but because there is such a great diversity of potentially harmful stimuli, that negative information needs more computational space to be encoded.

Empirically, the explanation allows novel predictions that are difficult to derive from the previously established explanations alone. It allows to quantify and predict the differential impact of positive and negative information. That is, given one knows inter-stimulus similarity, one may make precise predictions about the processing consequences. In addition, one may predict systematic reversals of these effects. For example, in a world where positive information is more diverse and negative information therefore more similar, the pattern of differential processes should shift; if negative information is more similar, it enjoys the same advantages and disadvantages as positive information. The similarity explanation thereby allows on the empirical level to delineate novel effects, to make quantitative predictions, to create reversals of typical asymmetries, and to solve old puzzles. We will provide examples of these implications in the following.

8.1 Novel insights: Halo effects

A novel finding implied by the presented approach is that "halo" effects are overall stronger for positive information. Halo effects are among the best-established phenomena in social psychology (see Cooper, 1981; for an overview). Thorndike (1920) introduced the term to describe the phenomenon that ratings of soldiers by their superiors correlated higher with one another as expected. This implied that the officers' ratings on separate dimensions influenced each other. The most famous example for halo effects is Dion, Berscheid, and Walster's (1972) statement that "what is beautiful is good". People infer positive traits and behaviors from physical beauty. However, this notion already implies a fundamental asymmetry which we would predict from the presented similarity explanation. What is beautiful is good, but what is ugly is not bad; or at least, to a lesser extent. Thus, inferences from people's traits and behaviors to other traits and behaviors should be more likely if these are positive.

As our short review of valence asymmetries in person perception and impression formation showed, negative information has more impact on how people perceive others because it is stronger, more informative, and most of the time, more diagnostic. Thus, observing negative behaviors or receiving information about bad traits should have more impact on people's impression of others (Fiske, 1980; Skowronski & Carlston, 1987). Given our similarity explanation, this prediction needs to be qualified. As negative behaviors and negative traits are dissimilar and thereby distinct, they should be informative and diagnostic only for the specific trait dimension, and for the overall impression only insofar, as the trait dimension is relevant for the overall impression. For predicting other behaviors and other traits, negative information should have less impact, just because it is so distinct. For example, prompting the cognitive system as depicted in Fig. 9 with a positive trait will have a much higher chance co-activate the other positive traits due to their high similarity (see Alves et al., 2016; Fiedler, 1996). Yet, a negative trait will co-activate other traits to a much lesser extent, which again follows from their lower similarity. Thus, we predicted that being truthful makes people appear industrious, but lying should not make them appear lazy.

Gräf and Unkelbach (2016) tested this prediction in four experiments of which we report Experiments 2 and 3 here in more detail. Experiment 2 used traits to investigate halo effects. The traits were selected from a study by Abele, Uchronski, Suitner, and Wojciszke (2008). The positive and negative traits in Experiment 2 either related to the dimension of communion

(e.g., honest-dishonest, sociable-solitary) or agency (e.g., tidy-chaotic, industrious-lazy). Experiment 3 used behavior descriptions that related to Experiment 2's traits. Each trait was represented by several behavioral descriptions selected based on pre-tests. For example, being honest was represented by the description "….does not speak ill of other people behind their back." and being dishonest by "…does speak ill of other people behind their back." The trait industrious was, for example, represent by "…studies and works hard and continuously.", and the trait lazy by "…studies and works little and not more than required." Experiment 2 used 16 traits, orthogonally manipulating valence (positive vs. negative) and dimension (communion vs. agency). Experiment 3 used 48 behavior descriptions, again orthogonally manipulating valence and dimension.

In both experiments), participants observed a trait describing a person or a person showing a behavior. In Experiment 2, they indicated the likelihood that this person possesses one of the other traits. Each trait was paired with all the other traits of the same valence. In Experiment 3, participants rated a trait dimension that was not indicated by the behavior; thus, these ratings indicated halo effects.

Fig. 12 shows the averaged likelihood ratings.[e] In both experiments, halo effects were stronger for positive compared to negative information; that is, the assumed likelihood of a trait given a presented trait was higher for positive traits and trait ratings were higher given positive behaviors. In addition, halo effects were stronger within a given dimension compared to halo effects across dimensions (i.e., being industrious influenced being tidy more than being honest).

The remaining two experiments in Gräf and Unkelbach (2016) showed similar patterns. Experiment 1 was similar in design, but manipulated information valence between participants; this manipulation therefore allowed the alternative explanation of potential mood effects. Experiment 4 investigated halo effects when participants saw a person characterized by several behaviors (positive or negative) and then rated the target person on dimensions not related to these behaviors; for example, they would observe communal behaviors and would rate the target on agency traits. Again, positive behaviors led to stronger positive effects while the negative behaviors had a much smaller influence on the trait ratings on the unrelated dimension.

[e] Both experiments also manipulated whether the target person was shown with a face or not. We omit this additional factor here for the sake of brevity.

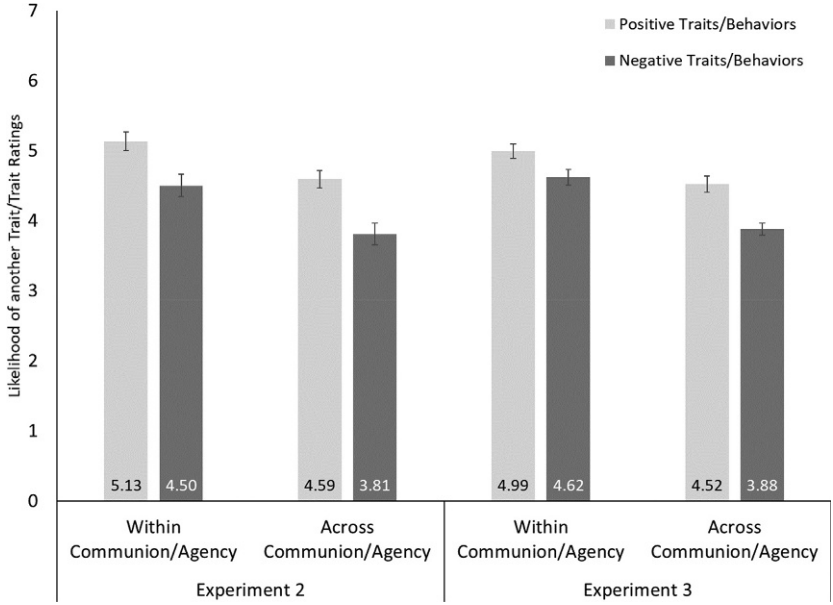

Fig. 12 The likelihood of another trait (Experiment 2) and trait ratings (Experiment 3) as a function of valence (positive vs. negative) and attribute dimensions (within vs. across warmth and competence). Higher values indicate higher likelihoods. Error bars represent standard errors of the means. *Data adapted from Gräf, M., & Unkelbach, C. (2016). Halo effects in trait assessment depend on information valence: Why being honest makes you industrious, but lying does not make you lazy.* Personality and Social Psychology Bulletin, 42*, 290–310.*

Thus, showing positive communal behaviors made targets also more agentic and vice versa, but showing negative communal behaviors made targets not less agentic.

This influence of information valence on halo effects was replicated by Gräf and Unkelbach (2018). They used the same behavior descriptions as Gräf and Unkelbach (2016), and again investigated participants' ratings of traits that were not implied by these descriptions (i.e., halo effects). In addition, they manipulated the targets' occupations. Occupations were pretested to be either defined by communal traits (e.g., kindergarten teacher) or agency traits (e.g., engineer). They predicted and found in three experiments that halo effects were stronger when occupations matched behaviors. That is, communal behaviors led to stronger halo effects from kindergarten teachers compared to engineers and vice versa for agentic behaviors. More importantly for the present purposes, they replicated the stronger halo effects of positive behaviors on traits; that is, when behaviors were not indicative of

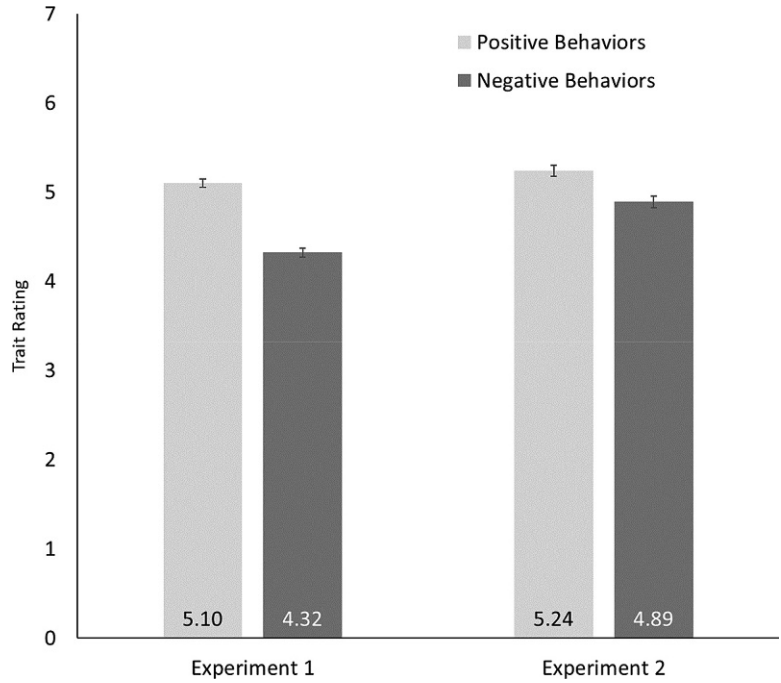

Fig. 13 Trait ratings based on behavior descriptions from Gräf and Unkelbach (2018) as a function of valence. Higher trait values indicate stronger halo effects. Error bars represent standard errors of the means. *Data adapted from Gräf, M., & Unkelbach, C. (2018). Halo effects from agency behaviors and communion behaviors depend on social context: Why technicians benefit more from showing tidiness than nurses do*. European Journal of Social Psychology, 48, *701–717.*

the traits. Fig. 13 shows the pattern for Experiment 1) and Experiment 2), collapsed across agency and communion behaviors and target occupations. The third experiment used the cover story of job applications, and did therefore not include a negative behaviors condition, which makes it less relevant here. As Fig. 13 shows, both experiments replicated the differential influence of information valence on halo effects.

This finding is not restricted to the present paradigm, but also apparent in other studies. For example, Carlston and Skowronski (2005; Experiment 3) investigated spontaneous trait inferences (STIs) and spontaneous trait transferences (STTs). Replicating the higher impact of negative behaviors on impression formation for behaviors that clearly imply a given trait, negative behaviors led to stronger effects on the trait dimension. However, ratings of traits not directly implied by the behavior were more strongly influenced by positive information (see Carlston & Skowronski, 2005; Fig. 5).

In addition, the findings may explain why the literature largely focuses on the positive side of evaluation when it comes to halo effects. Negative halo effects, also known as horn effects, are less frequently investigated because they are substantially smaller, or do not exist at all.

In summary, our underlying model predicts both positivity and negativity advantages in person perception and impression formation. Negative information has more influence when specific traits and behaviors are in the focus, due to its higher differentiation. In addition, when traits (e.g., being moral or helpful) are central for overall evaluations, negative information will have an overall stronger influence. However, positive information has more influence when inferences (e.g., halo effects) from presented information are investigated.

8.2 Quantitative predictions: Processing speed

As reviewed above, people classify evaluative information faster as "good" compared to "bad" (Bargh et al., 1992). They also make faster "word" and "non-word" classifications for positive compared to negative words (Unkelbach et al., 2010). The evidence as reviewed above for this differential effect is typically a negative correlation with evaluative ratings; thus, the more positive a stimulus, the faster its classification. However, for the 92 attitude objects we discussed above (Fazio et al., 1986; also, by Bargh et al., 1992; Klauer & Musch, 1999), the effect does not hold within the valence categories. For the 46 positively evaluated attitude objects, the correlation remains negative ($r = -0.59$, $P < 0.001$); the more positive, the faster the classification. For the 46 negatively evaluated attitude objects, the correlation becomes highly positive ($r = 0.71$, $P < 0.001$); the more negative, the faster the classification (Unkelbach et al., 2008). This is at odds with an unqualified main effect of valence, but in line with Fazio et al. (1986) assumption that evaluative extremity determines classification latency. It also shows that valence might not be responsible for the differential processing speed, but another underlying factor.

Our present similarity explanation, which Unkelbach et al. (2008) formulated as the "density hypothesis," predicts the processing speed difference. Due to the high similarity of positive information, presenting a positive stimulus does not only activate the specific stimulus, but "…a larger number of associated items will be affected as well, and the joint positive association of all these stimuli surrounding a positive concept will facilitate the response 'positive'." (p. 37). Conversely, for a negative stimulus "…a

smaller number of related, neighboring stimuli will be affected, and the joint associative reference to "negative" should thus be weaker. Rather than pointing to negativity as a common denominator of multiple prompted items, negative prompts may trigger the specific meaning and functional importance of distinct negative stimuli." This is in line with the sub-symbolic model we presented above (see also Alves et al., 2015, 2016).

To test the density hypothesis, and by implication, the similarity explanation for valence asymmetries, Unkelbach et al. (2008) asked participants to classified the 20 most positive and the 20 most negative attitude objects by Fazio et al. (1986). These latencies replicated the typical processing advantage for positive stimuli. To account for different explanations of this advantage, the authors predicted on the level of stimuli participants' average response latency from an attitude objects' valence (i.e., effect coded as "−1" for negative, and "+1" for positive attitude objects), extremity (i.e., absolute evaluation value), word length, and word frequency, based on the norm data by Klauer and Musch (1999). Thus, negative regression weights for valence indicate faster classifications for positive stimuli, negative regression weights for extremity indicate faster classifications for more extreme stimuli, positive regression weights for length indicate slower classifications for longer words, and negative regression weights indicate faster classifications for more frequent words. Table 4 upper half shows the results of this regression analysis (i.e., "Model I without density").

Table 4 shows the predicted relation of all variables with response latencies, with the exception of word frequency, which did not uniquely predict response latencies; however, the zero-order correlation shows the expected relation. The regression weights can be directly interpreted in terms of latency effects that are unique to a given predictor, as shared variance is not assigned to any predictor in multiple regression models. For example, independent of all other factors, if a word is one character longer, people need on average about 5 ms longer to classify it. For valence, the weight indicates that participants need on average 38 ms longer to classify a negative word. In addition, for each point of evaluation extremeness, participants are on average 25 ms faster.

Next, other participants did a multidimensional scaling of these stimuli; we already described this study within the "good is more alike than bad" section. Fig. 2 above presented the results of this scaling. Based on the stimuli's location in the resulting three-dimensional space, Unkelbach et al. (2008; Study 2) computed the average Euclidean distance of a given stimulus to all other stimuli within its valence cluster; for example, the stimulus "death" was very similar to all other negative attitude objects (e.g., war,

Table 4 Unstandardized multiple regression weights for predicting participants' response latencies for 40 attitude objects and zero-order correlations between latencies and predictors.

Predictor	*b*	*P* <	95%CI LL	95%CI UL	R^2	Adj. R^2	*r*	*P* <
Model I (without density)								
Intercept	745.84	0.001	708.19	783.49				
Valence	**−18.48**	0.003	−30.02	−6.94			**−0.373**	0.018
Extremity	**−24.80**	0.038	−48.08	−1.51			−0.294	0.065
Length	**5.40**	0.024	0.76	10.04			**0.362**	0.022
Frequency	−0.01	0.668	−0.04	0.03			−0.303	0.057
					0.396	**0.327**		
Model II (with density)								
Intercept	631.33	0.001	561.98	700.68				
Valence	−3.61	0.566	−16.28	9.06			**−0.373**	0.018
Extremity	−11.75	0.264	−32.79	9.29			−0.294	0.065
Length	**5.18**	0.012	1.22	9.14			**0.362**	0.022
Frequency	0.00	0.911	−0.03	0.03			−0.303	0.057
Density	**12.10**	0.001	5.60	18.59			**0.680**	0.001
					0.575	**0.513**		

Note: *b* indicates the unstandardized regression weight, *P* indicates the probability of the *b* parameters' *t*-value, and 95%CI LL and UL indicate the lower and upper limit of the *b* parameters 95% confidence limits. Regression weights that are significantly different from zero at $P < 0.05$ are present in bold.

bombs, funeral), while the stimulus "divorce" was rather dissimilar. They termed this the "density" index; this index is high for large spatial distances and thereby indicative of dissimilarity. It is low for small spatial distances and thereby indicative of similarity.

Then, they correlated the density index with participants' average response latencies on a stimulus level. Fig. 14 plots this correlation for the 40 stimuli. As Fig. 14 shows, there is a strong linear relation between similarity as indexed by density. The more centrally stimuli are located within their valence cluster, the faster people can classify them as "good" and "bad." Please note that these are not the spatial similarities; participants did not rate "holiday" and "death" to be similar. Rather, these two have a similar average distance relative to all stimuli of the same valence.

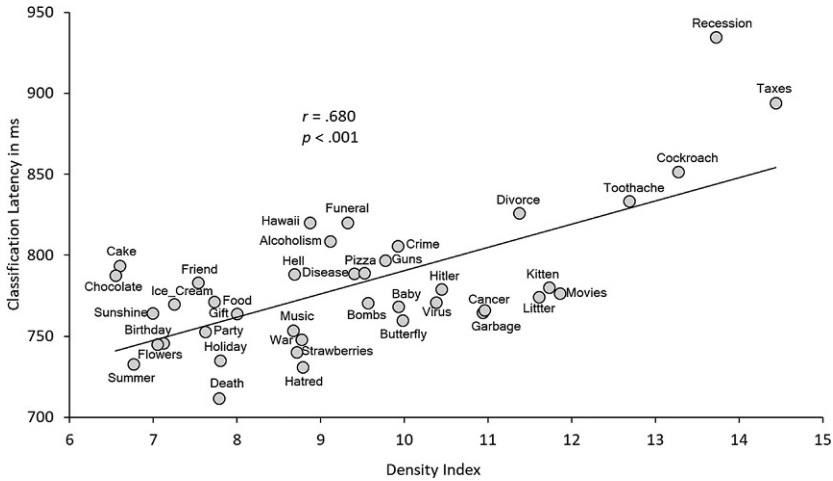

Fig. 14 The correlation between stimuli's spatial distance indexed as density and average classification latencies for 40 attitude objects. *Data adapted from Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis.* Journal of Personality and Social Psychology, 95, *36–49.*

Because similarity is substantially confounded with valence, as shown in Fig. 6, the valence asymmetry for stimulus classifications follows. However, the similarity explanation also allows for fast classifications of negative stimuli, given they are centrally located within their valence cluster (e.g., when two stimuli are sampled from the same quadrant in Fig. 6; e.g., "death").

Given the confounded nature of similarity and positivity, which is not by circumstances, but must follow from the range principle, it might still be the case that valence has a direct causal effect on latencies. Table 4 lower half therefore shows the results of the same multiple regression analysis as in the upper half, but with the density index that is plotted Fig. 14 as an additional predictor. First, this additional predictor substantially increases the model's fit, as indicated by the R^2 statistics. Second, adding density to the regression makes it the strongest predictor. Third and finally, density completely accounts for the variance explained in the model by valence and extremity, which are both no longer significant; that is, neither valence nor evaluative extremity contributes to predicting response latencies on a stimulus level beyond variance predicted by density.

In summary, the studies by Unkelbach et al. (2008) showed that it is possible to predict valence asymmetries quite precisely on a stimulus level, going substantially beyond the typical categorical effects (see our review above). They also

showed that similarity, here operationalized as stimulus density, fully accounts for the effect of evaluations and other variables such as extremity that are typically recruited to explain valence asymmetries in information processing.

8.3 Solving old puzzles: Recognition memory

As reviewed above, Ortony et al. (1983) reported a puzzle for valence asymmetries in recognition memory. Participants showed higher discrimination ability for negative stimuli, but lower thresholds to classify positive stimuli as "old." The present similarity approach provides a way to solve this puzzle. In addition, it allows again making quantitative predictions based on stimuli's overall similarity. Our model implies that positive information is more likely to co-activate other positive information, due to its high similarity within the same valence, while negative information should be easier to discriminate due to its lower similarity. In other words, participants should be more likely to discriminate "ugly" from "evil," while they should be more likely to confuse "beautiful" and "good," leading to higher false alarms for positive information. These differential false-alarm rates then lead to the observed effects on the signal detection parameters for recognition.

Alves et al. (2015) tested this explanation in two experiments using typical memory paradigms with recognition judgments (i.e., "old" vs. "new") as dependent variable. Thus, participants judged factually old and new stimuli as subjectively "old" and "new"; of interest are the "hits," correctly classifying the previously presented stimuli as "old," and "false alarms," incorrectly classifying new stimuli as "old." From the hit and false alarm rates, one may compute signal detection parameters, for example, d' for discrimination ability (i.e., how well one can discriminate between old and new stimuli), and C for response criterion (i.e., how likely one is to classify stimuli as old; sometimes also labeled a response bias; see Stanislaw & Todorov, 1999).

In Experiment 1, Alves et al. (2015) used the same stimuli as Unkelbach et al. (2008) and replicated the pattern by Ortony et al. (1983). Participants better discriminated old from new for negative stimuli, but showed a lower response threshold for positive items. To show that this "puzzle" emerges as a function of differential similarity, the authors predicted on a stimulus level the hits and false alarm rates from the stimulus density values from Unkelbach and colleagues.[f] Fig. 16 upper panel shows the scatter plots of

[f] One reviewer of Alves et al. (2015) noted that one should not compute signal detection parameters for stimuli, as stimuli cannot have discrimination ability or a response criterion. Alves and colleagues therefore reported the raw hit and false alarm rates.

the stimuli's hit and false alarm rates as a function of stimulus density. Density did not influence the hit rates, but the false alarm rates showed the expected negative relation: the less densely clustered a stimulus was within its valence cluster, the less likely it was to be falsely classified as "old." Given the natural confound of valence and similarity, this leads to the differential parameter estimates for positive and negative information. Positive stimuli are more likely to be judged as "old," while negative stimuli are easier to discriminate.

As in Unkelbach et al. (2008) analyses, this relation was not due to other stimulus features. In a regression analysis, Alves and colleagues predicted false alarms from density, word frequency, and evaluative ratings; density, that is, inter-stimulus similarity, was the only significant predictor in this analysis.

In Experiment 2, Alves et al. (2015) went one step further and manipulated the similarity of positive and negative stimuli by pre-selecting the stimuli. They created a "natural" condition, in which positive stimuli were more similar compared to one another relative to the negative stimuli's similarity. In a "reversed" condition, negative stimuli were more similar. Fig. 15 shows the d' and C estimates on the participant level. For both parameters, there is a clear interaction. The natural condition again replicates the pattern by Ortony et al. (1983). Yet, if negative information is marked by higher density, people have a lower response criterion and a higher discrimination ability for positive stimuli.

Fig. 16 bottom panel shows the relation of stimulus density and false alarm and hit rates on the stimulus level, collapsed across conditions. Replicating Experiment 1, independent of valence, stimulus density predicts

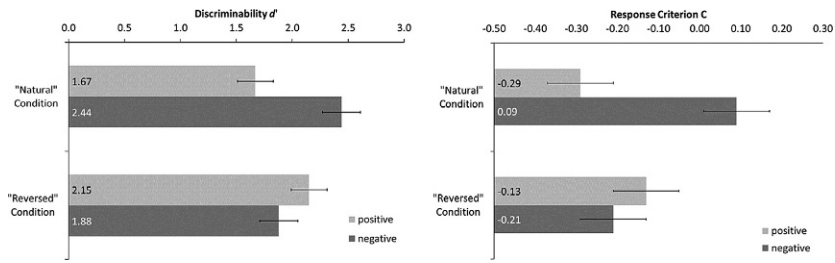

Fig. 15 Discrimination Ability d' (left side) and Response Criterion C (right side) in a recognition task as a function of stimulus valence and stimulus similarity. The natural condition presents the typical situation when similarity and valence are confounded. The reversed condition presents a situation where negative information is more alike compared to positive information. *Based on Alves, H., Unkelbach, C., Burghardt, J., Koch, A., Krüger, T., & Becker, V. D. (2015). A density explanation of valence asymmetries in recognition memory.* Memory & Cognition*,* 43*, 896–909; Experiment 2.*

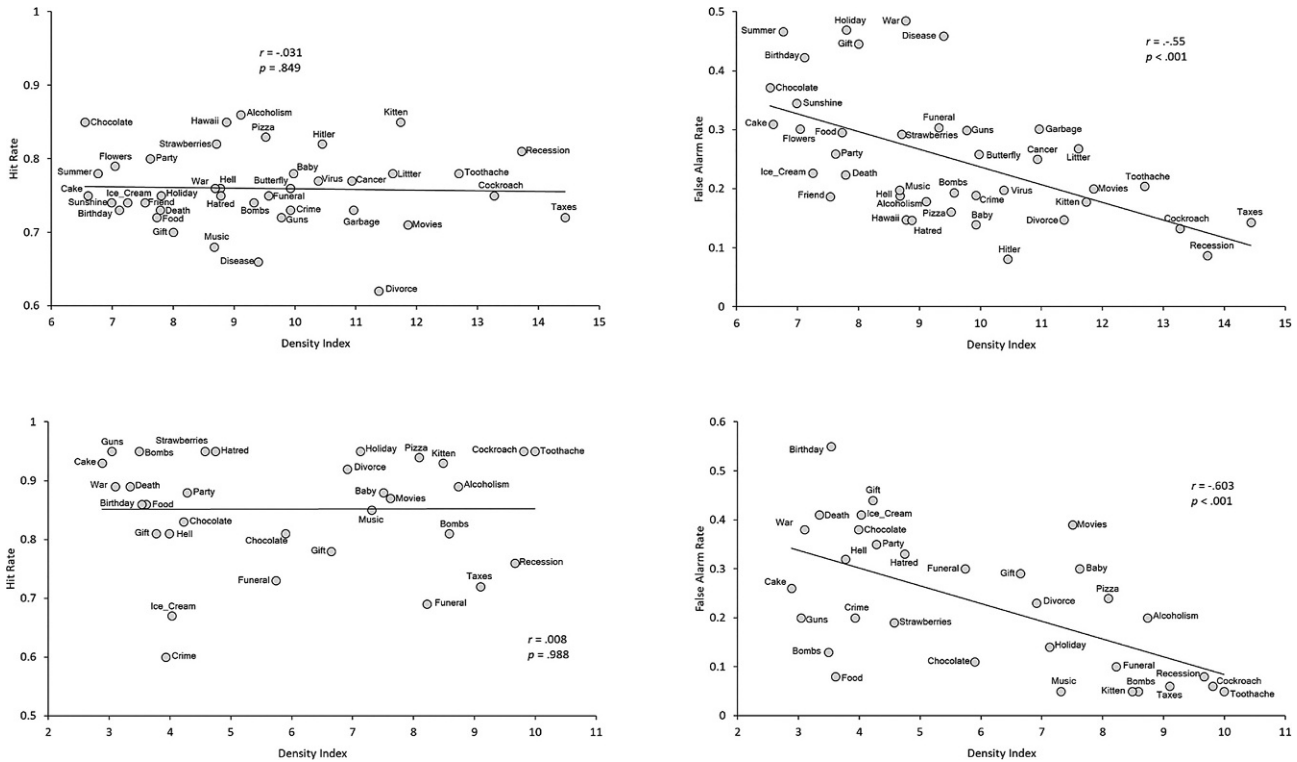

Fig. 16 Hit and false alarm rates on a stimulus level as a function of stimulus density. The upper half shows the rates for the stimuli used in Experiment 1 of Alves et al. (2015). The lower half shows the association of hit and false alarm rates and density for Experiment 2's stimuli, collapsed across the "natural" and "reversed" conditions in Experiment 2. As some stimuli appeared in both conditions, some stimuli are presented twice in this graph.

stimuli's false alarm rates. In the corresponding regression analysis predicting false alarms from density, word frequency, and evaluative ratings, density was again the only significant predictor.

In summary, both experiments showed that the similarity explanation solves riddles within the existing literature, it again allowed quantitative predictions, and by directly manipulating inter-stimulus similarity (i.e., density), one may flip the patterns that are typically obtained. The correspondence of Fig. 15 with the latency results shown in Fig. 13 is striking, although both dependent variables, response latencies and false alarm rates, are theoretically distinct. Considering the underlying similarity explanation for the differential processing of positive and negative information, the high correspondence lends credibility to a common underlying mechanism.

8.4 Summary of the similarity explanation

We presented evidence that good is on average more alike than bad (Koch, Alves, et al., 2016). This holds for both verbal stimuli, pictorial stimuli, and experiences (see Fig. 4). We also provide an explanation for this differential similarity (Alves et al., 2017a); based on the assumption that for physical and psychological attribute dimensions, the positive range on a dimension is framed by two negative ranges of "too much" and "too little" (see Fig. 6). We illustrated novel predictions of this explanation for halo effects, namely that positive information should lead to stronger halo effects (Gräf & Unkelbach, 2016, 2018). We showed that if one measures the differential similarity of positive and negative information, this similarity accounts for the effects of valence on processing speed (Unkelbach et al., 2008) and recognition memory (Alves et al., 2015). In addition, we showed that the similarity explanation allows to quantitatively predict valence asymmetries in evaluation speed, classification speed, or recognition memory.

There are many aspects of the assumed similarity explanation we did not address here, such as predicting association asymmetries or attributional thinking; we also focused on the processing level, and neglected differential similarity effects on the functional level (e.g., Alves, Koch, & Unkelbach, 2017b). There is still empirical work to be done. Yet, at the present stage, we believe the similarity explanation advances theorizing and empirical work on valence asymmetries in several ways. First, it simultaneously explains positivity and negativity advantages; second, it follows theoretically from a straightforward and well-established principle and it builds on

existing explanations; third, it answers both the *why* and the *how* questions of differential processing; and fourth, it allows quantitative predictions, going beyond the mere classification of information into "good" and "bad."

9. Conclusions

Valence asymmetries have been and still are a fascinating topic for psychologists in general and social psychologists in particular. Over the years, researchers investigated many phenomena and explanations, especially addressing advantages of negative information. Here, we presented a novel and potentially unifying principle that may explain processing advantages of negative information, but importantly, also positive information's advantages: Differential similarity.

Differential similarity of positive and negative information results from the range principle which assumes that for substance attributes, positively evaluated states are framed by negatively evaluated states of "too much" and "too little." Thus, negative information is more diverse, leading to higher similarity of positive information. This higher similarity may be measured and serves as the explanatory construct for processing differences.

Negative information's greater diversity also accounts for the apparent meta-pattern in the data. There is more variability of effects (see, for example, Section 4.1) and theories (see Section 6) for negative compared to positive information. By our analysis, this reflects reality. Given there is more variety on the negative input side, there should also be more variety and less coherence on the output side; that is, empirical effects and theoretical explanations.

We do not claim that other approaches are incorrect; in fact, our approach incorporates and builds on previous explanations. Yet, before one recruits motivational, emotional, or phylogenetic explanations, the basic structure of the information needs to be considered; it may suffice to explain the different advantages. As people learn that there are more ways to be bad than to be good, the cognitive system needs more units to code negative information, which leads to the observable valence asymmetries in the processing of good and bad.

Acknowledgment

Preparation of this manuscript was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft; DFG-Grant UN 273/4-2), awarded to C.U.

References

Abele, A. E., Uchronski, M., Suitner, C., & Wojciszke, B. (2008). Towards an operationalization of the fundamental dimensions of agency and communion: Trait content ratings in five countries considering valence and frequency of word occurrence. *European Journal of Social Psychology*, *38*, 1202–1217.

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

Alves, H., Koch, A., & Unkelbach, C. (2017a). Why good is more alike than bad: Processing implications. *Trends in Cognitive Sciences*, *21*, 72–82.

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

Alves, H., Koch, A., & Unkelbach, C. (2018). A cognitive-ecological explanation of intergroup biases. *Psychological Science*, *29*, 1126–1133.

Alves, H., Koch, A., & Unkelbach, C. (2019). The differential similarity of positive and negative information—An affect-induced processing outcome? *Cognition and Emotion*, *33*, 1224–1238.

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

Anderson, N. H. (1965). Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology*, *70*, 394–400.

Anisfeld, M., & Lambert, W. E. (1966). When are pleasant words learned faster than unpleasant words? *Journal of Verbal Learning and Verbal Behavior*, *5*, 132–141.

Anselmi, P., Vianello, M., & Robusto, E. (2011). Positive associations primacy in the IAT: A Many-Facet Rasch measurement analysis. *Experimental Psychology*, *58*, 376–384.

Aquino, J. M., & Arnell, K. M. (2007). Attention and the processing of emotional words: Dissociating effects of arousal. *Psychonomic Bulletin & Review*, *14*, 430–435.

Aristotle. (1999). *Nicomachean ethics*. (Ross, W.D., translator) Ontario, Canada: Batoche Books translated.

Baas, M., De Dreu, C. K., & Nijstad, B. A. (2008). A meta-analysis of 25 years of mood-creativity research: Hedonic tone, activation, or regulatory focus? *Psychological Bulletin*, *134*, 779–806.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*, 445–459.

Bar-Anan, Y., Nosek, B. A., & Vianello, M. (2009). The sorting paired features task: A measure of association strengths. *Experimental Psychology*, *56*, 329–343.

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

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

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

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

Becker, D. V., & Srinivasan, N. (2014). The vividness of the happy face. *Current Directions in Psychological Science*, *23*, 189–194.

Bohner, G., Bless, H., Schwarz, N., & Strack, F. (1988). What triggers causal attributions? The impact of valence and subjective probability. *European Journal of Social Psychology*, *18*, 335–345.

Boucher, J., & Osgood, C. E. (1969). The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior*, *8*, 1–8.

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings (technical report C-1)*. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, *5*, 73–99.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, *115*, 401–423.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review*, *1*, 3–25.

Carlston, D. E., & Skowronski, J. J. (2005). Linking versus thinking: Evidence for the different associative and attributional bases of spontaneous trait transference and spontaneous trait inference. *Journal of Personality and Social Psychology*, *89*, 884–898.

Charles, S. T., Mather, M., & Carstensen, L. L. (2003). Aging and emotional memory: The forgettable nature of negative images for older adults. *Journal of Experimental Psychology: General*, *132*, 310–324.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics*. New York, NY: Harcourt Brace Jovanovich.

Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, *90*, 218–244.

Corns, J. (2018). Rethinking the negativity bias. *Review of Philosophy and Psychology*, *9*, 607–625.

Davis, M. A. (2009). Understanding the relationship between mood and creativity: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *108*, 25–38.

Dawkins, R. (1976). *The selfish gene*. Oxford, UK: Oxford University Press.

Diamond, J. (1997). *Guns, germs, and steel: The fates of human societies*. New York, NY: Norton.

Dijksterhuis, A., & Aarts, H. (2003). On wildebeests and humans: The preferential detection of negative stimuli. *Psychological Science*, *14*, 14–18.

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

Eskreis-Winkler, L., & Fishbach, A. (2019). Not learning from failure—The greatest failure of all. *Psychological Science*, *30*, 1733–1744.

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., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013–1027.

Fazio, R. H., Pietri, E. S., Rocklage, M. D., & Shook, N. J. (2015). Positive versus negative valence: Asymmetries in attitude formation and generalization as fundamental individual differences. In J. M. Olson & M. P. Zanna (Eds.), Vol. 51, *Advances in experimental social psychology* (pp. 97–146). Burlington, MA: Academic Press.

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.

Feldman, S. (1966). Motivational aspects of attitudinal elements and their place in cognitive interaction. In S. Feldman (Ed.), *Cognitive consistency: Motivational antecendents and behavioral consequences*. London, UK: Academic Press.

Feltz, A. (2007). The Knobe effect: A brief overview. *The Journal of Mind and Behavior*, *28*, 265–277.
Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review*, *103*, 193–214.
Fiedler, K., Freytag, P., & Unkelbach, C. (2011). Great oaks from giant acorns grow: How causal-impact judgments depend on the strength of a cause. *European Journal of Social Psychology*, *41*, 162–172.
Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, *38*, 889–906.
Frischen, A., Eastwood, J. D., & Smilek, D. (2008). Visual search for faces with emotional expressions. *Psychological Bulletin*, *134*, 662–676.
Gaillard, R., Del Cul, A., Naccache, L., Vinckier, F., Cohen, L., & Dehaene, S. (2006). Nonconscious semantic processing of emotional words modulates conscious access. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 7524–7529.
Gräf, M., & Unkelbach, C. (2016). Halo effects in trait assessment depend on information valence: Why being honest makes you industrious, but lying does not make you lazy. *Personality and Social Psychology Bulletin*, *42*, 290–310.
Gräf, M., & Unkelbach, C. (2018). Halo effects from agency behaviors and communion behaviors depend on social context: Why technicians benefit more from showing tidiness than nurses do. *European Journal of Social Psychology*, *48*, 701–717.
Grant, A. M., & Schwartz, B. (2011). Too much of a good thing: The challenge and opportunity of the inverted U. *Perspectives on Psychological Science*, *6*, 61–76.
Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.
Halberstadt, J., & Rhodes, G. (2000). The attractiveness of nonface averages: Implications for an evolutionary explanation of the attractiveness of average faces. *Psychological Science*, *11*, 285–289.
Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology*, *54*, 917–924.
Harris, C. R., & Pashler, H. (2004). Attention and the processing of emotional words and names: Not so special after all. *Psychological Science*, *15*, 171–178.
Herring, D. R., White, K. R., Jabeen, L. N., Hinojos, M., Terrazas, G., Reyes, S. M., … Crites, S. L., Jr. (2013). On the automatic activation of attitudes: A quarter century of evaluative priming research. *Psychological Bulletin*, *139*, 1062–1089.
Hess, T. M., Popham, L. E., & Growney, C. M. (2017). Age-related effects on memory for social stimuli: The role of valence, arousal, and emotional responses. *Experimental Aging Research*, *43*, 105–123.
Hilbig, B. E. (2009). Sad, thus true: Negativity bias in judgments of truth. *Journal of Experimental Social Psychology*, *45*, 983–986.
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.
Ihmels, M., Freytag, P., Fiedler, K., & Alexopoulos, T. (2016). Relational integrativity of prime-target pairs moderates congruity effects in evaluative priming. *Memory & Cognition*, *44*, 565–579.
Inaba, M., Nomura, M., & Ohira, H. (2005). Neural evidence of effects of emotional valence on word recognition. *International Journal of Psychophysiology*, *57*, 165–173.
Isen, A. M., Shalker, T. E., Clark, M., & Karp, L. (1978). Affect, accessibility of material in memory, and behavior: A cognitive loop? *Journal of Personality and Social Psychology*, *36*, 1–12.

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, *75*, 887–900.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 219–266). Cambridge, MA: Academic Press.

Kanouse, D. E., & Hanson, L. R., Jr. (1972). Negativity in evaluations. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 47–62). Morristown, NJ: General Learning Press.

Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual Review of Psychology*, *31*, 457–501.

Kenrick, D. T., Griskevicius, V., Neuberg, S. L., & Schaller, M. (2010). Renovating the pyramid of needs: Contemporary extensions built upon ancient foundations. *Perspectives on Psychological Science*, *5*, 292–314.

Kensinger, E. A. (2009). *Emotional memory across the adult lifespan*. New York, NY: Psychology Press.

Klauer, K. C., & Musch, J. (1999). Eine Normierung unterschiedlicher Aspekte der evaluativen Bewertung von 92 Substantiven [A standardization of various aspects of the evaluation of 92 nouns]. *Zeitschrift für Sozialpsychologie*, *30*, 1–11.

Klauer, K. C., Teige-Mocigemba, S., & Spruyt, A. (2009). Contrast effects in spontaneous evaluations: A psychophysical account. *Journal of Personality and Social Psychology*, *96*, 265–287.

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

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

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

Koch, A., Speckmann, F., & Unkelbach, C. (2020). Q-SpAM: How to efficiently measure similarity in online research. *Sociological Methods and Research*. https://doi.org/10.1177/0049124120914937.

Król, M., & Król, M. E. (2019). A valence asymmetry in pre-decisional distortion of information: Evidence from an eye tracking study with incentivized choices. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *5*, 2209–2223.

Krull, D. S., & Dill, J. C. (1998). Do smiles elicit more inferences than do frowns? The effect of emotional valence on the production of spontaneous inferences. *Personality and Social Psychology Bulletin*, *24*, 289–300.

Kun, A., & Weiner, B. (1973). Necessary versus sufficient causal schemata for success and failure. *Journal of Research in Personality*, *7*, 197–207.

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

Leising, D., Scherbaum, S., Locke, K. D., & Zimmermann, J. (2015). A model of "substance" and "evaluation" in person judgments. *Journal of Research in Personality*, *57*, 61–71.

Leising, D., Scherbaum, S., Packmohr, P., & Zimmermann, J. (2018). Substance and evaluation in personality disorder diagnoses. *Journal of Personality Disorders*, *32*, 766–783.
Lewin, K. (1943). Psychological ecology. In D. Cartwright (Ed.), *Field theory in social science: Selected theoretical papers by Kurt Lewin* (pp. 17–187). London: Social Science Paperbacks. [Book published 1952].
Liu, J. H., Karasawa, K., & Weiner, B. (1992). Inferences about the causes of positive and negative emotions. *Personality and Social Psychology Bulletin*, *18*, 603–615.
Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*, 370–396.
Matlin, M. W., & Stang, D. J. (1978). *The Pollyanna principle: Selectivity in language, memory, and thought*. Cambridge, MA: Shenkman.
McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–457.
Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin*, *130*, 711–747.
Nasrallah, M., Carmel, D., & Lavie, N. (2009). Murder, she wrote: Enhanced sensitivity to negative word valence. *Emotion*, *9*, 609–618.
Ohira, H., Winton, W. M., & Oyama, M. (1998). Effects of stimulus valence on recognition memory and endogenous eyeblinks: Further evidence for positive-negative asymmetry. *Personality and Social Psychology Bulletin*, *24*, 986–993.
Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology*, *80*, 381–396.
Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*, 483–522.
Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, *97*, 315–331.
Ortony, A., Turner, T. J., & Antos, S. J. (1983). A puzzle about affect and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 725–729.
Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.
Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, *72*, 407–418.
Peeters, G. (1971). The positive-negative asymmetry: On cognitive consistency and positivity bias. *European Journal of Social Psychology*, *1*, 455–474.
Peeters, G. (1991). Evaluative inference in social cognition: The roles of direct versus indirect evaluation and positive-negative asymmetry. *European Journal of Social Psychology*, *21*, 131–146.
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.
Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.
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.
Robinson-Riegler, G. L., & Winton, W. M. (1996). The role of conscious recollection in recognition of affective material: Evidence for positive-negative asymmetry. *The Journal of General Psychology*, *123*, 93–104.
Rozin, P., Berman, L., & Royzman, E. (2010). Biases in use of positive and negative words across twenty natural languages. *Cognition and Emotion*, *24*, 536–548.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*, 296–320.
Schneider, I. K., & Schwarz, N. (2017). Mixed feelings: The case of ambivalence. *Current Opinion in Behavioral Sciences*, *15*, 39–45.
Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1–66.
Sherman, J. W., Calanchini, J., & Hehman, E. (2017). Intergroup bias (generally) reflects more positivity than negativity. In *Paper presented at the 18th General Meeting of the European Association of Social Psychology*. Spain: Granada.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127–190.
Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology*, *52*, 689–699.
Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*, 131–142.
Snodgrass, L. L., & Harring, K. E. (2004). Right hemisphere positivity bias in preconscious processing: Data from five experiments. *Current Psychology*, *23*, 318–335.
Sparks, J., & Ledgerwood, A. (2017). When good is stickier than bad: Understanding gain/loss asymmetries in sequential framing effects. *Journal of Experimental Psychology: General*, *146*, 1086–1105.
Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology*, *56*, 283–294.
Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*, 137–149.
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.
Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin*, *110*, 67–85.
Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*, 193–210.
Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*, 25–29.
Topolinski, S., & Deutsch, R. (2013). Phasic affective modulation of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 414–436.
Unkelbach, C. (2012). Positivity advantages in social information processing. *Social and Personality Psychology Compass*, *6*, 83–94.
Unkelbach, C., Bayer, M., Alves, H., Koch, A., & Stahl, C. (2011). Fluency and positivity as possible causes of the truth effect. *Consciousness and Cognition*, *20*, 594–602.
Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis. *Journal of Personality and Social Psychology*, *95*, 36–49.
Unkelbach, C., Koch, A., & Alves, H. (2019). The evaluative information ecology: On the frequency and diversity of "good" and "bad". *European Review of Social Psychology*, *30*, 216–270.
Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition*, *160*, 110–126.
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.

Vogt, J., De Houwer, J., Koster, E. H., Van Damme, S., & Crombez, G. (2008). Allocation of spatial attention to emotional stimuli depends upon arousal and not valence. *Emotion*, *8*, 880–885.

Von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld [The effects of field formation in the trace field]. *Psychologische Forschung*, *18*, 299–342.

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.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*, 548–573.

Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York, NY: Springer.

Wolford, G., & Morrison, F. (1980). Processing of unattended visual information. *Memory & Cognition*, *8*, 521–527.

Zajonc, R. B. (1968). Attitudninal effects of mere exposure. *Journal of Personality and Social Psychology*, *7*, 1–29.

Zeelenberg, R., Wagenmakers, E. J., & Rotteveel, M. (2006). The impact of emotion on perception: Bias or enhanced processing? *Psychological Science*, *17*, 287–291.