Manifestations and Consequences of Negative Information's Great Diversity

Inauguraldissertation
zur
Erlangung des Doktorgrades
der Humanwissenschaftlichen Fakultät
der Universität zu Köln
nach der Prüfungsordnung vom 10.05.2010
vorgelegt von

Hans Alves

aus
Viersen

Tag der Abgabe: 14.03.2016

Erklärung

Kapitel 1 beruht teilweise auf folgendem Manuskript:

Koch, A., Alves, H., Krüger, T., & Unkelbach, C. (2016). A General Valence Asymmetry in Similarity: Good Is More Alike Than Bad. *Journal of Experimental Psychology. Learning, Memory, and Cognition*.

Kapitel 2 beruht auf folgendem Manuskript:

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

Ich habe die Idee entwickelt, die Datenerhebung überwacht, die Datenanalyse ausgeführt und das Manuskript geschrieben. Sowohl bei Ideenausarbeitung, als auch beim Schreiben des Manuskripts haben der zweite und dritte Autor wertvolle Beiträge eingebracht.

Kapitel 3 beruht auf folgendem Manuskript:

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

Ich habe eine erste Idee mit dem zweiten Autor entwickelt. Eine Weiterentwicklung der Idee profitierte durch wertvolle Beiträge des vierten, fünften und sechsten Autors. Ich habe die Datenerhebung überwacht und die Analyse der Daten ausgeführt. Die Bereitstellung des Stimulusmaterials sowie methodische Beratung erfolgte durch die dritte Autorin. Ich habe das Manuskript geschrieben welches durch wertvolle Beiträge aller Autoren profitiert hat.

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

Hans Alves, Köln, Oktober 2016

Summary

In the present dissertation, I propose a general, robust, and objective characteristic of the information environment, according to which negative information is more diverse than positive information. I present an explanatory framework for this phenomenon based on the non-extremity of positive qualities. Specifically, most attribute dimensions host one “positive” range which is surrounded by two distinct “negative” ranges, resulting in a greater diversity of negative compared to positive attributes, stimuli, and information in general. Chapter 1 of my dissertation reviews evidence for the proposed greater diversity of negative information, showing that positive stimuli are consistently perceived as more similar than negative stimuli. Chapter 2 presents a line of research that demonstrates the same phenomenon in the domain of person perception. People perceive other people they like as more similar to one another than other people they dislike. I further suggest that the greater diversity of negative information as an “extrapsychic” phenomenon can account for a number of valence asymmetries in information processing that have previously been explained based on “intrapsychic” affective reactions. In line with this idea, Chapter 3 presents research showing that negative words enjoy a recognition advantage over positive words because they are more diverse. Chapter 4 discusses limitations, open questions and further implications of the current empirical findings and of the explanatory framework provided.

Keywords: valence asymmetries, information processing, similarity, ecology, affect, cognition, memory, recognition, person perception

Zusammenfassung

Die vorliegende Dissertation postuliert eine generelle, robuste, und objektive Asymmetrie der Informationsumwelt nach der negative Informationen diverser sind als positive Informationen. Für diese Asymmetrie formuliere ich ein theoretisches Erklärungsmodell basierend auf der „Moderatheit“ positiver Qualitäten. Demnach besitzen die meisten Attributdimensionen einen positiven (moderaten) Bereich, welcher umgeben ist von zwei distinkten negativen Bereichen. Dies führt zu einer größeren Diversität negativer Attribute, negativer Stimuli und negativer Informationen generell. Kapitel 1 meiner Dissertation gibt einen Überblick über empirische Evidenz für die postulierte größere Diversität negativer Informationen, welche zeigt, dass positive im Vergleich zu negativen Stimuli als ähnlicher wahrgenommen werden. Kapitel 2 präsentiert eine Forschungslinie die zeigt, dass die Diversitätsasymmetrie auch auf Personenwahrnehmung anwendbar ist. Es zeigt sich, dass beliebte Personen als ähnlicher wahrgenommen und ähnlicher beschrieben werden als unbeliebte Personen. Darüber hinaus behaupte ich, dass die geringere Diversität positiver Informationen als „extrapsychisches“ Phänomen eine Reihe von Asymmetrien in der Verarbeitung positiver und negativer Informationen erklären kann, welche in der Vergangenheit als Folge „intrapsychischer“ affektiver Reaktionen erklärt wurden. Als Beleg hierfür präsentiert Kapitel 3 eine Forschungslinie die zeigt, dass negative Wörter akkurater Wiedererkannt werden können weil sie diverser sind. Kapitel 4 diskutiert Einschränkungen, offene Fragen, und weitere Implikationen der empirischen Befunde und des vorgestellten theoretischen Erklärungsmodells.

Schlagwörter: Valenzasymmetrien, Informationsverarbeitung, Ähnlichkeit, Ökologie, Affekt, Kognition, Gedächtnis, Recognition, Personenwahrnehmung

Structure

Chapter 1 – Introduction

The central dualism within metaphysical disciplines such as religion or philosophy has always been the distinction between good and bad. From the religious distinctions between god and devil, paradise and hell, virtue and sin, to Immanuel Kant's formulation of universal rules of "good" behavior, the antagonism between good and bad is the primary theme of human existence. Good and bad are also among the first concepts that children learn and distinguishing between positive and negative is essential for humans to navigate complex social environments (Lewin, 1935). Therefore it is no surprise that the positive/negative dualism is a central concept within psychological theories as well. Psychologists have long recognized that the valence dimension is the most immediate, and basic psychological dimension on which people can easily locate any stimulus. The valence dimension is so important because it is closely related to behaviors such as approach, fight, or flight reactions which ensure survival in humans as well as in animals.

1.1 Good and bad are not symmetrical

At first glance, positive and negative, good and bad can be interpreted as two symmetrical poles of the same psychological dimension called valence. They are exact opposites like hot and cold, hard and soft or black and white. However, as noted by philosophers as well as psychologists, this symmetrical nature of good and bad seems not to hold. Schopenhauer (1844) noted that good things in life (e.g. pleasures) are often not more than the absence of bad things (e.g. pain). In Psychology, it is by now a well-known fact that positive and negative as the two dual valence poles vary in various respects. These asymmetries can be observed on different levels of psychological processing. The most general difference is that "bad is stronger than good" meaning that negative stimuli have a stronger psychological impact than positive stimuli (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001).

Research on basic learning principles shows that learning with negative reinforcement is faster than learning with positive reinforcement. Traumatic avoidance learning for example, occurs after a single electric shock (Solomon & Wynne, 1954) and eating something nauseating only once leads to an immediate taste aversion (Garb & Stunkard, 1974; Pelchat & Rozin, 1982). Likewise, negative events have a greater impact on people's mood than positive events (Taylor, 1991). Negative stimuli (e.g. cockroaches) are also highly contagious as they turn a positive stimulus (e.g. food) negative, while the reversed is not true (Rozin & Royzman, 2001). A similar "negativity bias" underwent extensive research in the domain of impression formation where the simultaneous presence of a positive and negative attribute leads to a negative overall impression (Asch, 1946; Skowronski & Carlston, 1989; Fiske, 1980; Peeters, 1971).

Not only does negative information have a greater impact but it also captures more attention (Pratto & John, 1991). Consistent with the stronger impact of negative stimuli, they also enjoy a memory advantage. In recognition memory for example, people are better at discriminating old from new stimuli when they are negative compared to when they are positive (Inaba, Nomura, & Ohira, 2005; Ohira, Winton & Oyama, 1998; Ortony, Turner & Antos, 1983; Robinson-Riegler, & Winton, 1996).

1.2 Valence (asymmetries) in the intrinsic and extrinsic world.

Given the importance of the valence dimension for human functioning, the question arises at what level it exists. From a strictly constructivist position, valence is an inherently psychological, and therefore subjective dimension that exists inside the mind only. As valence is preceded by some kind of evaluation which requires a human being that gives meaning first, there should not be any “objective” valence. However, as argued by Garner (1974) and others, the intrinsic world has a correspondence in the extrinsic world outside the human mind. Similarly, Gibson (1977) argued that stimuli themselves have specific affordances that trigger certain reactions or evaluations in the human being. From this “ecological” perspective, a psychological dimension such as valence might have its origin in the ecology surrounding the human mind. These two different perspectives can also be applied to the origin of valence asymmetries. The question whether valence asymmetries are intrapsychic or extrapsychic phenomena (Fiedler, 2014) is a central theme of my dissertation. I will argue that even though most psychological theories place valence asymmetries inside the human mind, many of them (if not all) can ultimately be traced back to ecological factors that lay outside the human mind and that are inherent in the information ecology itself.

1.2.1 Valence asymmetries as intrapsychic phenomena.

As described above, many of the most prominent valence asymmetries can be summarized as “bad is stronger than good” (Baumeister et al., 2001). Negative information’s stronger psychological impact as evident in the various phenomena described thus far is commonly thought to originate within the human being’s information processing system. The basic idea is that negative stimuli elicit and undergo deeper, more thorough processing as the result of internal affective reactions. The functional explanation for this processing advantage is traced back to evolutionary principles of survival. Accordingly, negative stimuli pose a potential threat to the organism which is why denoting special attention and deep processing to them constitutes a survival advantage for the organism (Taylor, 1991). The same idea is expressed in theories that describe processing differences under different affective states. Evidence form Clark and Isen (1982) suggests that negative mood leads to more effortful information processing compared to positive mood. While this notion is very similar to the idea that negative stimuli are processed deeper, it seems to provide evidence for a common intrapsychic mechanism underlying negativity biases. If deeper processing can be triggered independently from the stimulus via the induction of negative mood, the origin of the processing asymmetry should be intrapsychic. The most direct evidence for this idea comes from recent work by Topolinski and Deutsch (2013) who were able to vary phasic affective states on a trial to trial basis which then influenced processing of subsequent stimuli.

Bless and Fiedler provided a model that integrates affect-based processing differences into a general framework that distinguishes assimilative and accommodative processing styles (Bless & Fiedler, 2006; Fiedler, 2001). Accordingly, positive mood/affect leads to assimilative, top-down processing which relies on pre-formed internal schemas and heuristics while negative mood/affect leads to accommodative, bottom-up processing that is sensitive to the details of the external world. The functional explanation for this asymmetry is that affective states tune the organism to the demands of a given situation. Positive mood signals a benign environment in which the individual can afford to adopt an assimilative mindset which allows for abstraction and creative thinking (e.g. Isen & Daubman, 1984; Isen, Daubman, & Nowicki, 1987). Negative mood on the other hand, signals a potentially threatening environment in which the individual must adopt an accommodative mindset that gives rise to analytic, concrete, and stimulus-driven processing, and thereby prepares for defense reactions (e.g. Forgas, 2001).

1.2.2 Valence asymmetries as extrapsychic phenomena.

According to the perspective reviewed above, processing asymmetries elicited by positive and negative stimuli emerge as a result of an affective reaction within the information processing individual. The affective reaction is thereby a truly intrapsychic explanatory concept with strong intuitive appeal as it corresponds to people’s subjective experiences. In anticipation of what Garner (1974) later called the “extrinsic” world, Brunswick (1955) and Lewin (1951) argued that psychological theories should attempt to find the origin of intrapsychic processes outside of the human mind. They were primarily referring to the information environment in which c stimuli display certain “objective” properties. According to this perspective, the structure of the stimulus environment, or “ecology”, imposes constraints on the human mind, and ultimately determines intrapsychic processes (see also Fiedler, 2014). Examples for stimuli’s ecological properties are their frequency of occurrence and co-occurrence with other stimuli. These two properties are particularly important to understand valence asymmetries as extrapsychic phenomena. Referring back to the above described stronger impact of negative information during impression formation, several researchers have argued that at least part of it can be explained by negative information’s rarity in the real world. Fiske (1980) for example, argued that negative information receives greater weight because it is unexpected. Similarly, Skowronski and Carlston (1989) suggested that negative information is particularly diagnostic regarding a stimuli’s category membership, which is a function of the co-occurrences of different stimuli. The co-occurrence of stimuli in the environment is in fact the basis for most learning theories. It explains how stimuli derive meaning and how meaning is transformed from stimulus to stimulus. Consider the picture of a gun, which derives its affective potential because guns tend to co-occur with other threatening stimuli. Thus, the affective reaction to valenced stimuli itself might to some degree relate back to “objective” properties of the stimulus environment. Interestingly, many psychological concepts including that of an affective reaction tend to dismiss such an ecological perspective (Fiedler, 2000). In fact, ecological properties of information are often considered confounders in experimental set-ups. They are not of interest themselves but need to be controlled for, or are completely leveled out in order to provide the experimenter with a clear view on the “true” psychological processes at play. From the ecological perspective on the other hand, the confounders themselves are worth being investigated as they constitute ecological realities that directly influence human information processing. Experimental designs with stimuli that were pre-selected to control for natural confounders thereby lack ecological validity. For example, positive and negative stimuli typically appear with the same frequency in cognitive paradigms such as evaluative priming, conditioning or association tests. By doing so, researchers create mini-ecologies that do not correspond to real world ecologies in which negative stimuli appear a lot less frequently than positive stimuli.

A truly ecological theory should satisfy two main conditions as argued by Fiedler (2014). First, the explanatory concept should be located in the information environment (extrapsychic) and not within the individual (intrapsychic). Second, the process-determining variable should be “objectively” measurable. In order to test an ecological theory experimentally, a third condition must be added, namely that the ecology along with its process-determining variable should be manipulable. In the following, I will introduce the “Density Hypothesis” as an ecological theory of valence asymmetries that satisfies all three conditions. The theory suggests that positive stimuli are more similar to one another than negative stimuli. It further argues that this density asymmetry is the driving force behind various observable valence asymmetries that are typically attributed to internal affective reactions. After introducing the density hypothesis in detail I will present supporting empirical evidence. In a next step, I will introduce an explanatory framework which predicts that positive information is generally less diverse than negative information as the result of the non-extremity of positive qualities. In chapter 2, I present empirical work that tested the density hypothesis in the domain of person perception. The following chapter 3 presents an empirical investigation of one of the theory’s implications, namely its potential to account for valence asymmetries in recognition memory. In chapter 4, I discuss limitations, open questions, and implications of the density hypothesis and put it in a meta-theoretical context.

1.3 The density hypothesis

The density hypothesis argues that positive stimuli tend to be more similar to one another than negative stimuli. It thereby describes a structural property of the information environment that is general in the sense that it applies to various stimulus domains and that is robust meaning that it reliably occurs in given stimulus sub-samples. The density hypothesis was first formulated by Unkelbach and colleagues in 2008. It was derived to account for the effect that positive information is processed faster than negative information. This phenomenon was observed in evaluative priming where participants were faster at categorizing positive targets following positive primes than categorizing negative targets following negative primes. An intrapsychic explanation could again refer to the affective potential of negative stimuli that trigger deeper, and therefore more time-consuming processing. However, the authors discovered that in a sample of evaluative word stimuli frequently used in evaluative priming research (Fazio, Sanbonmatsu, Powell, Kardes, 1986; Bargh, Chaiken, Govender, & Pratto, 1992; Klauer & Musch, 1999), the positive stimuli were more similar to one another than the negative stimuli. Inter-stimulus similarity was measured by asking participants to provide pairwise similarity ratings which were analyzed using a multidimensional scaling procedure (MDS, Torgerson, 1965; Krumhansl, 1978). The MDS procedure estimates the coordinates of the stimuli in an *n*-dimensional space so that the Euclidean distances between stimuli are proportional to their similarities. Figure 1 depicts the 2-dimensional MDS solution from Unkelbach and colleagues (2008). The MDS procedure as implemented by the authors applies alternating least squares, an iterative procedure that alternates between applying permissible data transformations and estimating stimulus coordinates, while reducing a least squares measure of fit (Nesselroade & Cattell, 1988). Similar to factor analysis procedures, the meaning of the resulting dimensions are subject to the researcher's interpretation. As shown in Figure 1, the first dimension underlying the similarity structure of the words used by Unkelbach and colleagues (2008) is clearly a valence dimension, while the interpretation of the second dimension is less clear. It seems to combine various types of inter-stimulus similarity. I will return to the question of similarity content in chapter 4. For now, the important finding is that the positive words are more proximal to one another than the negative words suggesting that positive information is clustered more densely in mental representations than negative information. Using regression analysis, the authors then showed that inter-stimulus density could account for the faster processing of the positive words. This provided first evidence for the idea that inter-stimulus similarity as an external ecological property of evaluative information might account for effects that were formerly believed to originate in internal affective reactions.

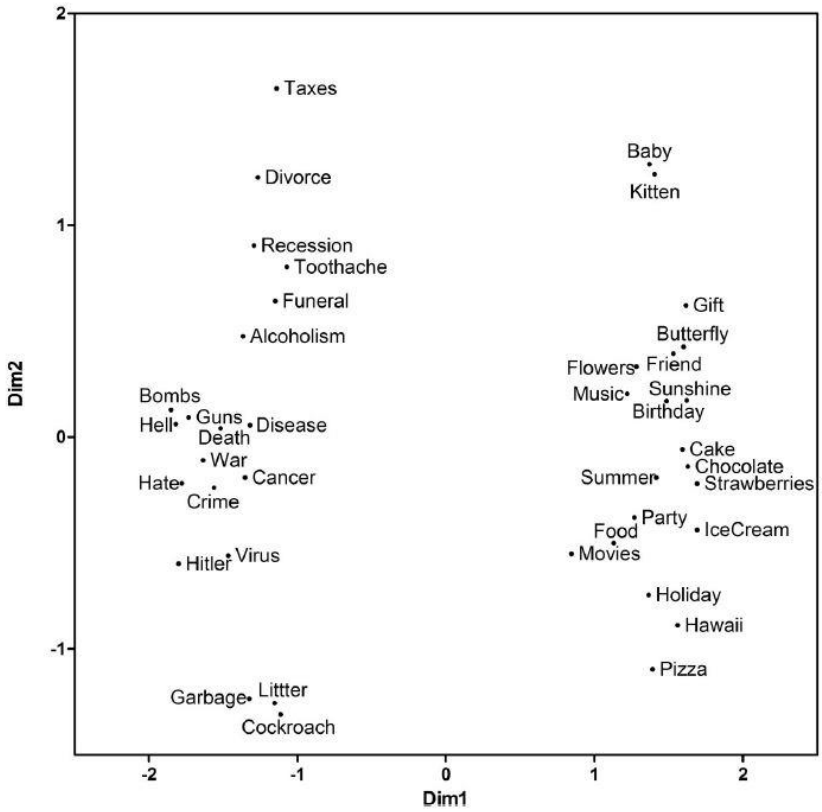

Figure 1. Positive and negative word stimuli in a 2-dimensional similarity space. The positive words are more densely clustered than the negative words (from Unkelbach et al., 2008).

Of course, there remained several open questions at this point that limited both, the generalization of the phenomenon itself as well as the formulation of an underlying theoretical framework. First, it was unclear whether the phenomenon itself would generalize to more exhaustive samples of word stimuli and to other stimulus domains. Second, as stimulus density was determined based on participants' subjective similarity ratings it was questionable to what extent stimulus density constituted an ecological, let alone objective variable. In other words, the greater diversity of negative stimuli could in fact mirror an objective property of the environment, but it could also be the result of a human mind that is tuned to denote special attention and differentiation to potentially threatening stimuli. Third, as the authors did not manipulate stimulus density experimentally, the evidence for its causal influence on downstream processing asymmetries remained correlations at this point. In the following sections, I will present findings from perceptual research, as well as recent empirical evidence from our own lab that suggests the density asymmetry is in fact a robust, general, and ecological phenomenon.

1.3.1 Perceptual prototypicality.

Evidence for the idea that positive stimuli are objectively more similar than negative stimuli can be found in perceptual research. A widely observed phenomenon in perception is that people prefer stimuli that are similar to certain prototypes (e.g. Halberstadt & Rhodes, 2003; Winkielman, Halberstadt, & Fazendeiro, 2006). As stimuli become more similar to a prototype they become more likeable, and importantly, they also become more similar to one another. The simple reason is that a prototype constitutes a rather specific evaluative-perceptual “end state” from which a stimulus can deviate in many different ways. This principle is especially illustrative in facial attractiveness. That is, faces become more attractive the more they possess “average”, or non-extreme features, thereby resembling a cultural ideal. Research has shown that the morph of a given number of faces is typically judged as more attractive than the individual faces (Langlois & Roggman, 1990; Rhodes, 2006). As a result, attractive faces are more similar to one another than unattractive faces (Potter, Corneille, Ruys, and Rhodes, 2007). These findings from perceptual research are particularly noteworthy because they show that desirable stimuli such as faces are more similar to one another regarding objectively quantifiable attributes such as nose length, eye position, ear size, and so on. I will later argue that the underlying principle according to which positive qualities as “non-extreme” on given attribute dimensions can serve as an explanatory framework for the density asymmetry in general, not only in the psycho-perceptual domain.

1.3.2 Recent empirical evidence.

In a recent line of research (Koch, Alves, Krüger, & Unkelbach, 2016), we built on and extended the seminal work by Unkelbach and colleagues (2008). We aimed at overcoming the empirical shortcomings that limited the conclusions that could be drawn from the initial evidence for the density asymmetry. Our goal was to answer three main questions. First, is the asymmetry general in the sense that it reliably occurs for different stimulus classes? Second, is there evidence suggesting that positive information is objectively more similar than negative information? Third, does the density asymmetry cause other valence asymmetries in the processing of positive and negative information?

A general phenomenon? To answer the first question we had to use stimulus sets that were more exhaustive and more representative regarding the universe of evaluative stimuli. The initial evidence for the density asymmetry by Unkelbach and colleagues (2008) rested on 40 word stimuli, taken from a larger set of 92 words which were compiled in a rather arbitrary fashion (Fazio et al., 1986). We used two strategies to create larger and more representative stimulus samples. In 4 studies, participants generated positive and negative word stimuli before they judged the similarities of their own, or of other participants’ word stimuli. Another study featured all positive and negative words from a database by Warriner, Kuperman and Brysbaert (2013) that contained around 14000 words. To test whether the density asymmetry extends to different stimulus domains, another study asked participants on seven consecutive days to name positive and negative events they had experienced before they rated their similarities. Yet another study featured around 1000 positive and negative pictures from the international affective pictures system (IAPS; Lang, Bradley, & Cuthbert, 2005).

Regarding the similarity measurement, we asked participants in most studies to spatially arrange the stimuli on a computer screen according to their similarities. This spatial arrangement method (SpAM; Goldstone, 1994; Kriegeskorte & Mur, 2012) is an effective method to obtain the similarities for large stimulus sets as it is a lot less time consuming then pairwise similarity ratings. In addition, the resulting inter-stimulus similarity estimates are highly consistent with those from pairwise similarity ratings (Hout, Goldinger, & Ferguson, 2013). This was also empirically confirmed in study one of our research line in which we used the original 40 word stimuli by Unkelbach and colleagues (2008) and found a strong correlation (*r* = .84) between similarity estimates from SpAM and those based on pairwise similarity ratings. Importantly, across all different stimulus sets that participants spatially arranged in 6 different studies, positive stimuli were always judged to be more similar to one another than negative stimuli. For the words from the large database by Warriner and colleagues (2013) and for the IAPS pictures, we calculated similarity indices based on their provided norm ratings. Again, the positive words as well as the positive IAPS pictures were more similar to one another than the negative ones.

An objective phenomenon? Relying on similarity estimates based on pairwise similarity ratings or spatial arrangement does not provide much evidence for the claim that the density asymmetry is an objective phenomenon of the extrinsic world. Similarity judgments themselves could be influenced by internal affective reactions within the perceiver elicited by the evaluative stimuli (e.g. Topolinski et al., 2013). As a more objective measure for inter-stimulus similarity we identified frequency of co-occurrence (e.g. Griffiths, Steyvers, & Tenenbaum, 2007; Jones & Meworth, 2007). Specifically, we measured how often all pairwise combination of the 40 words used by Unkelbach and colleagues co-occurred among web sites and among text passages from a collection of 738 books representative of the literature read by U.S. college students (Landauer & Dumais, 1997). Frequencies of co-occurence on web sites were determined by entering all pairwise word combinations into the Google search engine which then provided an estimated number of hits for each pair (see Cilibrasi & Vitanyi, 2007 for a similar approach). The resulting frequencies were analyzed using the MDS procedure which again resulted in a density index for each stimulus. Co-occurrences in book passages were determined using latent semantic analysis (LSA; Landauer & Dumais, 1997). As this technique directly translates co-occurences into similarity indices, no MDS procedure was necessary. Both these similarity measures were substantially correlated with the subjective similarity measures (pairwise and SpAM). The correlations between Google similarity and pairwise ratings was *r* = .56 and so was the correlation between Google similarity and SpAM ratings. Likewise, LSA similarity was substantially correlated with pairwise ratings (*r* = .73), and with SpAM ratings (*r* = .64). Importantly, the Google similarity measure as well as the LSA measure both confirmed the density asymmetry meaning that positive words occur in the same context more often than the negative words. While this shows that the density asymmetry has an objectively measureable manifestation in people's information environment, it does not necessarily mean that it originates in the same. That is, positive information's stronger co-occurrence could still be a mere reflection of an ultimately subjective asymmetry. However, the collection of converging evidence for the density asymmetry with different measures and stimuli suggests that it might actually root in objective properties of the information environment.

A causal factor for valence asymmetries? After having shown that the density asymmetry is a general phenomenon that also has an objective manifestation in the information environment, we then aimed at finding out whether it is meaningful in the sense that it is a causal predictor of other downstream processing asymmetries. We predicted and found five valence asymmetries for the 40 word stimuli used by Unkelbach and colleagues (2008) that occurred as a result of positive stimuli's higher density. The first two predictions concerned processing speed in evaluation and classification. We found that participants were faster at evaluating positive compared to negative words on a scale and that they were faster at classifying them as "positive" or "negative". The third and fourth predictions concerned recognition memory. We predicted and found that participant's recognition memory is more sensitive towards negative compared to positive words and that positive words elicit a stronger response bias. These empirical investigations are presented in chapter 3. The fifth prediction addressed categorization and we found that positive words are more likely to be subsumed under the same category than negative words. All of the five outcome measures were predicted by all four similarity measures. All predictions reached significance, except for Google similarity predicting recognition memory and categorization, and LSA similarity predicting categorization. This suggests that density might actually account for the valence asymmetries in all five outcome measures. For processing speed, this was already confirmed by Unkelbach and collagues (2008) when they showed that density predicted processing speed over and above valence. For recognition memory, the same point will be made in chapter 3. Participant's tendency to sort positive words into fewer categories than negative words was also predicted by density (here: pairwise ratings; $\beta_{density} = -.64$, $p = .003$) which again fully accounted for the effect of valence ($\beta_{valence} = -.18$, $p = .356$).

The fact that stimulus density predicts these cognitive processing variables and accounts for the effect of valence not only shows density's predictive superiority. It also adds to our idea that the density asymmetry is an ecological phenomenon and not the result of internal affective reactions. If the latter was true, the processing asymmetries should be a direct function of a stimulus valence as it should be the best proxy for its affective potential. In other words, the same affective potential causing participants to judge positive stimuli as more similar should also cause them to process positive stimuli faster, falsely remember them and to sort them into fewer categories. As this was not the case, I consider it likely that the density asymmetry constitutes an ecological reality that is independent of a stimulus' affective potential and therefore independent of participants' internal affective reactions. This reasoning will be discussed in more detail in chapter 3 which addresses valence asymmetries in recognition memory.

While I have reviewed converging evidence for the generality, objectivity, predictive power, and ecological nature of the density asymmetry, it lacks an explanatory concept at this point. If the asymmetry is an ecological property of the information environment the question remains why it exists. Why is positive information less diverse than negative information? In the following, I will present an explanatory model that builds on some ideas first introduced by Unkelbach and colleagues in 2008. Since then we have further elaborated on these ideas and integrated them into a comprehensive framework. Please note that the following part is partly redundant with the theoretical introductions in chapters 2 and 3.

1.3.3 Explanatory model.

The core principle of the present explanatory framework is that positive attributes are usually non-extreme. On most attribute dimensions, the positive range is located towards the middle of the dimension and is surrounded by two negative ranges towards the two ends of the dimension (see top part of Figure 2). Thus, for any quality that an object can possess there is usually only one positive range surrounded by two negative ranges. From this simple assumption, it follows that negative information should generally be more diverse than positive information. To clarify, this greater diversity has two manifestations. First, there should be more distinct negative than positive stimuli in the ecology. While one positive stimulus would fall in the positive range, the two negative ranges can each host another distinct negative stimulus. Second, in line with the density hypothesis, positive stimuli should be more similar to one another than negative stimuli because the distance between the two negative ranges always exceeds the distance within the positive range. While two positive objects necessarily have to lay within the same range, two negative objects can lay in two different ranges on a given attribute dimension that are highly distant from each other.

The non-extreme nature of positive qualities is not domain-specific but seems to be true across many if not all levels of observation. To start with a remote example, life is possible only within a single range of temperature, oxygen concentration, blood glucose level, and so on. For all these physical, chemical and biological dimensions there is a "too little" and a "too much". Thus, the diversity of unlivable conditions is greater than the diversity of livable conditions. Note that this reasoning is similar to the perceptual principles described above. Attractive faces are similar because they consist of features (e.g. nose) that lay within the same single range on a given attribute dimension ("good" nose size), while unattractive faces consist of features that can lay towards either ends of the dimension (nose is too long or too short). I argue that this principle is not restricted to the psycho-perceptual domain, but that it applies to the conditions that allow life to exist on physical, chemical, and biological levels. It is safe to assume that such a powerful principle reaches into psychological realities as well. This idea was expressed by no less a thinker than Aristotle:

> Both excessive and defective exercise destroys the strength, and similarly drink or food which is above or below a certain amount destroys the health, while that which is proportionate both produces and increases and preserves it. So too is it, then, in the case of temperance and courage and the other virtues. For the man who flies from and fears everything and does not stand his ground against anything becomes a coward, and the man who fears nothing at all but goes to meet every danger becomes rash; and similarly the man who indulges in every pleasure and abstains from none becomes self-indulgent, while the man who shuns every pleasure, as boors do, becomes in a way insensible; temperance and courage, then, are destroyed by excess and defect, and preserved by the mean.
> - Aristotle, (trans. 1999, p.229)

Aristotle precisely described the principle of negativity as deviations from non-extreme ideals which applies across various domains ranging from physical intake of food to psychological trait concepts. The same principle is mirrored by the fact that most human attributes including physical features, personality traits and abilities are normally distributed as first described by Francis Galton (1869; Simonton, 2003). The normal distribution indicates that non-extreme values on given attribute dimensions are more optimal for survival than extreme values assuming that attribute distributions are the result of Darwinian natural selection (Biktashev, 2014; Kimura, 1965; Lande, 1975).

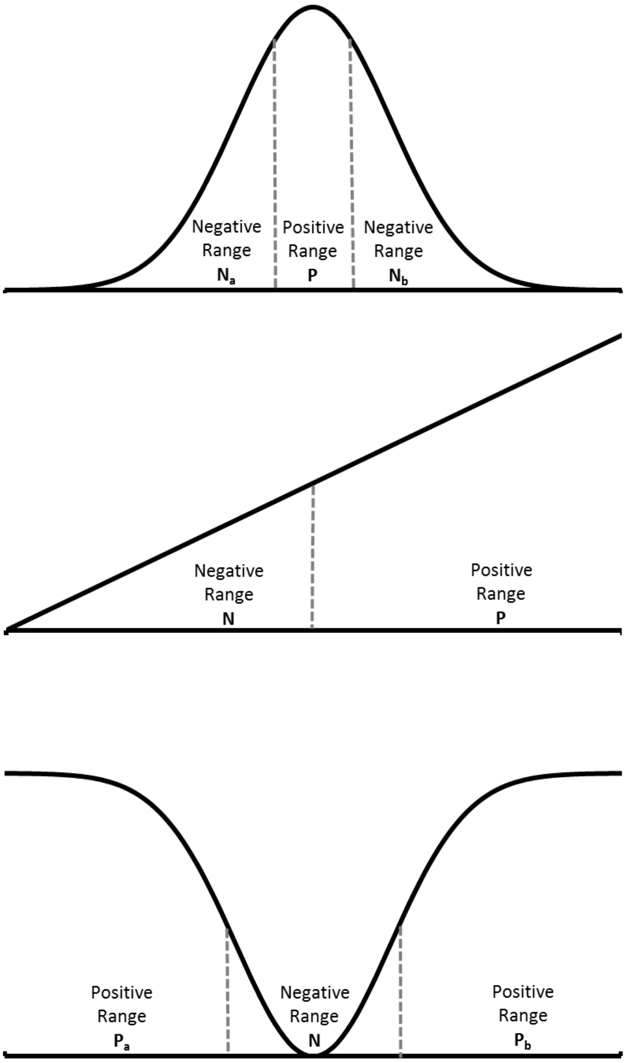

Figure 2. The curves depict three possible relations between positivity (y-axis) and a given attribute dimension (x-axis). The top part illustrates the common curvilinear relationship where non-extreme attribute values are positive, leading to a single positive range opposed by two negative ranges. In the middle part, positivity is a linear function of the attribute dimension, meaning that the attribute has a positive and a negative pole. This rather seldom relation results in one positive and one negative range. The lower curve illustrates the rarest case in which extreme attribute values are positive, resulting in two positive and only one negative range.

There might be some exceptions to the principle described. Some attribute dimensions might have a clear negative and a clear positive pole as illustrated in the middle part of Figure 2. Consider for example the concentration of poison in your blood. Here, the less the better should be the rule, which is why on a “poison concentration” dimension there is one positive and one negative range. While such attribute dimensions certainly exist, I argue that they are a lot less common than the one depicted in the top part of Figure 2. Again, most attributes are normally distributed. Even on attribute dimensions that allegedly have one positive and one negative pole (e.g. agreeableness), the positive range often reaches inflection points at which its effects turn negative (Grant and Schwartz, 2011). Even more unlikely to find is the reversed relation between positivity and attribute extremity as illustrated in Figure 2‘s lower part. On such an attribute dimension, extreme values in both directions would be positive while moderate values would be negative. While such attributes might exist (I do not know of any), they are certainly not as numerous as to make up for the prevalence of attributes with a non-extreme range of positivity (top part of Figure 2).

If we accept that most attribute dimensions relate to positivity in an inverted u-shaped manner, greater diversity of negative information should follow. How the diversity asymmetry unfolds when multiple attribute dimensions are considered can be illustrated in both, a spatial model of similarity (e.g. Torgerson, 1965) as well as in a feature-based model of similarity (e.g. Tversky, 1974). Note that the empirical evidence for the greater diversity of negative information reviewed above (Koch et al., 2016; Unkelbach et al, 2008) is based on a spatial model. This is mirrored by the description of the phenomenon as a “density” asymmetry, which directly refers to the representation of information in mental space.

Figure 3, taken from Koch and colleagues (2016), illustrates how the non-extremity of positivity leads to lower diversity using a spatial model of similarity. Spatial models assume that stimuli can be located in a space emerging from a given number of attribute dimensions. The similarities between two stimuli are given by the Euclidean distance between them, and more distant stimuli are less similar. By combining three attribute dimensions in Figure 3, a positive end-state emerges in the form of a bubble. The positive bubble is surrounded by a negative layer that occupies more space than the positive bubble. This means that the negative layer can host *more* distinct negative stimuli than the positive bubble and that the average Euclidean distance between given positive stimuli (P1, P2) is smaller than between negative stimuli (N1, N2).

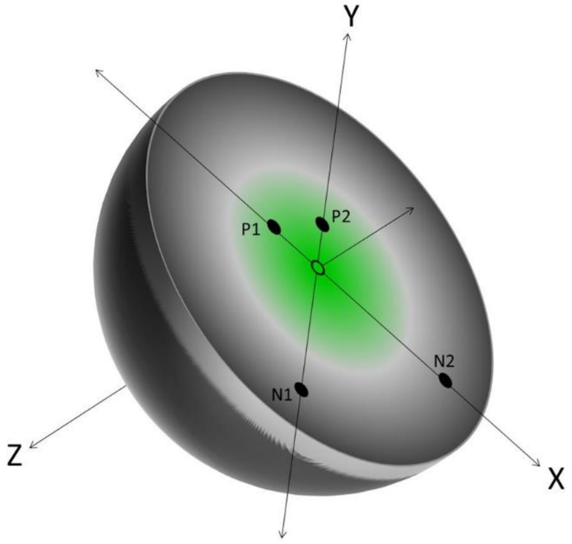

Figure 3. (from Koch, Alves, & Unkelbach, in prep.). Relationship between inter-stimulus similarity and stimuli's affective potential in a hypothetical 3-dimensional space. X, Y, and Z are stimulus dimensions while the positive range of each dimension lays towards the middle, surrounded by two negative ranges. P1 and P2 as well as N1 and N2 show positive and negative stimuli, respectively. The green area constitutes the positive stimulus space, while the grey area constitutes the negative space. The stimuli's affective potential is given by their beeline Euclidean distance to the neutral white layer. The inter-stimulus similarities are given by the beeline Euclidean distance between stimuli. Here, the affective potential is equal for all four stimuli P1, P2, N1, and N2, while the intra-valence similarity is larger for P1 and P2 than for N1 and N2.

The same principle can be illustrated in a feature-based model of similarity. It assumes that entities (e.g. objects or persons) can be described as collections of absent and present features. The similarity between two given stimuli is determined based on the number of shared (matching) and unshared (non-matching) features. To derive features from attribute dimensions, we simply consider each range in Figure 2's top part a distinct feature, resulting in one positive and two negative features per dimension. By combining three attribute dimensions we arrive at 3 positive (grey rectangles) and 9 negative features (black rectangles) as illustrated in Figure 4's left half. This again implies that there are more distinct negative than positive qualities. Regarding inter-stimulus similarity, Figure 4's right half illustrates the consequence for the feature-matching probability of two positive stimuli and of two negative stimuli. The positive stimuli (P1, P2), each consisting of two positive features (filled grey rectangles) and one negative feature (filled black rectangles), have a larger feature-matching probability than two negative stimuli (N1, N2), consisting of two negative and one positive feature. While the two positive features in this example share two features, the two negative stimuli only share one feature.

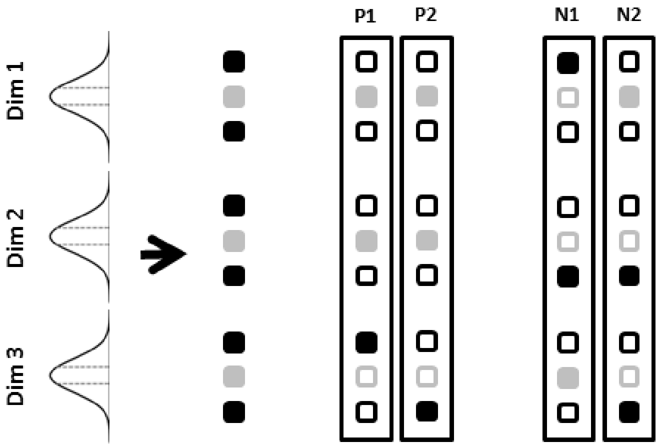

Figure 4. A feature-based illustration of negative information's greater diversity. The left part translates the non-extremity of positive qualities into a feature logic, resulting in twice as many negative features (black rectangles) than positive features (grey rectangles). The right part shows two positive (P1, P2) and two negative stimuli (N1, N2) and their respective feature profiles. Filled rectangles symbolize present features, unfilled rectangles symbolize absent features. Two positive features and one negative feature are present in the positive stimuli and two negative features and one positive feature are present in the negative stimuli. The resulting feature matching probability is larger for the positive than for the negative stimuli.

Hence, negative information's greater diversity (larger number of distinct stimuli and lower inter-stimulus similarity) should follow from the non-extremity of positive qualities independent of whether one conceptualizes similarity spatially or as a feature-matching process. The present explanatory framework thereby well accounts for the observation that positive stimuli are more similar to one another (Koch et al., 2016; Unkelbach et al., 2008). The framework's second prediction, namely a greater number of negative compared to positive stimuli, is also supported by empirical evidence from different psychological domains. When considering emotional experiences, it becomes evident that there are more distinct negative than positive emotions. Ekman (1971) for example, distinguished between four basic negative emotions (sadness, anger, fear, disgust) but described only one basic positive emotion (joy). While the exact number of positive and negative emotions varies depending on the theoretical model, they typically describe more negative than positive emotions (see also Ortony & Turner, 1990). Moreover, the greater differentiation of negative information is visible in language as there are more negative than positive words, which was shown for German, Spanish, and English words (Leising, Ostrovski, & Borkenau, 2012; Schrauf and Sanchez, 2004; Semin and Fiedler, 1992).

1.4 The Current Research

I propose that negative information is more diverse than positive information and that this constitutes a robust, general, and objective property of the information environment that accounts for downstream processing asymmetries. In the following two chapters, I present two lines of research that demonstrate a manifestation as well as a consequence of the density hypothesis.

In chapter 2, I apply the density hypothesis to person perception and predict that people perceive liked others as more similar than disliked others. This constitutes an especially strong test of the density hypothesis because it runs counter to a strong law of mental differentiation. That is, liking breeds differentiation as people collect more knowledge about objects and persons they like (e.g. Smallman & Roese, 2008). This well-documented phenomenon suggests a general dislike-homogeneity effect while the density hypothesis predicts the opposite based on the lower diversity of positivity.

In chapter 3, I present a line of research that tested the density hypothesis' potential to account for downstream valence asymmetries. Specifically, I predicted that negative information's recognition advantage is due to its greater diversity. While this demonstrates the predictive power of the density hypothesis, it also includes a first attempt to experimentally manipulate stimulus density. Note that so far demonstrations of the density hypothesis' potential to account for downstream processing asymmetries rested solely on correlative evidence (Unkelbach et al, 2008; Koch et al., 2016). As mentioned earlier, in order to establish a meaningful ecological theory of valence asymmetries, the causal variable (here: density) should be experimentally manipulable. Certainly, density, or inter-stimulus similarity, can be manipulated independent of stimulus valence. In doing so, we could experimentally test whether downstream processing asymmetries (here: recognition memory) are a function of valence or density.

Chapter 2 – My friends are all alike – the relation between liking and perceived similarity in person perception

Abstract

Past research showed that people accumulate more knowledge about other people and objects they like compared to those they dislike. More knowledge is commonly assumed to lead to more differentiated mental representations; therefore, people should perceive others they like as less similar to one another than others they dislike. We predict the opposite outcome based on the density hypothesis (Unkelbach, Fiedler, Bayer, Stegmüller, & Danner, 2008); accordingly, positive impressions are less diverse than negative impressions as there are only a few ways to be liked but many ways to be disliked. Therefore, people should perceive liked others as more similar to one another than disliked others even though they have more knowledge about liked others. Seven experiments confirm this counterintuitive prediction and show a strong association between liking and perceived similarity in person perception. We discuss the implications of these results for different aspects of person perception.

It seems evident that liking breeds differentiation. Wine lovers can differentiate between a Merlot and a Syrah, art enthusiast see the differences between a Monet and a Renoir, and soccer fans can distinguish between the playing styles of Lionel Messi and Thomas Müller. In social psychological research, this notion is found in broad phenomena like the outgroup homogeneity effect (Quattrone & Jones, 1980; Park & Rothbart, 1982) or the cross-race effect (Feingold, 1914; Young, Hugenberg, Bernstein, & Sacco, 2012) – people differentiate better between members of their usually preferred in-groups and between faces of their own ethnic identity compared to outgroup members and faces from other ethnicities. Liking breeds differentiation because information sampling follows a hedonic principle (e.g., Fazio, Eiser, & Shook, 2004; Thorndike, 1898). People seek interactions with persons they like and avoid interacting with disliked persons (Denrell, 2005). As a result, people's mental representations of liked others are highly differentiated as opposed to the rather shallow representations they have of disliked persons (e.g., Smallman & Roese, 2008). A direct implication of this liking-breeds-differentiation principle is that social perceivers should see liked persons as more diverse while disliked persons should all seem alike.

Despite the intuitive appeal of a general dislike-homogeneity phenomenon, a different line of research suggests that liking might go along with increased perceived similarity. According to the "Density Hypothesis", positive information is less diverse and thus more densely clustered in spatial representations compared to negative information (Unkelbach et al., 2008; Unkelbach, 2012). Similar to the principle observable in facial attractiveness, there are only a few possible ways to be liked but many different ways to be disliked (Potter, Corneille, Ruys, & Rhodes, 2007). Here, we apply this principle to person perception; based on the density hypothesis, we present a model that assumes perceived similarity among other people to be based on their matching and non-matching features. Because people should represent liked persons with predominantly positive features and because positive features are less diverse, a counterintuitive scenario follows: Social perceivers should see other people they like (e.g., their friends) as *more* similar to one another than other people they dislike even though they have collected more knowledge about liked others and therefore have a more differentiated representation of them. The goal of the present work is to examine the relation between liking and perceived similarity in person perception, and to test whether the liking-breeds-differentiation principle causes people to perceive liked others as more diverse, or whether the low diversity of positivity makes all liked others appear similar.

The nature of the relation between liking and perceived similarity has strong implications for social perception: Perceiving others as similar or different from one another is a determinant of many social cognitive processes including social categorization (Tajfel, Billig, Bundy, & Flament, 1971; Billig, Tajfel, 1973), generalization and stereotyping (Ames, 2004; Gawronski, & Quinn, 2013; Linville, Salovey, & Fischer, 1986), social comparison (Mussweiler, 2003), as well as person memory (Earles, Kersten, Curtayne, & Perle, 2008; Heathcote, Freeman, Etherington, Tonkin, & Bora, 2009).

In the following, we first introduce the concept of differentiation and explain why liking should breed differentiation, and accordingly, why social perceivers might perceive liked others as more diverse compared to disliked others. We then delineate the density principle, namely why positive information should display low diversity, and accordingly, why social perceivers might see persons they like as more similar, despite having more knowledge. The following empirical part presents data from seven experiments that systematically investigated the relation between liking and perceived similarity in person perception. Finally, we discuss implications from our research for different aspects of person perception such as mood effects, and social comparison processes.

2.1 Differentiation and evaluation

Social perception has two fundamental characteristics: it is driven by the process of differentiation, and its outcome is typically evaluative. Differentiation is a core concept of human perception and cognition, and it is essential for any kind of categorization and plays a particularly prominent role in social psychological theorizing. Seeing the differences and the similarities among individuals determines many aspects of social perception and behavior, such as perceived group membership (Campbell, 1958; Zárate & Sanders, 1999), social comparison (Mussweiler, 2003), or interpersonal interaction (Tajfel, 1982). For example, differentiation was described to be “at the heart of the stereotype concept” (Linville, et al., 1986, p.165) as stereotypes arise from a lack of differentiation between individuals (Park & Hastie, 1987). Consequently, much social psychological research is devoted to the perception of variability among group members (e.g. Judd & Park, 1988;

Judd, Ryan, Park, 1991; see Rubin & Badea, 2012 for an overview), including fascinating phenomena, such as the mentioned out-group homogeneity effect and the cross-race effect (e.g., Quattrone & Jones, 1980; Ostrom & Sedikides, 1992). While social perception is driven by differentiation, its outcome is typically evaluative. In order to navigate complex social environments, humans have to distinguish good from bad people and decide whom they like and whom they dislike (Lewin, 1935). In the present work we want to address how differentiation and evaluation, the two concepts most central to social perception, are related in people's mental representation of their social world.

More specifically, we ask whether people perceive others they like or others they dislike as more similar to one another. As introduced above, we believe there are two competing scenarios: (1) Liking goes along with increased perceived diversity (i.e., decreased perceived similarity); (2) Liking goes along with decreased perceived diversity (i.e., increased perceived similarity). We first review evidence for the first scenario building on the principle that liking breeds differentiation.

2.1.1 Liking-breeds-differentiation.

There is substantial evidence indicating that people have highly differentiated representations of the things and individuals they like. For example, people divide liked objects into more categories than disliked objects and use finer evaluative distinctions when expressing attitudes about liked vs. disliked stimuli (e.g., Smallman, & Roese, 2008; Smallman, Becker, Roese, 2014). In addition, research on the perception of group variability shows that perceptions of homogeneity tend to have negative associations while heterogeneity has positive associations. Groups that are perceived as homogenous "are usually low status, low power, minority groups, whose members are perceived in less individualistic terms, receive less attention, and display less positive emotions" (Badea, Brauer & Rubin, 2012, p.1094; see also Brauer & Bourhis, 2006; Fiske, Haslam, & Fiske, 1991).

But why should people have a more differentiated representation of the objects and persons they like? One answer is that human information sampling follows a hedonic principle, first described by Thorndike (1898) as the law-of-effect. Accordingly, people are more likely to engage in exploratory behavior (e.g., interactions with another person) if the behavior is expected to have positive consequences (see also Chen & Bargh, 1999; Fazio et al., 2004; Hirt, Melton, McDonald, & Harackiewicz, 1996). This hedonic principle of information sampling also serves the purpose of maintaining internal cognitive consistency (Festinger, 1954). As a result, people collect larger information samples (i.e., more knowledge) and derive a more differentiated representation about liked others than about disliked others (Denrell, 2005). Smallman and Roese (2008) explicitly expressed this idea as "to cherish a loved one is to relish the fine nuances of his or her personality" while " the rejected and forsaken are construed on a relatively surface level." (p. 1228).

As more differentiated representations typically go along with decreased perceived similarity (e.g. Goldstone & Steyvers, 2001; Nosofsky, 1986; Shepard, 1987), liked persons should be perceived as less similar to one another than disliked others. In line with this idea, Linville and colleagues (1989) showed that when people become more familiar with a person or a group, they perceive them as more diverse. Likewise, familiarity leads to more differentiated categorical constructions and evokes less generalizations among objects (Medin, Lynch, Coley, & Atran, 1997; Rota & Zellner, 2007).

Taken together, a large body of empirical research suggests that liking a person invites repeated exposure, increases the amount of knowledge, and thereby leads to a more differentiated representation of liked compared to disliked others, which indicates less similarity. From the perspective of a hedonic principle of information sampling, it seems evident that people should perceive liked others as diverse and disliked others as similar, constituting a general dislike-homogeneity effect.

2.1.2 Diversity of positive and negative impressions.

Despite the arguments reviewed so far, we believe that the seemingly obvious negative relation between liking and perceived similarity might not hold in most contexts. In fact, we suggest that liking often comes with increased perceived similarity. Clearly, information sampling follows a hedonic principle and people have rich representations of the people they like. However, these rich representations should entail predominantly positive information, while disliked others should be represented by negative information. Crucially, information valence is confounded with content diversity; negative information is more diverse than positive information. As a result, mental representations of different disliked people can be rather different while mental representations of different liked others should be rather similar.

The idea that positivity comes in low diversity was introduced as the “Density Hypothesis” by Unkelbach and colleagues (2008). The authors showed that positive stimuli display a higher density in spatial displays of mental representations than negative stimuli, an ecological phenomenon that is apparent across many different stimuli classes including evaluative words, self-generated nouns, trait words, IAPS-pictures, as well as facial features (Alves, Unkelbach, Burghardt, Koch, Krüger, & Becker, 2015; Bruckmüller & Abele, 2013; Koch, Alves, Krüger, & Unkelbach, 2015; Koch, Alves, & Unkelbach, 2015; Potter et al., 2007; Unkelbach, 2012; Unkelbach et al., 2008).

The limited diversity of positivity is not the result of biased information processing but reflects a robust principle of the world humans live in. Positive states usually constitute a norm state which is characterized by the absence of multiple different negative norm deviations (Clark & Clark, 1977). On any given dimension, the positive norm states typically occupies a single range close to the midpoint of the dimension, and is opposed by two negative deviations towards the two ends of the dimension. The top part of Figure 5 illustrates this principle. For example, humans are able to survive only within a narrow range of temperature, atmospheric oxygen concentration, and electromagnetic radiation given off by the sun. For each of these dimensions, there is a “not enough” as well as a “too much" leading to a greater diversity of unlivable compared to livable conditions. Likewise, the human body is constantly engaged in maintaining internal homeostatic conditions such as body temperature or the blood glucose level (Cannon, 1926). In general, life as we know it is possible only within tight boundaries. These examples might appear unrelated to psychological processes, but the same principle applies more generally to qualities humans prefer and therefore refer to as "positive". Facial beauty is an illustrative example of this principle as there are many ways to be ugly, but only one way to be beautiful (Potter et al., 2007; Rhodes, 2006).

If the low diversity of positivity can be applied to person perception, positive impressions should be less diverse than negative impressions. This suggestion is supported by research showing that "positive" personality profiles tend to be those that show average scores on personality dimensions, which is why the correlation between item means of personality tests and item desirability typically exceeds $r = .80$ (e.g., Edwards, 1953; Leising, Ostrovski, & Zimmermann, 2013).

Figure 5's lower half illustrates our application of the density hypothesis to person perception and why it should lead to greater perceived similarity among liked compared to disliked individuals. Let us assume three dimensions of social interactive behavior that one can observe during a social encounter (e.g., amount of talking, amount of laughing, amount of eye contact). Based on the reasoning above, the liked or "positive" range on each dimension is located around the midpoint of the dimension and constitutes a positive attribute, while there are two different negative ranges on each dimension that constitute two highly different attributes (e.g., barely talking vs. talking too much; not laughing vs. laughing all the time; no eye contact vs. constant eye contact). When combining the three dimensions, individuals can vary along three different positive attributes (grey rectangles) and along six different negative attributes (black rectangles). While a person can be perceived positively as talking, laughing, and making eye contact to an agreeable extent, he or she can be perceived negatively as incommunicative, *or* as a chatterbox; as humorless *or* clownish; and as making too little, *or* as making too much eye contact. Consequently, negative impressions come in greater diversity than positive impressions.

Figure 5 further illustrates the mental representation of four liked individuals that each are represented by two positive and one negative attribute, and the mental representation of four disliked individuals that are represented by two negative and one positive attribute. Assuming that similarity is a function of matched and non-matched attributes/features (Tversky, 1977), the mental representations of liked individuals necessarily include more matching attributes than the representations of disliked individuals.

This principle should apply to "objective" attributes that target persons possess as well as to "subjective" representations of these attributes. In Garner's (1974) terminology, the principle is inherent in the intrinsic as well as in the extrinsic structure of the social world. That is, persons that display more likeable attributes than others should factually be more similar to one another, while at the same time persons that are perceived as more likeable than others by a given observer should also be perceived as more similar to one another. In the experiments that follow we investigate the latter case, that is, the subjective perception of similarities among liked and disliked others.

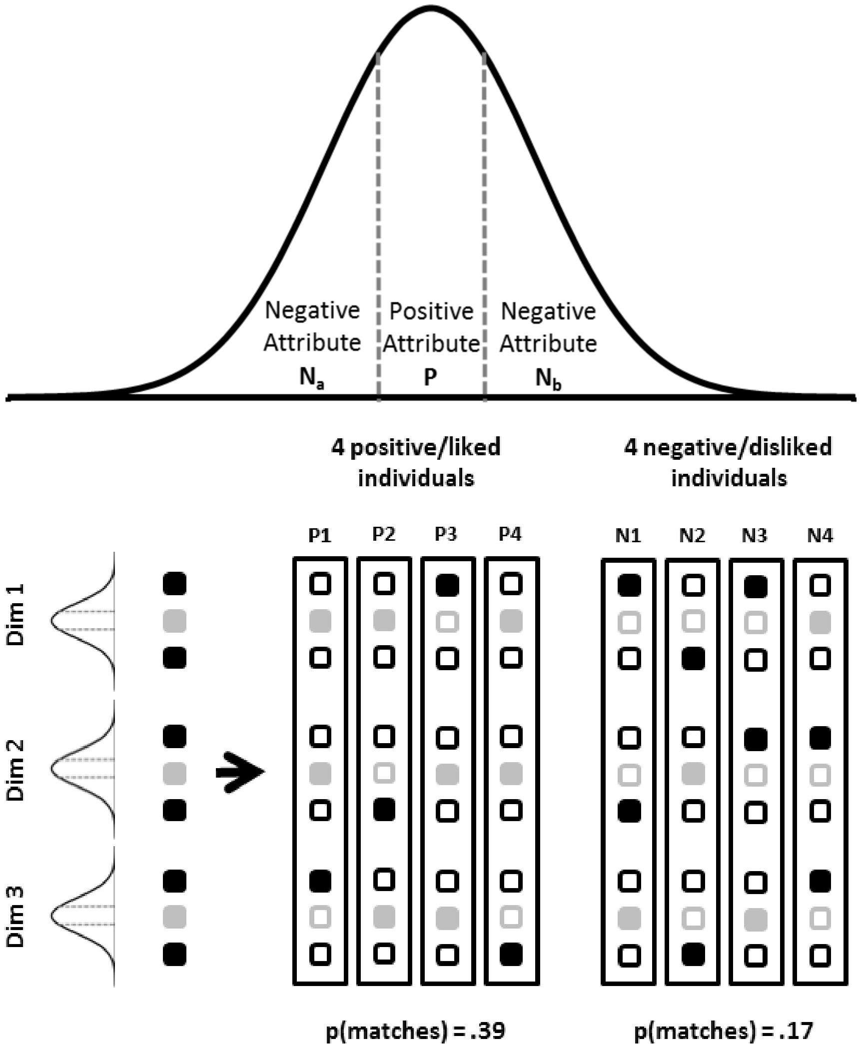

Figure 5. The top part illustrates a normal distribution of values on a single dimension. Any positive quality is typically located in the center of the distribution and is opposed by two different negative qualities. The bottom part illustrates how this principle leads to higher similarity among positive/liked individuals compared to negative/disliked individuals when three attribute dimensions are combined. On each dimension, there is one positive attribute (grey square) and two negative attributes. Individuals that have mostly positive attributes (grey sqares) vary along less attributes than individuals that have mostly negative attributes. In this example, the positive individuals on average have 1.16 out of 3 attributes in common, indicating a matching probability of .39, while the negative individuals have only 0.5 out of 3 attributes in common which equals a

matching probability of .17. This principle can be applied to objectively present attributes as well as to subjective representations of other people's attributes.

There is already empirical support for our model from research showing that the vocabulary for describing disliked persons is larger and more differentiated than the vocabulary for describing liked persons (Leising, Ostrovski, & Borkenau, 2012). Furthermore, research from Leising and colleagues (2013) suggests that people like others for similar reasons but dislike them for different reasons, a phenomenon that logically follows from our model (see also Leising, Erbs, & Fritz, 2010). When two people have a positive impression about another person, their impressions are likely to be highly similar, while negative impressions can be quite different.

In sum, the present model suggests a highly counterintuitive but theoretically predictable scenario: The people one likes and therefore knows the best will seem less diverse than the people one dislikes and therefore barely knows.

2.3 Overview of empirical investigation

We collected data from 7 experiments that investigated the relation between liking and perceived similarity in person perception. We employed different experimental designs (within, between), using different target persons (personally known to participants, celebrities), outcome measures (spatial arrangement, pairwise comparisons, trait overlap), and participants (students in Germany, Mturk participants in the US). All experiments applied a representative stimulus sampling approach (Brunswik, 1955), and asked participants to generate their own (liked and disliked) target persons, in order to increase external validity as well as construct validity (Wells & Windshitl, 1999; Judd, Westfall, & Kenny, 2012).

We used three different measures of perceived similarity. Experiments 1 and 2 used the spatial arrangement method (SpAM; Goldstone, 1994; Hout, Goldinger, Ferguson, 2013), which builds on the intuitive and robust association between similarity and spatial distance (e.g. Casasanto, 2008; Lakoff & Johnson, 1980). SpAM asks participants to create their own mental map of a given set of stimuli by moving the stimuli around the computer screen so that distances between stimuli resemble their similarities (Goldstone, 1994; Hout et al., 2013; Koch et al., 2015; Kriegeskorte & Mur, 2012).

We used pairwise similarity ratings as a second measure of perceived similarities in Experiments 3 to 7. This procedure presents participants with all possible pairwise combinations of stimuli and asks them to rate each pair's similarity. The similarity ratings serve either directly as a measure or the resulting similarity matrix can be analyzed using multidimensional scaling (MDS; Torgerson, 1965; Krumhansl, 1978). The MDS procedure estimates the coordinates of the stimuli in an *n*-dimensional space so that the Euclidean distances between stimuli are proportional to their similarities.

We employed a third measure of similarity in Experiments 4 and 7. Participants were asked to generate traits that applied to different target persons and we calculated the proportion of traits that applied to multiple target persons simultaneously (shared traits) as a proxy for perceived similarity.

In total, we conducted seven experiments to investigate the relation between liking and perceived similarity in person perception, and to test whether people perceive liked or disliked others as more similar to one another. All experiments we conducted are reported here, along with all experimental conditions and all collected variables.

2.4 Experiment 1

We started our empirical investigation with a simple design that asked participants to sample 4 personally-known others they liked and 4 personally-known others they disliked. First, we tested whether the hedonic principle of information sampling applied, namely that participants indeed collect more information about people they like than about people they dislike. Specifically, we expected participants to have spent more time with the liked target persons than with the disliked target persons, and that, accordingly, they have more knowledge about the liked compared to the disliked target persons. Second, we tested whether participants perceive liked or disliked target persons as more similar to one another.

2.4.1 Method.

Participants and design. We had no specific prediction regarding the expected effect size in the first experiment; to achieve sufficient statistical power to detect small to medium effects (Cohen, 1988), we aimed for data from 70 participants. We factually collected data from 71 students of the University of Cologne (45 female, 26 male), who participated for a compensation of 3 Euros or for course credit. There was only one within-participants factor, as each participant provided ratings for the liked and disliked people they knew. Similarity assessment was realized using an adaptation of the spatial arrangement method (SpAM; e.g., Hout et al., 2013).

Procedure. Participants arrived in the laboratory and the experimenter seated them in front of a computer. If they agreed to participate after reading an informed consent, the experimenter started a Visual Basic program that presented instructions along with an adaption of the spatial arrangement method. First, participants were shown 8 text boxes and asked to provide forenames of four persons they knew and liked and forenames of four persons they knew but did not like. The order in which the program asked to provide names of liked and disliked persons was counterbalanced. The next screen showed participants 8 text labels displaying the provided names, arranged next to one another in the center of the screen. Instructions asked participants to create a similarity map by dragging and dropping the persons around the screen so that the distances between them indicated the similarities of their personalities (distant = dissimilar, close = similar).

Only after each name label was moved at least once, an "OK" button was activated, which enabled participants to submit their similarity maps. Figure 6 shows an example of the spatial arrangement method. On the next screen, participants used sliders, ranging from 0 to 100, to indicate their liking for each of the 8 persons ("not at all" to "extremely"). This item served as a manipulation check. Participants also indicated the amount of knowledge they had about each person ("no knowledge at all" to "very much knowledge"), and the amount of time they had spent with each person ("no time at all" to "very much time") using the same type of sliders. At the end of the experiment, participants were thanked, paid, and debriefed about the purpose of the experiment. Experimental sessions lasted about 5 minutes.

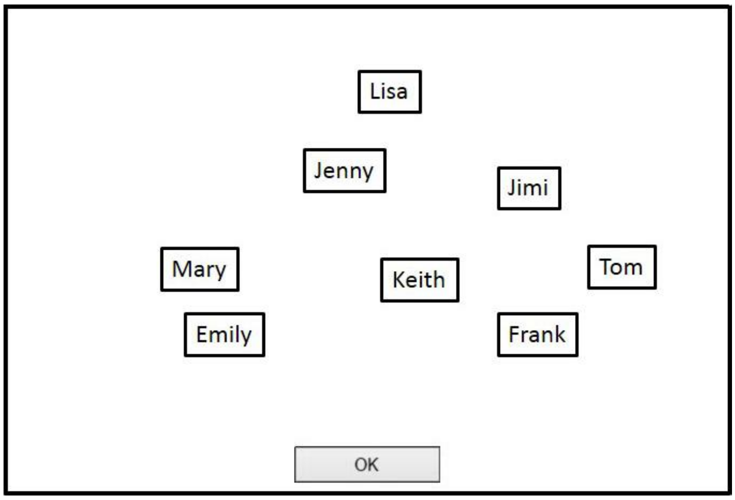

Figure 6. Illustration of the spatial arrangement method (SpAM). Participants arranged liked and disliked others according to the similarities of their personalities.

2.4.2 Results.

Liking, knowledge, and time spent together. We computed mean liking ratings for the liked and the disliked persons for each participant as a manipulation check. A paired-samples t-test confirmed that participants adhered to the instructions, as liking substantially differed between liked and disliked persons ($M_{liked} = 93.54$, $SD_{liked} = 6.54$ vs. $M_{disliked} = 18.97$, $SD_{disliked} = 9.96$), $t(70) = 46.70$, $p < .001$. We then computed mean values for knowledge and time spent together for the liked and disliked persons within each participant. As shown in Table 1, participants reported to have more knowledge about liked compared to disliked persons, $t(70) = 22.89$, $p < .001$, $d = 3.43$; and to have spent more time with liked compared to disliked persons in the past, $t(70) = 19.47$, $p < .001$, $d = 2.98$, confirming the expected pattern.

Similarity. The critical test regards the similarity (i.e., distance) between liked and disliked persons. The spatial arrangement method provides distances in percentage of the screen size. We calculated the mean distances between the four liked persons (i.e., six distances) and the mean distances between the four disliked persons for each participant. As shown in Table 1, the mean distance between the liked persons was smaller than the distance between the disliked persons, $t(70) = -5.33$, $p < .001$, $d = 0.75$, as predicted by the density hypothesis and the present model (cf. Figure 5).

	Liked	**Disliked**
Experiment 1a: Spatial Arrangement Within		
Knowledge	90.93 (8.08)	47.55 (15.95)
Time Spent Together	88.80 (10.16)	40.07 (20.81)
Distance (screen size %)	11.08 (7.17)	18.81 (12.64)
Experiment 2: Spatial Arrangement Between		
Knowledge	85.94 (7.51)	51.73 (18.56)
Time Spent Together	84.36 (7.47)	49.27 (17.81)
Distance	13.83 (8.67)	20.61 (12.34)
Experiment 3: Pairwise Similarity Ratings		
Knowledge	90.54 (8.27)	42.98 (17.97)
Time Spent Together	89.69 (8.43)	39.84 (17.80)
(Mean Euclidean) Distance	1.91 (1.06)	2.64 (1.06)
Experiment 4: Trait Generation		
Knowledge	90.01 (10.07)	42.41 (17.41)
Time Spent Together	88.67 (9.72)	37.88 (17.70)
(Mean Euclidean) Distance	1.44 (0.86)	2.40 (0.93)
Experiment 5: Celebrities		
Knowledge	56.25 (21.41)	40.63 (16.55)
(Mean Euclidean) Distance	2.25 (1.02)	2.90 (1.05)
Experiment 6: Celebrities Sampled		
(Mean Euclidean) Distance	2.31 (1.14)	2.64 (1.22)

Table 1. Mean and standard deviations of the dependent variables for liked and disliked target persons across Experiments 1 to 6.

2.4.3 Discussion.

Experiment 1 provides first support for the predicted counterintuitive scenario in person perception. First, the hedonic principle of information sampling clearly applies to participants' mental representations of other people. Participants indicated to have spent more time with liked than with disliked others and to have more knowledge about liked others. Second, despite this knowledge asymmetry, participants arranged liked others closer together indicating that they perceive them as more similar to one another regarding their personalities.

However, Experiment 1 allows a number of alternative explanations. First, the observed similarity asymmetry might hinge on the simultaneous arrangement method which requires participants to locate liked and disliked others within the same similarity space. A second concern regarding the spatial arrangement method is that participant's similarity dragging solutions might be influenced by other factors than perceived similarity. For example, they might be motivated to group liked others closely together to indicate that they stand together as friends. Likewise, participants might express that their liked others are "close with one another", meaning that they hold relationships with one another. A third concern relates to possible differences in the processing of positive and negative information (e.g. Taylor, 1991). It is a well-known phenomenon that negative stimuli trigger deeper and more elaborate processing which could increase differentiation and lead to the perception of dissimilarity. Fourth, we argue that despite the fact that participants have more knowledge about liked others, this knowledge indicates greater similarity. However, we do not know whether participants use this knowledge when judging between-person similarities. It is possible that they actually retrieve more knowledge about disliked others from their memory as would also be implied by a processing depth explanation. A final concern is that personally known liked others are likely to include persons that are related to one another, (i.e., family members), or who come from the same group, which increases their likelihood of actually possessing more similar personalities.

The following experiments will address these concerns. Experiment 2 first tests the possibility that our results hinge on the simultaneous arrangement of liked and disliked targets in the same similarity space. Simultaneous arrangement implies that similarities of liked and disliked targets should be judged on the basis of the same trait dimensions using equal resolutions of these dimensions. One might argue that people usually compare liked others on different and more specific trait dimensions than disliked others and perceive subtle differences on these dimensions as magnified.

Thus, liked targets might appear more similar to one another *only* in the context of disliked targets while the reverse might be true if both liked and disliked others are compared separately. To test this possibility, Experiment 2 asked for liked and disliked target persons between conditions. This allows participants in each condition to base their similarity judgments on different trait dimensions and to use different resolutions of these dimensions.

2.5 Experiment 2

Experiment 2 repeated Experiment 1 with a between-participants design. Instead of comparing both liked and disliked others, participants compared either liked or disliked others.

2.5.1 Method.

Participants and design. Based on the data from Experiment 1, we expected a medium to large effect size for the similarity asymmetry. Therefore, we again aimed for data from 70 participants which provided us with sufficient statistical power for a between-participants comparison (Cohen, 1988). We factually collected data from 71 participants (47 females, 24 males) who were students from the University of Cologne and participated for a compensation of 3 Euros or for course credit. Procedure and materials were similar to Experiment 1, except that participants were randomly assigned to a “like” or “dislike” condition. Participants in the “like” condition were asked to provide forenames of four persons they knew and liked while participants in the “dislike” condition were asked to provide forenames of four persons they knew but did not like. Participants than performed the spatial arrangement task before they indicated liking, amount of knowledge, and amount of time spent together for each of the four persons.

2.5.2 Results.

Liking, knowledge, and time spent together. As a manipulation check, we compared the liking ratings in the “like” condition with the ratings in the “dislike” condition. This difference was highly significant (M_{liked} = 91.50, SD_{liked} = 6.99 vs. $M_{disliked}$ = 26.51, $SD_{disliked}$ = 14.13), $t(69) = 24.46$, $p < .001$. As shown in Table 1, participants in the like condition also indicated to have more knowledge about the target persons than participants in the dislike condition, $t(69) = 10.13$, $p < .001$, $d = 2.44$. Participants in the like condition also reported to have spent more time with the persons than participants in the disliked condition, $t(69) = 10.77$, $p < .001$, $d = 2.59$, replicating the finding that liking goes along with more knowledge and exposure in a between-participants comparison.

Similarity. We calculated the mean distances between the four persons in the like condition and the mean distances between the four persons in the dislike condition. Table 1 shows that distances between target persons in the like condition were smaller than in the disliked condition, $t(69) = 2.67$, $p = .009$, $d = .64$. Liked others were again perceived as being more similar regarding their personalities than disliked others.

2.5.3 Discussion.

Experiment 2 showed that the effects found in Experiment 1 did not hinge on a simultaneous comparison of liked and disliked others. When participants compared liked or disliked others in two conditions, they still perceived liked others as more similar to one another. Participants were not restricted to compare liked and disliked others within the same similarity space. Instead, participants in the liked condition had the possibility to base their similarity judgments on completely different trait dimensions and to use finer resolutions than participants in the disliked condition. In addition, participants still indicated to have more knowledge about liked others, and to have spent more time together with liked others.

Having ruled out a possible comparison set artifact, we now address other limitations related to the spatial arrangement method. Even though SpAM offers an elegant tool for simultaneous similarity assessment, this method leaves room for some alternative explanations regarding the similarity asymmetry we found.

First, participants might be *motivated* to spatially arrange liked others close together because they want them to “stand together” as friends. Second, the dense spatial arrangement of liked others might reflect participants’ experience that their friends often appear together or even live close together. Further, participants might simply indicate that their liked others are “close with one another”, or form a “circle of friends”. To address these concerns, we conducted another experiment using pairwise similarity ratings instead of SpAM.

Yet another concern is that results might reflect affect-induced differences in the processing of positive and negative stimuli (e.g. Taylor, 1991; Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). Accordingly, negative affect triggers deeper and more elaborate processing than positive affect. Assuming that the comparison of liked and disliked others triggers affective reactions in participants, it is possible that participants engage in deeper processing when comparing disliked others. As a result, they might even retrieve more knowledge about disliked others from memory and might conclude that they are dissimilar, while the more shallow processing when comparing liked others might give rise to a heuristic judgment that they are similar. To test such a processing explanation, we measured response latencies for each pairwise similarity ratings as a proxy for processing (e.g., Craik & Tulving, 1975). Assuming that response latencies are a measure for processing depth, this allows to test whether processing asymmetries can account for the differential similarity effect.

2.6 Experiment 3

Experiment 3 asked participants to compare the personalities of four liked and four disliked others using pairwise similarity ratings while we recorded response latencies for each comparison. We expected to replicate the similarity asymmetry from the previous two experiments and expected this asymmetry to be unrelated to possible processing depth differences as indexed by response latencies.

2.6.1 Method.

Participants and design. Assuming a similar effect size, we again collected data from 70 participants (54 females, 16 males). All participants were students from the University of Cologne and participated for a bag of gummy bears or for course credit. There was only one within-participants factor; each participant rated similarities of liked as well as disliked persons.

Procedure. Similar to Experiment 1, participants first provided forenames of four liked and four disliked persons they personally knew. The next screen informed participants about the study's purpose to measure how similar these persons were regarding their personalities. Participants were further instructed that they would be asked to compare the persons with one another sequentially. Thus, each participant performed a total of 28 comparisons. Comparison orders were randomized for each participant. Each comparison trial showed a text label on the screen stating "*How similar are [Person X] and [Person Y] to each other regarding their personalities?*" Participants then provided their rating on a 9 point scale with endpoints labeled "not at all similar" (0) and "extremely similar" (9). For each comparison, the computer recorded the time it took participants to respond. After completing the 28 comparisons, participants indicated their liking for each of the 8 persons, the amount of knowledge they had about that person, and the amount of time they had spent with that person. Then experimenters thanked, paid, and debriefed participants about the purpose of the experiment. Experimental sessions lasted about 5 minutes.

2.6.2 Results.

Liking, knowledge, and time spent together. Participants adhered to the instruction, as liking substantially differed between liked and disliked persons ($M_{liked} = 93.98$, $SD_{liked} = 6.33$ vs. $M_{disliked} = 19.50$, $SD_{disliked} = 12.05$), $t(69) = 42.62$, $p < .001$. Table 1 shows that participants again indicated to have more knowledge about liked compared to disliked persons, $t(69) = 19.59$, $p < .001$, $d = 3.40$, and to have spent more time with liked compared to disliked persons, $t(69) = 20.28$, $p < .001$, $d = 3.58$).

Similarity. To compare similarities among liked and disliked persons, we used the MDS procedure (Krumhansl, 1978) provided by the SAS system. We assumed an ordinal structure of the similarity ratings and submitted participants' ratings to separate MDS procedures. Analogous to the spatial arrangement method, we defined a two-dimensional similarity space and the MDS procedure estimated coordinates for each person in this space by minimizing a loss function. Based on each participant's MDS solution, we then calculated the mean Euclidean distances from each person to all other persons of the same valence (i.e., a "density" index for liked vs. disliked others; see Unkelbach et al., 2008). A high index indicates low density; the person is dissimilar to others. A low index indicates high density; the person is similar to others. For each participant, we then computed the mean Euclidean distance among the four liked persons and the mean distance among the four disliked persons. As shown in Table 1, the mean distance between the liked persons was substantially smaller than the distance between the disliked persons, $t(69) = -4.33$, $p < .001$, $d = .69$.

Next, we wanted to assure that the observed density difference did not hinge on the two-dimensionality of the MDS procedure. Yet, the differential density was also present when we estimated distances based on three-dimensional ($t(69) = -4.26$, $p < .001$), and four-dimensional scaling solutions ($t(69) = -4.27$, $p < .001$).

Response latencies. Each participant performed six comparisons of two liked persons and another six comparisons of two disliked persons. A paired-samples *t*-test found that participants performed the liked-liked comparisons faster than the disliked-disliked comparisons ($M_{liked} = 5.25$ s, $SD_{liked} = 1.96$ s vs. $M_{disliked} = 6.03$ s, $SD_{disliked} = 2.33$ s), $t(69) = -3.76$, $p < .001$, $d = .36$. This pattern is in line with affect-based processing theories which suggest that people engage in deeper processing when confronted with negative stimuli. Thus, we investigated if these latency differences accounted for the observed similarity differences.

Regression coefficient analysis. We conducted the reaction time analysis on the level of the individual pairwise similarity ratings. We analyzed whether response latency of a given comparison was related to its respective (similarity) rating and if so, whether this could account for the larger perceived similarity among liked vs. disliked targets. In other words, if processing asymmetries explain the observed pattern, longer latencies should indicate lower similarity. Conversely, if our model holds, similarity should simply be predicted by the valence of the target persons.

We tested whether response latency and valence predicted similarity ratings using regression coefficient analysis (RCS; Lorch & Myers, 1990; Thompson, 2008; see also Baayen, Davidson, & Bates, 2008, for a discussion). This procedure also enabled us to test whether the density asymmetry was evident at the level of individual similarity ratings that did not undergo a multidimensional scaling procedure. For each participant, there were twelve similarity ratings of interest (i.e., six like-like and six dislike-dislike comparisons), along with their respective response latencies. In addition, we included a dichotomous valence variable that contrasted these six liked-liked and the six disliked-disliked comparisons. For each participant we then specified a simultaneous regression model predicting the 12 similarity ratings by their respective response latencies and target valences. Regression could not be performed for one participant whose similarity ratings were always 1. We excluded this participant and then calculated mean standardized regression coefficients for reaction time and for valence across participants. Contradicting a processing depth explanation, response latencies did not predict similarity ratings (*Mean Beta* = -.04, $SD = 0.31$), $t(68) = -0.95$, $p = .34$. Thus, while it took participants longer to compare the personalities of disliked persons, this had no influence on their similarity ratings. In accordance with the density hypothesis, only target valence significantly predicted similarity ratings (*Mean Beta* = .20, $SD = 0.49$), $t(68) = 3.46$, $p = .001$, $d = .42$.

2.6.3 Discussion.

Experiment 3 replicated the previous findings using pairwise similarity ratings instead of the spatial arrangement method. The data suggest that the differential similarity of liked and disliked persons is no artifact of spatial arrangement.

The experiment also provided some deeper insights. Participants took longer to make disliked-disliked comparisons than liked-liked comparisons, indicating a possible processing difference. While the observed latency difference mirrors the phenomenon that positive information is processed faster than negative information (e.g. Unkelbach et al., 2008; 2010), regression coefficient analysis showed that response latencies did not predict similarity ratings, ruling out a processing depth explanation. In line with the model presented in Figure 5, the only significant predictor for participant's raw similarity ratings was whether the target persons were liked or disliked.

The experiments so far showed that participants report more knowledge about liked others, but they seem to perceive them as more similar than disliked others regarding their personalities. However, it is unclear if participants use their knowledge when assessing the similarities, either pairwise or by spatial arrangement. The next experiment therefore aimed to activate participants' knowledge structures prior to similarity judgments by including a task that asked participants to generate for each of the persons as many traits as they could come up with.

This procedure also provides us with a more direct measure of participants' knowledge structures and it allows us to test our predictions in several new ways. First, we could test whether participants are able to generate more traits for liked than for disliked others as predicted by the hedonic principle of information sampling. Second, we could test whether these larger information samples are actually more similar; that is, we would expect participants to describe their liked others with relatively more of the same traits and their disliked others with different traits.

Third, analyzing the trait structure allows for another critical test of our model that can discount two other alternative explanation for the results we obtained thus far. That is, our model predicts that positive impressions are more similar than negative impression even across different perceivers. This means that we expect a participant's liked targets to show a stronger trait overlap with other participants' liked targets as well. Such a larger between participants trait overlap would discount the possibility that our effects are due to systematic differences between liked and disliked targets that are specific for a given participants. For example, participants might know more liked than disliked targets, which enables them to sample liked targets from the same context (e.g. family, group of friends) but forces them to sample disliked targets from different contexts. Further, as known from research on interpersonal attraction, people like others who are similar to themselves (Byrne, 1971; Berscheid, 1985; Montoya, Horton & Kirchner, 2008). From this perspective one could argue that liked others might be perceived as more similar to one another because they are more similar to the perceiver.

These two alternative explanations predict a larger trait overlap among liked targets only within participants, but not across different participants. The present density model on the other hand predicts a larger trait overlap among liked targets both within participants as well as across different participants.

2.7 Experiment 4

Experiment 4 was similar to Experiment 3 except that participants were asked to generate for each target person as many traits as they could come up with prior to providing similarity judgments. We thereby tested whether knowledge activation influenced perceived similarity and whether the target persons’ trait structure was in line with the present density model.

2.7.1 Method.

Participants and design. Based on the previous effects, we collected data from 70 participants (48 females, 22 males). All 70 participants were students from the University of Cologne and participated for a bag of gummy bears or for course credit. There was only one within-participants factor, as each participant generated traits for and rated similarities of liked as well as disliked persons they knew.

Procedure. The procedure was almost identical to the previous experiments. Yet, participants were asked to provide for each person as many traits as they could come up with. Participants used textboxes to provide the traits while the order of the eight target persons was randomized. As in Experiment 3, participants then provided pairwise similarity ratings as well as liking, knowledge, and time spent together ratings. Finally, participants were thanked, paid, and debriefed about the purpose of the experiment. Experimental sessions lasted about 10 minutes.

2.7.2 Results.

Liking, knowledge, and time spent together. Again, liking ratings differed substantially between liked and disliked target persons ($M_{liked} = 95.26$, $SD_{liked} = 5.07$ vs. $M_{disliked} = 18.50$, $SD_{disliked} = 13.85$), $t(69) = 41.00$, $p < .001$. Table 1 shows that participants indicated to have more knowledge about liked compared to disliked persons, $t(69) = 17.83$, $p < .001$, $d = 3.35$, and to have spent more time with liked than with disliked persons, $t(69) = 19.81$, $p < .001$, $d = 3.56$.

Similarity. We submitted each participant’s pairwise similarity ratings to the same density analyses as in Experiment 3. As shown in Table 1, the mean Euclidean distance between liked persons was again smaller than the mean distance between disliked persons, $t(69) = -6.80$, $p < .001$, $d = 1.07$. Liked persons clustered more densely then disliked persons. The differential density was also present when we submitted similarity ratings to three-dimensional ($t(69) = -6.15$, $p < .001$, $d = 1.12$), and four-dimensional solutions ($t(69) = -6.53$, $p < .001$, $d = 1.12$).

Response latencies. We again computed the mean response latencies for the liked-liked and disliked-disliked comparisons. A paired-samples *t*-test found that participants again performed liked-liked comparisons faster than disliked-disliked comparisons ($M_{liked} = 4.74$ s, $SD_{liked} = 1.64$ s vs. $M_{disliked} = 5.53$ s, $SD_{disliked} = 1.85$ s), $t(69) = -3.49$, $p = .001$; $d = .45$.

Regression coefficient analysis. We again performed RCA to test if response latencies of the individual comparisons predicted the respective similarity ratings. As in Experiment 3, response latencies did not predict similarity ratings (*Mean Beta* = -.02, *SD* = 0.35), $t(69) = -0.37$, $p = .71$. However, as in the previous analyses, valence of the comparison targets did predict similarity ratings (*Mean Beta* = .27, *SD* = 0.48), $t(69) = 5.06$, $p < .001$, $d = .63$.

Trait overlap within participants. Experiment 4's trait generation task affords a new test of the knowledge asymmetry and of the density asymmetry. We excluded one participant from the following trait analysis as this participant failed to generate any traits for three of the four disliked persons. However, including this participant did not alter any of the following results.

In line with the proposed knowledge asymmetry, participants generated substantially more traits for liked persons ($M_{liked} = 6.67$, $SD_{liked} = 3.13$) than for disliked persons ($M_{disliked} = 3.91$, $SD_{disliked} = 1.76$), $t(68) = 11.59$, $p < .001$, $d = 1.09$.

To test for the proposed higher similarity of these larger information samples, we then computed the proportion of traits that participants assigned multiple times to different liked and different disliked others (i.e., how often participants used the same trait within a valence group). We calculated for each participant the mean proportion of traits that are shared among their liked targets and the proportion of traits shared among their disliked targets. Indeed, the proportion of shared traits was larger for the liked pairs ($M = .33$, $SD = .17$) than for the disliked pairs ($M = .22$, $SD = .17$), $t(68) = 5.06$, $p < .001$, $d = .65$. These numbers mean that 33% of the traits generated for a given liked target are on average shared with another liked target, while only 22% of a disliked target's traits are shared with another disliked target. Thus, participants generated *more* traits for the liked targets than for the disliked targets while they also generated a larger proportion of *the same* traits for their liked targets, supporting the prediction derived from our model.

There is one possible caveat arising from the fact that participants generated more traits for the liked than for the disliked others. We propose that this mirrors the knowledge asymmetry that participants also report on the explicit knowledge measure. However, this asymmetry might indicated that negative traits have a stronger impact on impressions than positive traits (see Baumeister et al., 2001 for a review). In order to be liked, many positive traits might have to be present while in order to be disliked, a few or even only one negative traits might be sufficient. This would in turn restrict the variation of liked individuals' traits as many positive traits have to simultaneously be present in them. There would simply be less degrees of freedom for variation among liked compared to disliked targets. If this alternative explanation holds, liked targets' larger trait overlap (proportion of shared traits) should be accounted for by the larger total number of traits that they are represented with.

We ran another regression coefficient analysis in order to test for this alternative explanation. For each of the 6 liked-liked and each of the 6 disliked-disliked target pairs, within each participant, we calculated the proportion of shared traits, serving as regression criterion, and we calculated the total number traits that each pair was described with, serving as predictor. Thus, for each of these twelve target pairs, we now had the proportion of shared traits, the number of traits and their valence (liked vs. disliked). We defined a regression model within each participant, predicting trait overlap (proportion of shared traits) by number of traits and valence. Ruling out the alternative explanation and in line with the density hypothesis, valence did predict trait overlap (*Mean Beta* = .32, $t(68) = 3.39$, $p = .001$, $d = .82$), but number of traits did not (*Mean Beta* = -.03, $t(68) = -.35$, $p = .72$). Thus, liked others show a stronger trait overlap independent of the number of generated traits.

Trait overlap across participants. As discussed above, a specific prediction of the density hypothesis and the present model is that trait overlap should also be larger for liked compared to disliked targets across different participants. We calculated the across-participants trait overlap separately for the eight target persons. For example, the traits of the first liked target that a given participant generated was compared to all first liked targets that the other participants generated, and so on. The resulting four liked and four disliked proportions were then averaged.[2] In line with the present density model, liked targets showed a larger trait overlap across participants compared to disliked targets ($M_{liked} = 0.07$, $SD_{liked} = 0.03$ vs. $M_{disliked} = 0.03$, $SD_{disliked} = 0.02$), $t(68) = 10.68$, $p < .001$; $d = 1.57$. These values mean that a liked target on average shared 7% of its traits with another participant's liked target, while a disliked target shares only 3% of its traits with another participant's disliked target. These differential proportions across participants only follow from the present model, but neither from the alternative that participants might sample liked others from the same and disliked others from different contexts (as this is not possible across participants), nor from the possibility that participants sample liked others that are similar to themselves and therefore similar to one other (as similarity to the self cannot play a role in trait overlap across participants).

2.7.3 Discussion.

Experiment 4 replicated the findings from the previous experiments; participants rated liked persons as more similar than disliked persons. This effect occurred after participants initially engaged in a trait generation task which required them to thoroughly think about all persons. The trait generation task ensured that participants' knowledge structures were activated. This, however, did not weaken the similarity asymmetry; rather, the effect increased numerically ($d_{Exp.3}$ = 0.69 vs. $d_{\text{Exp.4}}$ = 1.07), z = 1.53, p = .06. Thus, even when participants carefully and explicitly considered what they knew about liked and disliked others, liked others appeared more similar to one another than disliked others. Participants again performed the liked-liked comparisons faster than the disliked-disliked comparisons; but again, response latencies did not predict similarity ratings while target valence did.

The trait generation task provided new insights regarding participants' knowledge structures. First, participants were able to generate almost twice as many traits for the liked targets compared to the disliked targets. Second, participants perceived liked targets' personalities as more similar to one another, as they assigned a larger proportion of the same traits to them. Thus, the asymmetry in participants' similarity judgments was in accordance with the traits they generated. This shows that even though participants have more knowledge about the personalities of people they like (i.e., larger information samples), this knowledge contains highly similar attributes (i.e., similar content); while knowledge about disliked people contains less but rather diverse attributes.

In addition, liked targets showed a larger trait overlap than disliked targets even across participants. That is, the liked targets of different participants were described with more of the same traits than the disliked targets. This is strong evidence for our model's prediction of a generally small positive diversity. In addition, this finding rules out the possibility that liked others have a larger trait overlap because they stem from the same context or because they are more similar to the perceiver. If this was true, liked others should be more similar than disliked others only within participants, but not across participants.

To summarize so far, Experiments 2-4 ruled out alternative explanations of the basic effect in Experiment 1 and supported the model put forth in Figure 5. We now turn to a concern regarding the type of target persons used in the previous experiments. So far we showed the association between liking and similarity for the mental representation of persons that participants personally knew. The question remains to what extent the results are specific to these target persons. For example, participants might have selected multiple family members as personally known liked others. Liked others would then have more similar personalities due to the fact that they are relatives. In addition, liked others are more likely to know one another and to hold relationships with one another, while the same might not be true for personally known disliked others. Consequently, liked others might display stronger entitativity as a group than disliked others, which may cause participants to judge them as more similar (e.g., Lickel, Hamilton, Wieczorkowska, Lewis, Sherman, & Uhles, 2000). Although the trait overlap data from Experiment 4 is not in line with this alternative, it seems prudent to rule out this alternative experimentally.

In addition, disliked targets displayed a much smaller variability in terms of their likeability than disliked targets even within participants (Experiment 1: $Mean(SD_{liked})$ = 6.16 vs. $Mean(SD_{disliked})$ = 13.62; Experiment 2: $Mean(SD_{liked})$ = 7.67 vs. $Mean(SD_{disliked})$ = 16.14; Experiment 3: $Mean(SD_{liked})$ = 6.01 vs. $Mean(SD_{disliked})$ = 14.21; Experiment 4: $Mean(SD_{liked})$ = 5.17 vs. $Mean(SD_{disliked})$ = 13.34). This indicates that liked others are perceived as equally likeable while disliked others vary in terms of their likeability. Liked others might then be judged as more similar because they are similar in the sense that they are all liked to the same extent.

If any of these alternative explanations apply, results from the previous experiments would not be evidence for our model of limited positive diversity in person perception, but merely be an artifact of the target persons generated by participants. To rule out these alternative explanations and to put our model to another test, Experiment 5 used celebrities instead of personally known others as target persons.

2.8 Experiment 5

Experiment 5 tests the generality of our model by moving from the representation of personally known people to the representation of publicly known people. If liking goes along with decreased diversity, participants should perceive liked celebrities as more similar to one another than disliked celebrities.

2.8.1 Method.

Participants and design. Based on the previous effects, we collected data from 71 participants (57 females, 14 males). All participants were students of the University of Cologne and participated for a bag of gummy bears or for course credit. There was only one within-participants factor, as each participant rated similarities of liked as well as disliked celebrities.

Procedure. The procedure followed Experiment 3 except that we asked participants to provide names of four liked and four disliked celebrities. Participants then provided pairwise similarity ratings while we recorded response latencies. Participants also provided liking and knowledge ratings for the celebrities, while time spent together ratings were omitted for obvious reasons. At the end, participants were thanked, paid, and debriefed about the purpose of the experiment. Experimental sessions lasted about five minutes.

2.8.2 Results.

Liking and knowledge. Participants followed instructions, as liking substantially differed between liked celebrities and disliked celebrities (M_{liked} = 80.40, SD_{liked} = 11.79 vs. $M_{disliked}$ = 14.94, $SD_{disliked}$ = 12.50), $t(70) = 30.00$, $p < .001$. Table 1 shows that participants reported to have more knowledge about liked celebrities compared to disliked celebrities, $t(70) = 8.47$, $p < .001$, $d = .82$.

Similarity. Table 1 shows the results of the same density analysis as in Experiments 3 and 4. The effect did not hinge on the dimensionality of the MDS solution as it was also present for three-dimensional ($t(70) = -3.19$, $p = .002$, $d = 0.55$), and four-dimensional solutions ($t(69) = -3.18$, $p = .002$, $d = .54$).

Response latencies. We again computed mean response latencies for the comparisons involving liked celebrity pairs and the comparisons involving disliked celebrity pairs for each participant. Different from the comparisons involving personally known others, participants performed liked-liked and disliked-disliked comparisons equally fast ($M_{liked} = 5.15$ s, $SD_{liked} = 2.08$ s vs. $M_{disliked} = 5.06$ s, $SD_{disliked} = 2.06$ s), $t(70) = 0.44$, $p = .664$.

2.8.3 Discussion.

Experiment 5 asked participants to judge the similarities of celebrities. As Experiment 5 was identical to Experiment 3 in all other respects including sample size, comparing results from both experiments is of particular interest.

Participants indicated that they had more knowledge about liked celebrities than about disliked celebrities, but this effect was substantially smaller than for personally known others in Experiment 3 ($d_{\text{Exp.5}} = 0.82$ vs. $d_{Exp.3} = 3.40$), $z = 8.16$, $p < .001$. This suggests that the association between liking and knowledge is weaker when target information is acquired indirectly (e.g., via television) compared to when target information is acquired directly (via social interactions).

Importantly, participants also perceived liked celebrities as more similar to one another than disliked celebrities. This effect was of similar size as for personally known others in Experiment 3 ($d_{\text{Exp.5}} = 0.63$ vs. $d_{Exp.3} = 0.69$). As predicted, the density asymmetry is not limited to the representation of personally known people but applies to publicly known people as well.

This finding also excludes the possible alternative explanations that personally known others are perceived as more similar because they are related or because they display a larger entitativity due to personal relationships or group memberships they hold. Further, and in contrast to personally known target persons, liked and disliked celebrities were equal regarding their liking variability within participants ($Mean(SD_{liked}) = 11.13$ vs. $Mean(SD_{disliked}) = 11.52$), while the similarity difference was of similar size as in the previous experiments. This rules out the possibility that liked persons are judged as more similar because they are more similar in likeability. Different from the previous experiments, participants performed the liked-liked and disliked-disliked comparisons equally fast for celebrities. Thus, the density asymmetry occurred independently of the reaction time asymmetry, further ruling out processing depth explanations.

While results from the present experiment using celebrities make it unlikely that liked targets hold relations with one another possibly causing the similarity asymmetry, there remains a final concern. Possibly, participants know more liked celebrities than disliked celebrities, which enables them to sample liked target from the same context (e.g. TV show) but forces them to sample disliked targets from different contexts. Experiment 6 therefore asked participants to draw a larger and more exhaustive sample of liked and disliked celebrities to increase context diversity of target persons.

2.9 Experiment 6

Experiment 6 tests whether previous results were due to participants retrieving liked and disliked targets from unequally diverse contexts. To account for this possible asymmetry, we instructed participants to sample 12 liked and 12 disliked celebrities before only four of each group were randomly determined as target persons. We expected the density asymmetry to hold even when target persons are drawn from a larger and thus more divers sample of celebrities.[1]

2.9.1 Method.

Participants and design. We expected the random selection of target persons from a larger sample to introduce error variance as similarity ratings are likely to be sensitive to target persons' proximity in the generated list. That is, if targets appear close to one another in the initial list, they are likely to stem from the same context and are thus more likely to be similar to one another. We therefore increased the sample size and collected data from 121 participants (63 females, 58 males). All participants were recruited online via the Mturk platform; all were located in the US and were compensated with $ 0.70.

Procedure. We asked participants to provide names of twelve liked and twelve disliked celebrities. Four of the liked and four of the disliked celebrities were randomly determined as target persons for which participants then provided pairwise similarity ratings. We omitted liking and knowledge ratings in order to keep the online experimental sessions short and increase the likelihood of successful completion. At the end, participants were asked to rate their level of concentration during the experiment (1 = "extremely low" to 6 = "extremely high"), before they were thanked, and debriefed about the purpose of the experiment. Experimental sessions lasted about 15 minutes.

2.9.2 Results.

Similarity. We removed data from one participant who rated his level of concentration as "extremely low". We then conducted a similar density analysis as in Experiments 3, 4, and 5. As Table 1 shows, the mean Euclidean distance between liked celebrities was again smaller than the distance between the disliked celebrities, $t(119) = -2.73$, $p = .007$, $d = 0.28$. The effect did not hinge on the dimensionality of the MDS solution as it was also present for three-dimensional ($t(119) = -2.54$, $p = .012$), and four-dimensional solutions ($t(119) = -2.53$, $p = .013$).

2.9.3 Discussion.

As in the previous experiments, liked target persons were again perceived to be more similar to one another than disliked target persons. Crucially, the present experiment ruled out the possibility that this effect was due to differences in retrieving a small number of liked and disliked celebrity targets. That is, even when participants generated a larger sample of 24 celebrities, random sub-samples of 4 liked and 4 disliked celebrities showed the same asymmetry. The observed effect was smaller compared to the previous experiments, which might be due to the procedural differences, that is, the online data collection process and/or the sampling of 24 liked and disliked celebrities. Some participants reported that they found it difficult to come up with 24 celebrities they feel strongly about. Despite these variations, the predictions from the present model hold and the density asymmetry seems to be a general characteristic of people's mental representation of liked and disliked people.

2.10 Experiment 7

So far, we have tested our model's predictions by comparing the perceived similarities of liked and disliked target persons, and we have addressed a number of possible alternative explanations for the similarity asymmetry. However, we cannot completely rule out that the liked and disliked target persons that participants retrieve differ on some unassessed variables that create the observed similarity asymmetry, meaning that confounds remain a possibility. Experiment 7 therefore puts our model to a final critical test that does not rely on contrasting liked and disliked targets. From our model's assumption that positive traits are less diverse than negative traits it follows that different people have rather similar positive traits, but different negative traits; positive traits should have a higher matching probability across different people. In other words, positive traits can be expected to constitute people's similarities, while negative traits constitute their differences. If this is true, focusing on people's positive or negative attributes should make them appear more or less similar to one another.

Therefore, Experiment 7 asked participants to think about two persons they personally know, without specifying whether target persons had to be liked or disliked. Participants then generated positive or negative traits for the target persons before they were asked to judge the similarity of the target persons' personalities. Based on our model of small positive diversity, we predicted that participants focusing on positive traits generate a larger proportion of the same traits for both target persons (more shared traits), and that this leads participants to perceive them as more similar compared to participants focusing on negative traits. In addition, we predicted a larger overlap of positive traits across participants as well.

2.10.1 Method.

Participants and design. We had no specific prediction regarding the expected effect size; for sufficient statistical power to detect small to medium effects in a between-participants design (Cohen, 1988), we aimed for data from 100 participants. We factually collected data from 101 participants (49 females, 52 males). All participants were recruited online via the Mturk platform; all were located in the US and were compensated with $ 0.70. We asked participants to either generate positive or negative traits for two target persons.

Materials and procedure. We realized the experiment using the Qualtrics software. Participants first read and agreed with a consent form before answering demographical questions. On the next screen, participants used textboxes to provide the forenames of two people they personally knew. Depending on condition, participants then provided as many positive (positive condition) or negative (negative condition) traits as they could come up with for each person by typing them into two separate textboxes. The next screen asked participants, “What do you think, how similar are [*Person 1*] and [*Person 2*] regarding their personalities?” Participants used a slider ranging from 0 (very dissimilar) to 100 (very similar) to indicate the target persons’ similarity. At the end, participants were asked to rate their level of concentration during the experiment (1 = “extremely low” to 6 = “extremely high”), before they were thanked, and debriefed about the purpose of the experiment. Experimental sessions lasted about six minutes.

2.10.2 Results.

Similarity. As predicted, participants who generated positive traits judged the target persons as more similar (M = 56.78, SD = 24.39) than participants who generated negative traits (M = 38.78, SD = 26.81), $t(99) = 3.53$, $p = .001$; $d = .71$.

Trait overlap within participants. We excluded one participant from the trait analysis because she described target persons using whole sentences instead of individual trait words. We counted the number of traits participants generated and calculated the mean number of traits generated in the positive and negative conditions. Participants in the positive condition generated almost twice as many traits compared to participants in the negative condition ($M_{pos} = 6.42$, $SD_{pos} = 2.84$ vs. $M_{neg} = 3.78$, $SD_{neg} = 1.66$), $t(98) = 5.69$, $p < .001$, $d = 1.15$.

We then computed the proportion of traits that participants assigned to both target persons. For each participant, we counted the number of traits assigned to both persons simultaneously (shared traits) and divided this by the mean number of assigned traits. Indeed, the proportion of shared traits was larger in the positive condition compared to the negative condition ($M_{positve} = .21$, $SD = .18$ vs. $M_{negative} = .08$, $SD = .15$), $t(98) = 5.69$, $p < .001$, $d = .79$. Thus, participants in the positive condition were more likely to assign the same traits to both target persons compared to participants in the negative condition.

Regression analysis. We then conducted a regression analysis to test our prediction that participants in the positive condition perceived target persons as more similar because the traits they generated had a higher matching probability, that is, they were more likely to be shared by both target persons. In a simultaneous regression we predicted similarity ratings by experimental condition (positive traits vs. negative traits), and by the proportion of shared traits generated. The zero order correlation between the two predictors valence and proportion of shared traits was $r = .36$, $p < .001$; the criterion (similarity) was positively correlated with valence ($r = .33$, $p = .001$) and proportion of shared traits ($r = .54$, $p < .001$). In a simultaneous regression, valence was no longer a significant predictor of perceived similarity ($\beta = .15$, $p = .108$), while proportion of shared traits remained a significant predictor ($\beta = .49$, $p < .001$). Hence, the proportion of shared traits that participants generated accounted for the effect of trait valence on perceived similarity of target persons.

Given that participants in the positive condition also generated a larger total number of traits this might cause them to generate a larger proportion of shared traits. If the trait sample size increases sampling becomes more exhaustive and traits might be more likely to overlap. This possibility was ruled out by the fact that when we entered the total number of traits into the regression model as an additional predictor, it did not predict similarity ($\beta = - .03$, $p = .737$), while proportion of shared traits remained a significant predictor of similarity ($\beta = .50$, $p < .001$).

Trait overlap across participants. Similar to Experiment 4, we then tested whether trait overlap across participants was also larger for positive than for negative traits. We calculated the proportion of shared traits among different participants' first and among different participants' second target persons in the positive traits condition and in the negative traits condition.[3] As predicted by the present density model, the mean proportion of shared traits was larger among participants in the positive traits condition compared to those in the negative traits condition ($M_{positve} = .08$, $SD = .03$ vs. $M_{negative} = .01$, $SD = .01$), $t(98) = 14.28$, $p < .001$, $d = 2.88$. This means that different participants describe their target persons in a much more similar way when they name positive compared to negative traits.

2.10.3 Discussion.

Experiment 7 investigated the flip side of our prediction that liked persons should be more similar than disliked persons; namely, that generating positive attributes makes people more similar, and generating negative attributes makes them more dissimilar. In contrast to the previous experiments, these results preclude any possible confounds related to self-sampling liked and disliked targets.

When participants generated positive traits, they described target persons with more of the same traits compared to participants in the negative condition. Again in line with our model, positive traits had a higher matching probability across participants' target persons, leading to higher perceived similarity in the positive condition as shown by the regression results. In addition, positive compared to negative attributes had a higher matching probability across different participants, providing strong support for our model shown in Figure 5.

2.11 General Discussion

The present work examines the relationship between liking and perceived similarity in person perception. We introduced two principles that make opposite predictions regarding this relationship, namely liking-breeds-differentiation and the density hypothesis. The liking-breeds-differentiation principle builds on the assumption that information sampling follows a hedonic principle (Denrell, 2005; Fazio et al, 2004). People spend the majority of their social interactions with persons they like, their partners, friends, and family members. In addition, people usually avoid the unpleasant experience of interacting with someone they do not like. As a result, people acquire more knowledge about the persons they like and therefore have highly differentiated representations of them. Because differentiation typically reduces perceived similarity (e.g. Nosofsky, 1986; Goldstone & Steyvers, 2001; Linville et al., 1989; Shepard, 1987), liked others should appear less similar to one another in people's mental representation than disliked others.

On the other hand, liking is based on positive representations, which according to the density hypothesis, are less diverse compared to negative representations (Unkelbach, 2012, Unkelbach, et al, 2008). That is, each positive quality is usually defined by the absence of several different negative qualities. While the ways in which an object or a person can meet certain "positivity" criteria are limited, the ways to diverge from them are numerous. We summarized this in the model displayed in Figure 5. Regarding the mental representation of other people, knowledge about liked others should contain highly similar (i.e. matching) attributes and therefore indicate greater similarity, while knowledge about disliked others should contain rather diverse attributes.

Results from seven experiments support the density hypothesis and our model. Even though participants had more knowledge about liked others they perceived liked others as more similar to one another than disliked others.

The hedonic principle of information sampling was also evident as participants consistently indicated to have more knowledge about liked others than about disliked others. In addition, Experiment 4 found that participants were able to describe liked others with almost twice as many character traits than disliked others. It seems to be the case that people are true experts regarding the people they like.

However, this knowledge asymmetry did not lead to the perception that liked others are less similar to one another than disliked others. In six out of the seven experiments we asked participants to compare the personalities of four liked and four disliked others, and liked others were clearly perceived as more similar to one another. Experiment 1 showed this asymmetry when participants simultaneously positioned liked and disliked others in a spatial arrangement procedure. Experiment 2 found the same asymmetry in a between-participants design which allowed participants in the liked and disliked conditions to base their similarity judgments on different dimensions with different resolutions. Experiment 3 replicated the effect using pairwise similarity ratings, thereby ruling out a number of alternative explanations related to the spatial arrangement method. Experiment 4 replicated the effect after participants engaged in a trait generation task, ensuring that their knowledge about the target persons was activated. This knowledge activation procedure did not weaken but strengthen the similarity asymmetry. Moreover, the greater perceived similarity among liked others was also visible among the traits that participants generated. Liked targets showed a larger trait overlap than disliked targets, both within and between participants. Experiment 5 found that the density asymmetry applies to personally as well as publicly known people ruling out additional alternative explanations regarding the relatedness of personally known others. Experiment 6 replicated the effect when participants generated a larger sample of liked and disliked celebrities from which targets were randomly chosen in order to increase variability among targets. Finally, Experiment 7 experimentally manipulated valence of participants' focus and found that focusing on target persons' positive traits compared to focusing on their negative traits increased perceived similarity.

In Experiments 3 to 5, we measured response latencies for the similarity judgments to test whether results were due to an affect-induced processing asymmetry (Taylor, 1991). It took participants indeed longer to compare personally known disliked others than liked others which might indicate deeper processing. Crucially, the response latencies did not influence the similarity ratings which were however predicted by the valence of the target persons. The reaction time difference was not present in Experiment 5 where participants compared celebrities, while the density asymmetry was still present.

To summarize, people adhered to a hedonic principle of information sampling and reported and showed more knowledge about liked others. Yet, the content of these larger samples consistently indicated higher similarity between liked compared to disliked others.

2.11.1 Implications.

We believe the observed relationship between liking and perceived similarity affects many different aspects of person perception. First and foremost, our research shows that the liking-breeds-differentiation principle does not lead to the perception that disliked persons are generally more homogenous than liked persons. Given our data, the idea that people's preference to seek positive interactions and to avoid negative ones creates the perception that disliked members of their social world are all the same is not valid. Instead, people are surrounded by a social world in which negativity comes in various different forms and impressions about disliked persons contain little, but highly diverse information.

Positive similarities and negative differences. Another of our model's implication is that negative attributes tend to be different across people while positive attributes tend to be similar across people. Thus, while their negative traits make people unique, their positive traits make them similar. We directly tested this prediction in Experiment 7 and showed that focusing on the good (bad) aspects of other people makes them appear more (less) similar to each other. This notion has a powerful implication for apparent effects of mood on perceived variability. Accordingly, positive mood increases the perception of inter- and intragroup homogeneity. For example, happy participants viewing behavioral descriptions of highly variable groups perceived these groups as less diverse than participants in neutral moods exposed to the same descriptions (Queller, Mackie, & Stroessner, 1996). The authors suggested that happy perceivers focus on similarities rather than differences (Stroessner, Mackie, & Michalsen, 2005). When information that differentiates group members is made salient, happy participants no longer perceive them as more homogenous than their neutral mood counterparts. In line with this idea, Estes, Jones, and Golonka (2012) found that positive primes increased similarity ratings of unrelated social categories (e.g., musicians & dentists), and in an experiment by Isen and Daubman (1984), happy participants perceived exemplars as more similar to a given category than participants in neutral mood. The present research might provide a new explanation for these effects by assuming that participants in positive mood attend more strongly to positive and thus mood congruent information (e.g., Forgas & Bower, 1987). As our research shows that positive attributes are usually similar among objects and persons, representing stimuli (e.g., a group of people) primarily by their positive attributes will make them appear more similar to one another. The result that happy perceivers focus on similarities would therefore not be the explanation, but the outcome. Future research should test the possibility that the density principle together with the simple notion of mood-congruent processing can account for the well-documented finding that people in a good mood perceive the world in a more inclusive and integrative way (e.g., Bless & Fiedler, 2006).

Social comparison processes. The established positive relationship between liking and perceived similarity also has implications for social comparison processes. According to the selective accessibility model of comparison (Mussweiler, 2001; 2003), initial assessments of similarities between targets and standards determine whether judgments about the target are assimilated towards or contrasted away from the standard. If target and standard are perceived to be similar, target-congruent information is rendered accessible during the subsequent comparison process resulting in assimilation; when target and standard are perceived as dissimilar, target-incongruent information is more accessible resulting in contrast. Our results suggest that assimilation effects are more likely to occur when people compare others they like while contrast effects should more frequently occur when people compare others they dislike. At the same time, focusing on the similarities between other people might lead to more positive evaluations than focusing on the dissimilarities between people. This follows from the notion that people more strongly differ regarding their negative attributes than their positive attributes as a result of the greater diversity of negative attributes.

2.11.2 Reconciliation of two contradicting principles.

Results from the present work follow the density hypothesis and contradict the liking-breeds-differentiation principle. However, we believe that both principles co-exist. To reconcile both theoretical accounts, we have to ask why on the one hand, liking is usually associated with greater differentiation and why on the other hand positive impressions are less diverse than negative impressions.

As outlined earlier, there is direct evidence for the influence of liking on differentiation. Smallman and Roese (2008) demonstrated that participants sort liked objects into more categories than disliked objects, indicating that liked objects are perceived as less similar. However, the authors used an evaluative conditioning procedure where neutral symbols (CS) were paired with either positive or negative pictures (US). Thus, the valence of the to-be-categorized target stimuli did not stem from the stimuli's own attributes. In this case, the density principle does not apply as it is based on an ecological diversity of attributes that positive and negative stimuli display.

A well-studied type of phenomena where liking also goes along with greater differentiation includes the outgroup homogeneity effect and the cross-race effect (Feingold, 1914; Quattrone & Jones, 1980; Park & Rothbart, 1982; Young, et al., 2012). Contrary to the density principle, people typically perceive members of their preferred in-group as less homogenous than members of an out-group. However, the preference for the in-group is most likely not based on certain positive and negative *attributes* of in- and out-group members (as proposed in our model), but on familiarity. People simply like others who they encounter frequently (Zajonc, 1968), which is also visible in people's general preference for objects, faces, or persons that are prototypical or “average” regarding a given environment (e.g., Langlois & Roggman, 1990). In other words, in-group members might be liked because social perceivers are repeatedly exposed to them, making them more familiar.

While familiarity leads to in-group preference it also decreases perceived similarity among in-group members (Linville et al., 1989). Given the simultaneous influence of familiarity on liking and perceived similarity, liking is sometimes negatively associated with perceived similarity. That is, on a group level analysis, a given in-group is perceived as more diverse than a given out-group. However, this magnifying effect does not even out the similarity asymmetry inherent in the evaluative information environment and thus does not lead to a general dislike-homogeneity effect. The fact that positivity exists within tight boundaries limits the possible diversity of positive impressions.

In our view, the density asymmetry and outgroup homogeneity are not mutually exclusive but co-exist. For example, men should perceive female target persons as more similar to one another than male target persons (outgroup homogeneity). This effect should be even larger when men compare liked women and disliked men (outgroup homogeneity + density asymmetry). However, when men compare liked men and disliked women, the effect should vanish or even reverse (outgroup homogeneity – density asymmetry). Research designs that demonstrate the existence of both, the density asymmetry and outgroup homogeneity within the same paradigm would be a promising route for future research. Such a line of research could help to identify the boundary conditions of both effects and ultimately determine which effect dominates in which context.

One viable boundary condition that determines whether liking goes along with higher or lower perceived similarity is the type of attributes that similarity judgment are based on, more specifically, whether attributes are evaluative or non-evaluative. As argued above, people have more differentiated representations of other people they like as they accumulate a great amount of knowledge about these persons. This knowledge might often not be per se evaluative in nature. For example, one might learn about liked person’s jobs, hobbies, family situation and eating habits. Liked others might therefore appear highly diverse regarding these non-evaluative attributes. Yet, in terms of evaluative attributes, and in particular their personality traits, the representation follows our model’s prediction of a limited diversity of positive impressions.

2.12 Conclusion

In the opening of his novel “Anna Karenina”, Leo Tolstoi stated, “Happy families are all alike; every unhappy family is unhappy in its own way.” He thereby recognized the low diversity of positivity which we suggest to be a robust phenomenon that is inherent in the world humans live in, including their social world. Those people we represent with predominantly positive attributes and therefore like, appear highly similar. In the end, despite “relishing the nuances” of their personalities, our friends all seem alike.

Chapter 3 – A density explanation of valence asymmetries in recognition memory

Distinguishing between positive and negative information is essential for humans to navigating complex environments (Lewin, 1935); unsurprisingly, this distinction fundamentally influences human cognition as well. Previous research identified numerous asymmetries in the perception, processing, elaboration, storage, and retrieval of positive and negative information. These valence asymmetries are commonly explained by the affective potential of evaluative information (see Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001, for a review). Accordingly, the affective reaction of the organism alters cognitive information processing. A prominent example is the notion that negative information triggers deeper and more accommodative processing styles (Fiedler, & Bless, 2006; Taylor, 1991).

A different perspective on valence asymmetries is provided by the density hypothesis (Unkelbach, Fiedler, Bayer, Stegmüller, & Danner, 2008), which claims that positive and negative information differ ecologically regarding their diversity. That is, besides the “hot” potential of evaluative information to influence emotions, motivations, and behavior (e.g., such as approach and avoidance), it is assumed that there are systematic “cold” differences between positive and negative information. These differences should not depend on the information’s energetic potential, for example, due to its self- or survival relevance. Additionally, these differences exist independent of the emotional, motivational, or behavioral states of the organism (e.g., Lepper, 1994). Specifically, Unkelbach and colleagues suggested that there is a lower diversity and therefore a higher similarity among positive information compared to negative information, leading to higher “density” of positive information in mental representations. They argued that this ecological difference might explain observed valence asymmetries in processing of evaluative information. For example, the authors showed that positive information is processed faster than negative information; not because of differential affective reactions, but because of the differential density of positive and negative information (Unkelbach et al., 2008; Exp. 2). Here, we test whether the differential density of evaluative information influences recognition performance.

There is substantial evidence that stimulus similarity influences recognition memory. For example, perceptual recognition research shows that recognition is less accurate for prototypical stimuli (e.g. Busey & Tunnicliff, 1999); it is difficult to distinguish old from new stimuli when stimuli are highly similar. Specifically, similarity causes false recognition of stimuli that were not presented during a study phase, most prominently evident in the Deese-Roediger-McDermott (DRM) paradigm (Roediger & McDermott, 1995); for example when people falsely recall the word “sleep” after studying the words “bed”, “rest”, and “awake”. Given that positive information is overall more similar to other positive information, recognition of positive information should be less accurate than the recognition of negative information. Positive information should provoke more false alarms and thus weaken recognition performance. In terms of signal detection analysis (Stanislaw & Todorov, 1999), this translates to better discriminability for negative stimuli and a stronger response bias for positive stimuli.

A considerable amount of empirical evidence for such differential valence effects in recognition is already available (Inaba, Nomura, & Ohira, 2005; Ohira, Winton & Oyama, 1998; Ortony, Turner & Antos, 1983; Robinson-Riegler, & Winton, 1996). However, these effects have been traced back to the affective reaction of the organism in response to evaluatively and affectively connotated stimuli. Here, we will test whether positive and negative information's differential density accounts for the observed recognition asymmetry, over and above evaluative and affective influences. In the following, we provide an outline of the density hypothesis, explain how density and memory performance should be related, and compare predictions from density and affect-based explanations of recognition asymmetries.

3.1 The density hypothesis

The density hypothesis proposes that positivity comes with decreased diversity (Unkelbach et al., 2008; Unkelbach, 2012). Differential diversity of positivity and negativity is evident in many different domains. Most generally, positive states typically constitute the norm while negative states constitute some kind of deviation from that norm (Clark & Clark, 1977). There is usually one normal and thus positive state which is characterized by the absence of many abnormal and thus negative states (e.g. being healthy means not having any of many health-related abnormalities). As a result, negative states display a greater diversity than positive states. This principle reaches into language as there is a larger vocabulary for negative states than for positive states. This was shown for English verbs (Semin and Fiedler, 1992), German personality traits (Leising, Ostrovski, & Borkenau, 2012), as well as for English and Spanish emotion words (Schrauf and Sanchez, 2004). Another example is facial-attractiveness; attractive faces are alike, while there are many different ways to be unattractive (Potter, Corneille, Ruys, & Rhodes, 2007). The same principle extents to person perception in general as likable, or "positive" persons are perceived as more similar to one another compared to disliked persons (Alves, Koch, & Unkelbach, 2015; Leising, Ostrovski, & Zimmermann, 2013). And finally, the effect is also present in emotional experiences as there is one basic positive emotion ("joy"), but multiple distinct negative emotions (anger, disgust, fear, sadness; Ekman & Friesen, 1971; see also Ortony & Turner, 1990).

Based on lower diversity, positive information displays higher density in spatial models of mental representations than negative information. The original measure of the density construct consisted of pairwise similarity ratings that were analyzed using a multidimensional scaling procedure (MDS; Krumhansl, 1978). Stimulus density was defined as the average Euclidean distance of a stimulus to all other stimuli of the same valence in a given stimulus set (Potter et al., 2007; Unkelbach et al., 2008). As of now, we and others have found this density asymmetry for a variety of different stimulus classes including nouns, trait words, self-generated words, as well as IAPS-pictures using different measures like multidimensional scaling the spatial arrangement method (SPAM; Goldstone, 1994) and Google co-frequency analysis (Bruckmüller & Abele, 2013; Koch, Alves, Krüger, & Unkelbach, 2014; Unkelbach, Guastella, & Forgas, 2008b). In all these studies, participants judged positive stimuli as more similar to one another than negative stimuli. The density asymmetry seems to be a general and robust phenomenon of evaluative information ecologies (see Unkelbach, 2012, for a discussion).

Similarity as used in the current theoretical framework is defined by its experimental operationalization as spatial distance. When doing pairwise comparisons or spatial arrangements of stimuli, it is likely that participants rely on different components of similarity when making their judgments such as semantic similarity, feature overlap, associative strength, and frequency of co-occurrence (Maki, 2008). For example, results from the Google co-frequency analysis show that positive words more frequently co-occur across web pages than negative words while this co-occurrence is substantially correlated to participant's similarity ratings (*r* = .60) (Koch et al., 2014, see also Lund & Burgess, 1996). In a similar vein, people also group positive words into fewer categories than negative words (Koch et al., 2014). We suggest that higher perceived similarity, stronger associative relations, more frequent co-occurences, and more inclusive categorizations of positive compared to negative information are all different observable phenomena of the same common principle. In the present work we refer to this principle as density in a spatial model of representation (e.g. Goldstone, 1994; Nosofsky, 1992; Shepard, 1987).

We believe the differential density of positive and negative information is ecological, not affective in nature. Affective reactions are located within the organism – that is, organisms react to stimuli depending on their individual state, their current motivational situation, or their learning history. Ecological effects are located in the environment and depend on its contextual structure. That is, an ecological property changes with regard to the environment, while an affective property changes with regard to the organism. Applied to evaluative information and the construct of density, the same stimulus might be similar to other stimuli in one environment, but dissimilar to other stimuli in another environment. Consequently, the differential effects of stimulus valence might shift depending on the structure of the information environment. However, as the data described above suggests, across most environments, positive stimuli seem to be more similar to one another than negative stimuli.

The proposition that information valence is ecologically confounded with information density has a number of implications for information processing. Stimulus density or similarity is a fundamental determent of cognitive processes ranging from attention (Nosofsky, 1986; Ward, Duncan, & Shapiro, 1997), visual search (Phillips, Takeda, Kumada, 2006), storage (Mate & Baqués, 2009), retrieval (Nosofsky, 1988, 1991; Glanzer, Knoppenaal, Nelson, 1972; Lewandowsky, Farrell, 2008), and processing speed (Unkelbach et al., 2010) to evaluative judgments (Montoya, Horton & Kirchner, 2008). The differential density might therefore account for a number of observed valence asymmetries in information processing (Unkelbach, 2012). The present work tests the explanatory power of the density hypothesis in the domain of recognition memory.

3.2 Density and memory performance

Increasing the similarity or relatedness among stimuli increases false alarm rates. This effect appears for stimuli such as alphanumeric characters (Flagg, 1976; Reitman & Bower, 1973), geometric shapes (Medin & Schaffer, 1978; Nosofsky, 1991; Nosofsky, Clark, & Shin, 1989), pictures (Koutstaal & Schacter, 1997; Strack & Bless 1994), faces (Busey & Tunnicliff, 1999; Vokey & Read, 1992), words (Brainerd, Reyna, Mojardin, 1999; Dyne, Humphreys, Bain & Pike, 1990; Montefinese, Zannino, & Ambrosini, 2014; Postman, 1951; Roediger & McDermott, 1995), and sentences (Cantor & Engle, 1993; Holmes, Waters & Rajaram, 1998).

One explanation for this similarity effect is provided by global activation theories of recognition memory. They assume that stimulus classifications as old or new depend on the signal strength or the echo that stimuli evoke from memory (Gillund & Shiffrin, 1984; Hintzman, 1988; Wixted, 2007). If the echo exceeds a recognition threshold, participants provide an “old” response. The echo strength of a given stimulus is a function of the summed similarity with presented (“old”) stimuli stored in memory (Fiedler, 1996; Nosofsky, 1988, 1991). The left part of Figure 6 illustrates this principle in a simple subsymbolic memory network. The activation vectors represent information stored in memory and new information (e.g., new positive vs. new negative words); stimuli that are more similar to other stimuli produce stronger echos, and therefore, high similarity produces false alarms. Note again that similarity can have different components (semantic similarity, associative strength, co-occurrence) that might all increase echo strength.

As we assume that positive stimuli are more similar to one another than negative stimuli, the likelihood of falsely responding “old” to a new stimulus should be higher for positive stimuli. The right part of Figure 6 illustrates the same principle for a symbolic network. However, this effect should arise independent of the specific memory architecture (e.g., symbolic vs. sub-symbolic networks), as long as it allows to model similarity. The association between similarity and false alarms also follows from dual process models of recognition which assume false alarms to depend on the familiarity that a stimulus evokes (Mandler, 1980, Yonelinas, 1994). Similarly to the echo strength concept, familiarity is supposed to increases with increasing similarity among stimuli (Verde, 2004).

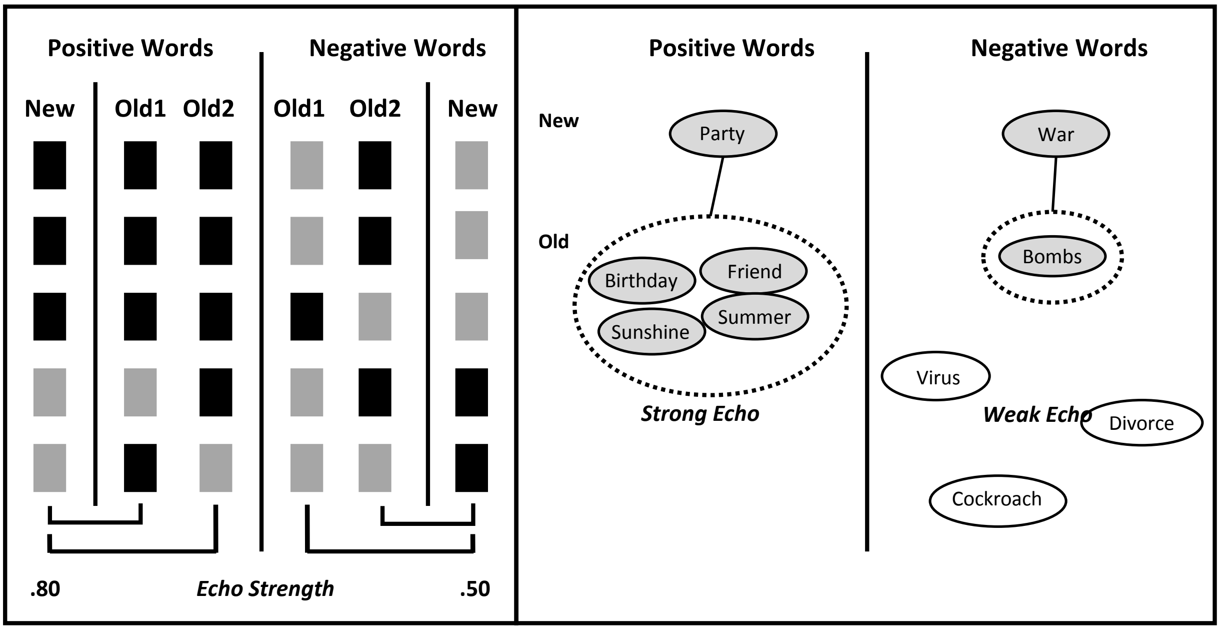

Figure 6. The left part illustrates a subsymbolic memory model along with the memory echo strength (simple matching coefficient) for a new positive and a new negative word elicited by old positive and old negative words. Assuming positive words are more similar to one another than the negative words, the resulting echo is stronger for the new positive word than for the new negative word and so is the likelihood for a false alarm. The right part illustrates the same principle in a symbolic memory model.

While similarity causes false recognition, it typically does not lead to an equal increase in hits which is why recognition performance is impaired for highly similar stimuli (Anderson & Reder, 1999; Cantor & Engle, 1993; Dyne et al. 1990; Shiffrin et al., 1995; Verde, 2004). Using a global activation model, Zaki and Nosofsky (2001) predicted and found that a larger proportion of new stimuli are pushed past the response criterion than old stimuli because participants usually correctly classify the majority of stimuli[4]. Another explanation is provided by dual process models as they suggest similarity affects familiarity and recollection in opposite ways (Mandler, 1980, Yonelinas, 1994). Similarity supposedly increases familiarity and thus false alarm rates; but similarity makes the recollection of a specific stimulus more difficult (Gillund & Shiffrin, 1984; Wixted, Ghadisha, & Vera, 1997). As the correct recognition of old stimuli ("hit") is a function of both, familiarity and recollection, similarity always produces a stronger increase in false alarms than in hits (Joordens & Hockley, 2000). Assuming that high similarity among stimuli increases false alarms without a corresponding increase in hits, we expected recognition advantages for negative stimuli over positive stimuli.

3.3 Affect or density?

Several studies already reported a similar recognition advantage for negative information over positive information. In a study by Ortony and colleagues (1983), false alarm rates for positive sentences were larger than for negative sentences, while hit rates were unaffected by valence. Using Jaboby's process dissociation paradigm (1991), Robinson-Riegler and Winton (1996) reported that positive words were more likely to be falsely recognized under an exclusion instruction than negative words. Ohira and colleagues (1998) replicated this pattern using Japanese word stimuli. Inaba and colleagues (2005) also reported a recognition advantage for negative words as they found larger false alarm rates for positive words and equally large hit rates for positive and negative words.

All these studies provide an affective explanation for the recognition advantage of negative information. Accordingly, valenced stimuli elicit affective reactions within participants, which influence encoding. Specifically, "…negative events appear to elicit more physiological, affective, cognitive, and behavioral activity and prompt more cognitive analysis than neutral or positive events." (Taylor, 1991, p.67) As deeper encoding benefits memory performances (e.g., Craik & Tulving, 1975), negative information is expected to show the described recognition advantage.

A related idea suggests that negative mood/affect leads to accommodative, bottom-up processing that is sensitive to the details of the external world, while positive mood/affect leads to assimilative, top-down processing which relies on pre-formed internal schemas and heuristics (Bless & Fiedler, 2006). A number of findings suggest that an accommodative processing style benefits memory accuracy while an assimilative style produces false memories (Forgas, Goldenberg, & Unkelbach, 2009; Fiedler, Asbeck, & Nickel, 1991).

Besides processing depth and processing style, Monin (2003) postulated the "warm-glow" heuristic, claiming that people mistake positive affect for familiarity. The author showed that participants perceived attractive faces as more familiar (see also Garcia-Marques, Mackie, Claypool, & Garcia-Marques, 2004). Using an old-new recognition task, Monin (2003, Study 3) also found that participants erroneously recognized positive words more often than negative words. Such misperceptions of positive affect as familiarity could thus increase familiarity for positive stimuli and thereby cause false recognition.

The processing depth explanation, the processing style explanation, and the "warm-glow" heuristic all suggest the recognition asymmetry is caused by "hot" valence–based affect, that is, the energizing potential of evaluative information. In contrast, as delineated above, we predict the asymmetry to be based on "cold" stimulus density, a property that, albeit ecologically confounded with valence, does not require affect to exert an influence. Again, density is a property of a stimulus within its information ecology and is thus not fixed (Unkelbach, 2012). For example, the usually very distinct stimulus "bombs" has high density in the context of war and weapons, and accordingly will produce high rates of false alarms. But in the context of spiders and snakes, the stimulus “bombs” has low density and will produce few false alarms. While affect-based explanations predict general differences between positive and negative information, the present account allows for *a priori* predictable alternative outcomes depending on the respective information ecology.

3.4 Overview and predictions for the following experiments

We conducted two old-new recognition experiments to disentangle density effects in recognition from affective influences. Based on the density hypothesis, we predicted that false alarm rates are higher for positive words than for negative words, while the same should not be true for hit rates. Consequently, discriminability should be higher for negative words while positive words should elicit a stronger response bias. Crucially, we predicted this effect to depend on the similarity among the word stimuli (density) and not on the valence of the words as suggested by affect based-explanations. We tested these hypotheses using the same sample of word stimuli that were used by Unkelbach and colleagues (2008) in their empirical test of the density hypothesis. It contains the 20 most positive and the 20 most negative words from a set of 92 words that are frequently used in experimental social psychology (Fazio et al., 1986; Bargh et al., 1992; Klauer & Musch, 1999). Using this stimulus sample ensures that stimuli have strong valence, and has the advantage that density parameters for the individual words, that is, their average similarity to the other stimuli in the set, are already available. In addition, it represents a standard and often used set of evaluative stimuli.

3.5 Experiment 1

Experiment 1 employed an old-new recognition task. First, we aimed at replicating a recognition advantage for negative words on the participant level. Next, regression analysis on the item level aimed to identify the underlying processes, by regressing stimuli’s false alarm and hit rates on their valence and density. Additionally, the regression included word frequency which has been shown to strongly influence recognition performance as well (Arndt & Reder, 2002; Glanzer & Adams, 1990).

3.5.1 Method.

Participants and Design. 183 students (106 women and 77 men) of the University of Cologne participated for 3€ or course credit. All participants were native German speakers. Stimulus valence was the only experimental factor and was manipulated within participants.

Stimulus Materials. As described, the 20 most positive and the 20 most negative words based on German norm ratings by Klauer & Musch (1999) from a set of 92 frequently used attitude objects (Fazio et al., 1986) served as stimuli.

Procedure. Participants arrived in the laboratory and were seated in front of a computer. After completing a consent form, the experimenter started a Visual Basic program that presented instructions, stimuli, and recorded the dependent variables. The program instructed participants to pay close attention while they would see several word series. The experiment had 3 phases, separated by 2 filler tasks. The first phase presented all 40 words in randomized order and was administered to familiarize participants with the stimuli and thus make the subsequent recognition task more difficult[5]. Each stimulus appeared for 1000 ms followed by blank screen that appeared for 1200 ms. Participants then worked on a filler task for about 8 minutes. The subsequent learning phase instructed participants to pay close attention while they would be presented with 20 out of the 40 word stimuli from the first phase. The program then presented 10 positive and 10 negative randomly selected words from the original 40 items. After a second filler task, which took about 5 minutes, the test phase presented all 40 stimuli and participants decided for each word whether it was present during the learning phase or not. Participants indicated their decision by pressing either the “L” key (“The word was presented in the second phase”) or the “A” key (“The word was NOT presented in the second phase”). Each phase of the experiment presented stimuli in a newly-randomized order for each participant. At the end, participants were thanked, paid, and informed about the purpose of the experiment.

3.5.2 Results.

Analysis across participants. Prior to inferential analysis, we removed data from seven participants because their memory performance did not exceed chance. For the remaining participants, we calculated participants’ false alarm and hit rates separately for positive and negative stimuli. Figure 7 illustrates that the mean false alarm rates were significantly higher for positive stimuli than for negative stimuli ($M_{\text{pos}} = 0.29$, $SD_{pos} = 0.19$ vs. $M_{\text{neg}} = 0.23$, $SD_{neg} = 0.17$), $t(175) = 3.68$, $p < .001$, $d = 0.33$, while the hit rates for the positive words were only slightly higher than for the negative words ($M_{\text{pos}} = 0.76$, $SD_{pos} = 0.16$ vs. $M_{\text{neg}} = 0.75$, $SD_{neg} = 0.16$), $t(175) = 1.18$, $p = .241$.

Based on the hit- and false alarm rates, we calculated participants’ signal detection parameters *d’* and *C* for the positive and negative words (Stanislaw & Todorov, 1999). Higher *d’* values indicate better discrimination ability, while $d' = 0$ indicates inability to distinguish between old and new stimuli. The response bias parameter *C* indicates the tendency to respond “new” or “old”. *C* values of 0 indicate no bias; *C* values higher than 0 indicate a tendency to respond “new”, while *C* values lower than 0 indicate a tendency to respond “old”. The data confirmed our hypothesis as *d’* was higher for the negative words ($M = 1.59$, $SD = 0.81$) than for the positive words ($M = 1.46$, $SD = 0.80$), $t(175) = 2.10$, $p = .038$, $d = 0.17$. In addition, *C* was higher for the negative words ($M = 0.04$, $SD = 0.39$) than for the positive words ($M = -0.09$, $SD = 0.42$), $t(175) = 3.32$, $p = .001$, $d = 0.33$. As predicted, participants’ discrimination ability was higher for negative words, while participant’s tendency to respond “old” was stronger for the positive than for the negative words.

Analyses across stimuli. Having established the basic pattern, we tested our core hypothesis that stimulus density influences recognition performance independent of and beyond valence. We therefore calculated hit and false alarm rates separately for each stimulus[6]. In addition, we obtained density indices for each word from Unkelbach and colleagues (2008). The density index designates the mean Euclidean distance of a word to all words of the same valence in a multidimensional space; thus, it is a metric of similarity. The multidimensional space is calculated based on pairwise similarity ratings using a multidimensional scaling procedure. To obtain a continuous measure for word valence, 8 research assistants rated the words' valences on scales ranging from-5 (negative) to +5 (positive). In addition, we obtained word frequencies estimates from Klauer and Musch (1999).

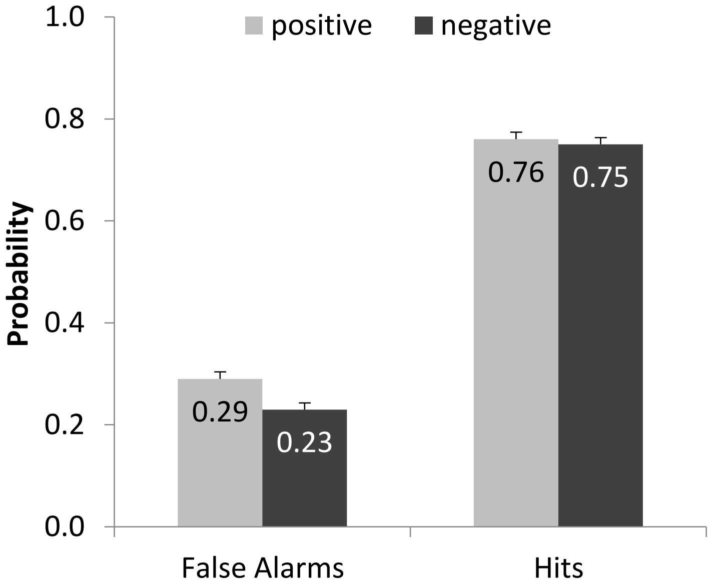

Figure 7. False alarm and hit probabilities among the positive and negative word stimuli. The error bars represent standard errors of the means.

Two regression analyses tested the influences of density and valence on false alarms and hits. A first model predicted false alarms from word valence and frequency and found an influence of both predictors. More frequent words had a higher chance of being falsely recognized (β = .36, *p* = .020), a well-known phenomenon (e.g., Glanzer, Adams, Iverson, & Kim, 1993). In addition, positive words elicited more false alarms than negative words (β = .29, *p* = .055), even though this effect did not reach conventional levels of significance. We attribute this to a lack of power, as the item-level analysis had only 39 degrees of freedom. A second regression model added density as a predictor. Table 2a shows that when density was included, it was the strongest and only significant predictor of false alarms. Specifically, the partial correlation between valence and false alarms dropped from *r* = .31, *p* = .055 to *r* = .05, *p* = .772. Thus, density accounted for the effect of a word's valence on its probability to be falsely recognized.

Next, we conducted a similar regression procedure with hit rates as the criterion. Table 2b shows that neither valence nor density predicted hit rates. Frequency showed a marginally significant influence on hit rates, indicating that infrequent stimuli had a higher chance of being correctly recognized. This reflects typical word frequency effects in recognition (see Murdock, 2003, for a review).

Table 2a

Results of a multiple regression analysis predicting a word's probability to be falsely classified as "old" (false alarm rate) from valence, density, and frequency across the 40 word stimuli.

Predictor	*β*	*t*	*p*	Partial r^2	Simple Correlations
Density	-.469	-2.82	.008	-.426	-.554
Valence	.046	0.29	.772	.048	.284
Frequency	.222	1.56	.128	.252	.351

Table 2b

Results of a multiple regression analysis predicting a word's probability to be correctly classified as "old" (hit rate) from valence, density, and frequency across the 40 word stimuli.

Predictor	*β*	*t*	*p*	Partial r^2	Simple Correlations
Density	-.059	-0.30	.766	-.050	.470
Valence	.102	0.54	.592	.090	.137
Frequency	-.288	-1.71	.091	-.275	-.274

3.5.3 Discussion.

Experiment 1 replicated the standard valence asymmetry in recognition performance, that is, higher false alarm rates among positive stimuli and equal hit rates among positive and negative stimuli. As a result, discriminability was higher for negative stimuli and response bias was larger for positive stimuli. Regression analyses on the item level explored the cause of this asymmetry. As expected, stimulus density was the best predictor of false alarms and fully accounted for the effect of valence on false alarms. Thus, Experiment 1 supported the idea that higher similarity among positive stimuli, as indexed by stimulus density, causes false recognition and thereby produced an apparent valence asymmetry.

Experiment 1 utilized the "natural" covariation of valence and density (i.e., positive information clusters more densely). Experiment 2 provides a stronger test and varied density and valence orthogonally between a "natural" and a "reversed" condition. If the density explanation holds true and the recognition asymmetry is not a function of valence *per se*, it should be possible to create ecologies in which negative stimuli produce more false alarms than positive stimuli.

3.6 Experiment 2

Experiment 2's "natural" condition featured positive words that were similar and negative words that were dissimilar. The "reversed" condition featured positive words that were dissimilar and negative words that were similar, while mean stimulus valence was constant across conditions ($M_{\text{natural}} = 4.79$, $SD = 3.44$ vs. $M_{\text{reversed}} = 4.75$, $SD = 3.69$, $t(30) = 0.04$, $p = .97$). Figure 8 presents the stimuli plotted in a 2-dimensional similarity space based on similarity ratings from Unkelbach et al. (2008). This selection directly tested the affect and density explanations for recognition asymmetries. Affect-based explanations predict recognition to be a function of valence, which statistically translates into a main effect in the present design. The density-based explanation predicts recognition to be a function of stimulus similarity, which statistically translates into an interaction effect between valence and condition.

3.6.1 Method.

Participants and Design. 74 students (58 women, 16 men) of the University of Cologne participated for 3€ or course credit. All participants were native German speakers. Experiment 2 did not hinge on correlational variation of the variables, but manipulated them via pre-selection. Accordingly, the necessary participant sample was smaller than in Experiment 1. Participants were randomly assigned to the "natural" and "reversed" conditions; in the former condition, positive stimuli were more similar to one another than negative stimuli, while in the latter condition, negative stimuli were more similar to one another than positive stimuli.

Stimulus Materials and Procedure. We created four stimulus subsets which realized the orthogonal combinations of valence and similarity (positive/similar; positive/dissimilar; negative/similar; negative/dissimilar). Each set contained 8 stimuli from the 40 word stimuli used in Experiment 1. We selected stimuli based on density indices and additional visual inspection of the multidimensional scaling solution. Visual inspection was necessary as two stimuli might be very dissimilar to the whole set (resulting in low density), but very similar to each other. Visual inspection ensured that in such cases, only one of them was included in the subset. Using similarity ratings from Unkelbach et al. (2008), we calculated new density indices (i.e., average Euclidean distances) for each word in each subset with lower values representing higher density. The respective means confirmed that words in the positive/similar ($M = 3.89$, $SD = 0.53$) and negative/similar ($M = 3.89$, $SD = 0.92$) subsets were more similar than those in the positive/dissimilar ($M = 7.34$, $SD = 0.81$) and negative/dissimilar ($M = 8.88$, $SD = 1.01$) subsets ($t(30) = 12.09$, $p < .001$). The dissimilar sets did not fully dissolve the confound between valence and density, as the positive/dissimilar words were still more similar compared to the negative/dissimilar words, *t*(14) = 3.36, *p* = .005. Yet, as Figure 8 shows, the experimental density difference within a given condition was fully established. The four subsets did not differ in word frequency (*F*(3, 28) = 1.66, *p* = .198).

We combined the positive/similar and negative/dissimilar subsets to serve as stimuli in the "natural" condition, while the positive/dissimilar and the negative/similar subsets served as stimuli in the "reversed" condition. Thus, each condition contained 16 word stimuli as illustrated in Figure 8. For each participant, the computer randomly determined 4 of the 8 positive and 4 of the 8 negative stimuli as "old" stimuli and the remaining as "new" stimuli. Except for the stimuli, the procedure of the recognition task was identical to Experiment 1. That is, the task consisted of three phases separated by to filler tasks. The first phase presented all 16 word stimuli in order to familiarize participants with the stimuli. After a first filler task, the subsequent learning phase presented 8 randomly selected words (4 positive and 4 negative) from the original 16 items. After a second filer task, the test phase presented all 16 stimuli and participants decided for each word whether it was present during the learning phase or not.

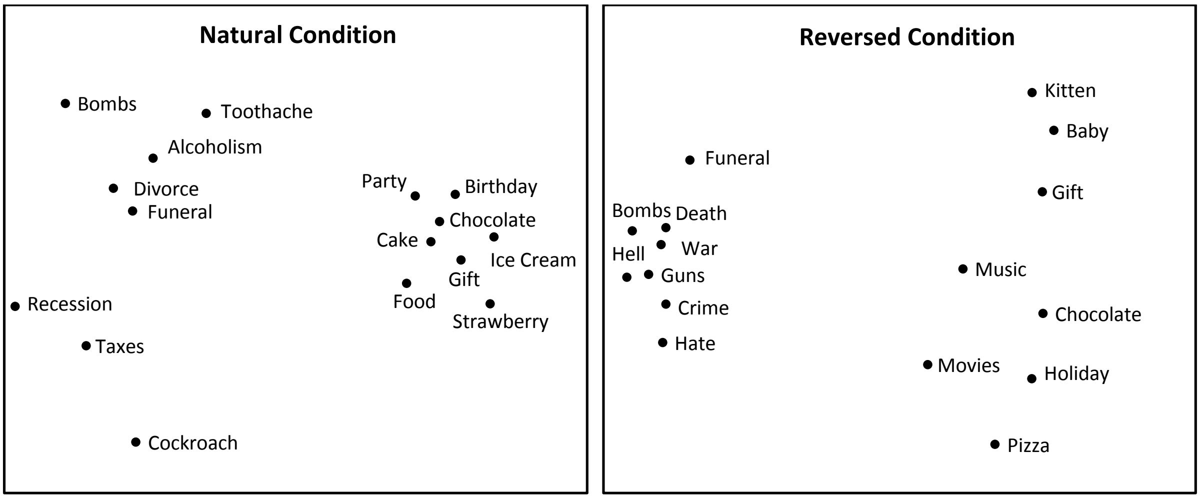

Figure 8. Spatial density differences for the 16 stimuli in the natural condition and the 16 stimuli in the reversed condition in a 2-dimensional similarity space.

3.6.2 Results.

Analysis across participants. Prior to inferential analyses, we removed data from three participants because their memory performance did not exceed chance. For the remaining participants, we calculated false alarm and hit rates separately for the positive and the negative stimuli. We conducted a 2(condition: natural vs. reversed) x 2(valence: positive vs. negative) ANOVA with repeated measures on the last factor and false alarm rates as the dependent variable. The analysis yielded a main effect of valence: false alarm rates were higher for the positive words than for the negative words (M_{pos} = 0.28, SD_{pos} = 0.21 vs. M_{neg} = 0.21, SD_{neg} = 0.21), *F*(1, 69) = 5.01, *p* = .028,

η^2 = 0.07. However, the predicted interaction explained a much larger part of variance. As the upper-left part of Figure 9 shows, false alarm rates were higher for positive words than for negative words in the natural condition, while the opposite was true in the reversed condition, $F(1, 69) = 25.14$, $p < .001$, η^2 = 0.27. Similar to Experiment 1, hit rates did not differ between experimental conditions (see upper-right part of Figure 9), all *F*s < 1, *ns*.

We then computed participants' SDT estimates separately for positive and negative words. First, we conducted a 2(condition: natural vs. reversed) x 2(valence: positive vs. negative) mixed ANOVA with repeated measures on the last factor and *d′* as the dependent variable. The only significant effect was the predicted interaction of condition and valence. As Figure 9's lower-left part illustrates, *d′* was higher for negative words than for positive words in the natural condition, while in the "reversed" condition, *d′* was larger for positive words than for negative words, $F(1, 69) = 17.62$, $p < .001$, $\eta^2 = .20$.

We conducted the same analyses with *C* as dependent variable. This analysis yielded a main effect for valence: *C* was larger for negative words ($M = -0.06$, $SD = 0.48$) than for positive words ($M = -0.21$, $SD = 0.44$), $F(1, 69) = 4.68$, $p = .034$, $\eta^2 = .06$. However, a much larger part of variance was explained by the predicted interaction which is illustrated in Figure 9's lower-right part: *C* was smaller for the positive words than for the negative words in the natural condition, while in the reversed condition, *C* was smaller for the negative words than for the positive words, $F(1,69) = 11.27$, $p = .001$, $\eta^2 = .14$.

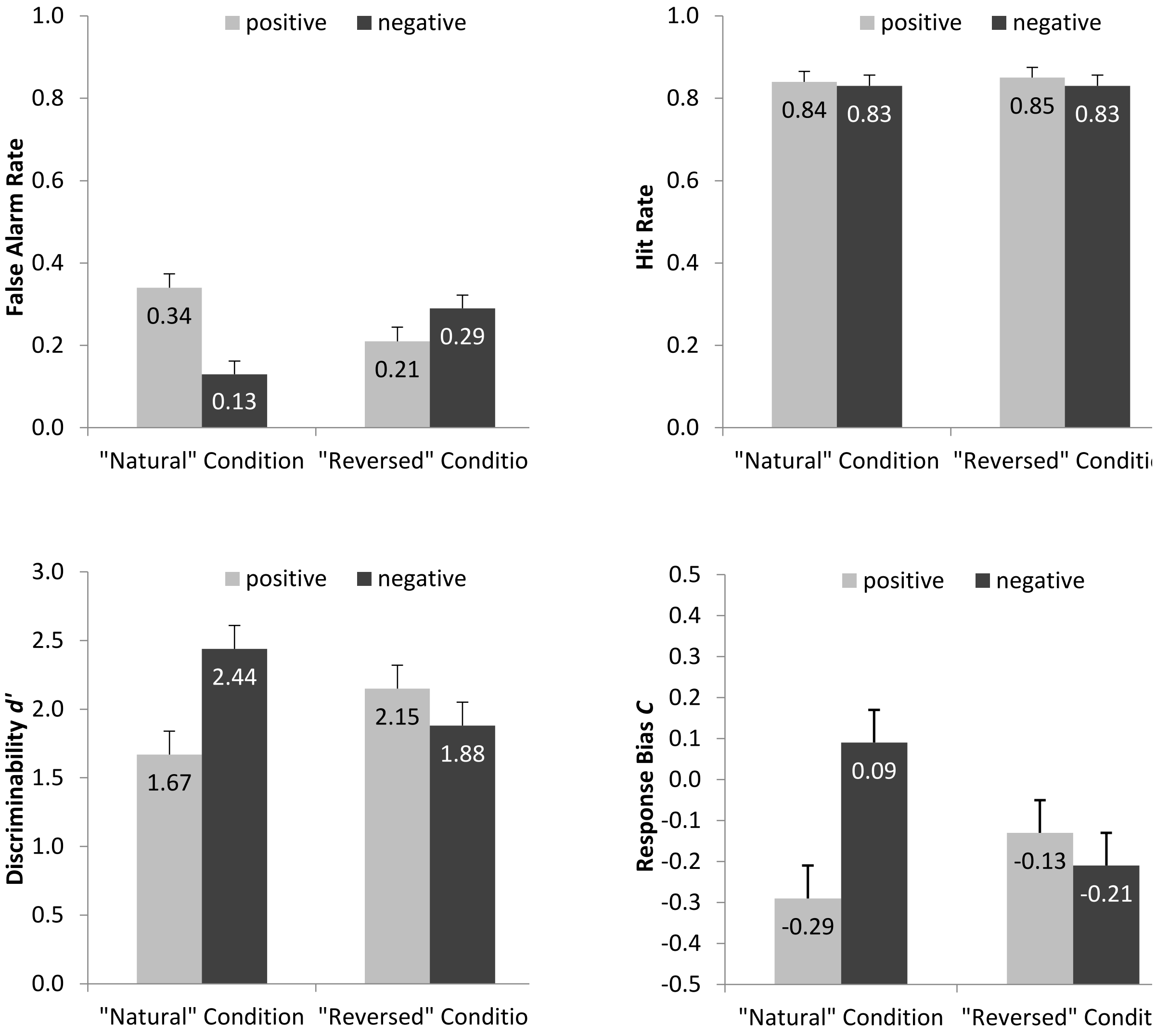

Figure 9. False alarm rates, hit rates, discriminability and response bias for the positive and negative words when positive words are more similar than negative words (“natural” condition) and when negative words are more similar than positive words (“reversed” condition). Larger *d’* values indicate better discriminability. *C* values lower than 0 indicate a tendency to respond “old”. The error bars represent standard errors of the mean.

Analysis across stimuli. Similar to Experiment 1, we tested the influence of density and valence on the stimulus level. Therefore, we again calculated hit and false alarm rates separately for each stimulus across both conditions. As there were 16 stimuli in each condition we calculated a total of 32 hit and false alarm rates. Two of the same positive and two of the same negative stimuli were used in both conditions simultaneously (chocolate, gift, funeral, bombs). Thus, these stimuli appeared twice among the 32 stimuli. While they had the same valence values in both conditions, their density value varied due to the differential ecologies. Similar to Experiment 1, we conducted two regression analyses predicting false alarm rates and hit rates from valence, density, and frequency across all 32 words. As in Experiment 1, the density index was the only significant predictor for false alarm rates, while neither valence nor density had a significant influence on hit rates (Tables 3a and 3b).

Table 3a

Results of a multiple regression analysis predicting a word's probability to be falsely classified as "old" (false alarm rate) from valence, density, and frequency across the 32 word stimuli.

Predictor	B	*t*	*p*	Partial r^2	Simple Correlations
Density	-.570	-3.56	.001	-.558	-.603
Valence	.162	1.07	.293	.199	.209
Frequency	.049	0.30	.765	.057	.227

Table 3b

Results of a multiple regression analysis predicting a word's probability to be correctly classified as "old" (hit rate) from valence, density, and frequency across the 32 word stimuli.

Predictor	*B*	*t*	*p*	Partial r^2	Simple Correlations
Density	.049	0.24	.812	.045	.008
Valence	.073	0.38	.708	.071	.054
Frequency	.096	0.47	.643	.088	.068

3.6.3 Discussion.

Experiment 2 supports the proposed influence of density on recognition memory: With increasing similarity among stimuli, false alarm rates increase while hit rates remain mostly unaffected. Consequently, discriminability decreases and response bias increases. Furthermore, by varying density and valence across two conditions, the influences of the two competing predictors were directly tested against each other. We created two stimulus ecologies; one included the natural confound between valence and density while the other represented a reversed ecology in which negative words were more similar to one another than positive words. The natural ecology produced higher false alarm rates for positive words, while the reversed ecology produced higher false alarm rates for negative words. Consequently, discriminability and response bias also changed as a function of the ecology as evident by the interaction effects of valence and condition. This pattern directly follows from the density explanation, but affect-based explanations would predict a main effect of stimulus valence on recognition performance. For example, if the "warm-glow" of positive words increased familiarity and thus false alarms, this should also be the case in a reversed ecology. Indeed, we found a main effect of valence on false alarm rates and response bias as well; however, this effect was much smaller than the interaction effect, and might be due to the still present confound between valence and density in our stimulus subsets. Item-level analyses confirmed this explanation; across both conditions, the density index predicted false alarm rates, but a continuous measure of valence did not.

3.7 General discussion

The density hypothesis states that positive information is less diverse than negative information, resulting in differential density in mental representations (Unkelbach et al., 2008). This differential density may account for apparent valence asymmetries in the processing of evaluative information (Unkelbach, 2012). While valence asymmetries are commonly explained as “hot”, affect-induced processing asymmetries, the density hypothesis points to a “cold”, ecology-related explanation of the same asymmetries.

The present work shows that higher density among positive stimuli impairs memory accuracy, and thereby creates a recognition advantage for negative stimuli. Our account thereby provides an alternative explanation for the recognition advantage of negative information that does not rely on the affective reaction of the organism (Inaba et al., 2005; Ohira, et al., 1998; Robinson-Riegler, & Winton, 1996). Two experiments showed that the density asymmetry creates a recognition advantage for negative stimuli over and above evaluative and affective influences.

Experiment 1 found the standard recognition valence asymmetry for 20 strongly positive and 20 strongly negative nouns frequently used in research on evaluative information processing. Similarly to past findings (Ortony at al., 1983; Inaba et al., 2005), false alarm rates for positive words were higher than for negative words while hit rates were mostly unaffected. Regression analyses supported the density explanation, as stimulus density was the best predictor for false alarm rates and fully accounted for any effects of valence. Hit rates were unaffected by both density and valence. Experiment 2 varied density and valence between conditions, and again, memory performance was a function of density and not of valence. Results showed that in a reversed ecology, in which negative stimuli were more similar than positive stimuli, false alarm rates were higher for negative words than for positive words and consequently, discriminability was higher for positive words and response bias was stronger for negative words. This pattern is incompatible with a processing depth, processing style, and a “warm-glow” explanation, but directly follows from the density explanation. Although we do not claim that these explanations are generally wrong, the density effect dominated the valence influences at least within the present paradigm and stimulus-set.

Across both experiments, hit rates were unaffected by density and valence on the participant as well as on the stimulus level. While we predicted similarity to exert its influence mostly on false alarms and not on hits, the absence of an association between valence and hits casts further doubt on a processing depth/style explanation. To our understanding, deeper and more accommodative encoding of stimuli should make it easier to correctly classify new and old stimuli likewise, and thereby produce an effect similar to the strength-based mirror effect (Glanzer & Adams, 1990). This effect describes the widely observed phenomenon that false alarm rates are smaller and hit rates are higher for strongly encoded stimuli compared to weakly encoded stimuli (e.g. Stretch & Wixted, 1998).

3.7.1 Limitations and open questions.

The present work is limited by the fact that our observations are based on a sample of only 40 word stimuli. This poses the question of generalizability of our results in regard to the density asymmetry as well as to the recognition asymmetry. However, we did not generate the present set for our purposes, but used an existing set introduced by Fazio et al. (1986), which is frequently used in cognitive research.

As of now, we are confident that the density asymmetry in this stimulus sample and the following processing differences are general phenomena observable across many different domains; the density asymmetry has been found for nouns, trait words, self-generated words, IAPS-pictures, and mental representations of other people (Alves, et al., 2015; Bruckmüller & Abele, 2013; Koch et al., 2014; Leising et al., 2012; Potter, et al., 2007; Unkelbach, 2010; Unkelbach et al., 2008; Unkelbach et al., 2008b). However, given the limitation of the stimulus sample used in the present work it remains unclear to what extent the present findings can be generalized to different stimulus sets and procedures other than the old-new recognition task. Future research should address possible density-driven recognition asymmetries in domains other than language such as person perception, faces, pictures, sounds, tastes, and life events, using more exhaustive stimulus sets, and alternative recognition paradigms such as DRM lists or two alternative forced-choice tasks (e.g., Smith & Duncan, 2004).

Another open question relates to the different components of the density asymmetry and their contribution to the recognition effects reported here. As mentioned earlier, we argue positive information is less diverse than negative information, resulting in a higher density of positive information in mental representations. Higher density correlates with semantic similarity, associative strength, frequency of co-occurrence, and category inclusiveness (Koch et al., 2014), and may serve as the latent cause for observable valence asymmetries on these measures. The present experiments show that density strongly predicts recognition performance and accounts for valence effects; however, it is not clear which of the different components of density influences recognition. As argued earlier, there is good evidence that semantic similarity among stimuli increases false recognition (e.g. Montefinese et al., 2014). However, some researchers have challenged the idea that such item noise itself influences recognition performance (e.g., Dennis & Humphreys, 2001; Maguire, Humphreys, Dennis, & Lee, 2010). Specifically, it was suggested that participants rely on category labels (i.e., animals vs. vegetables) that are associated with stimuli as part of the recognition process

(Humphreys, Murray, & Koh, 2014). This would suggest that it is not the larger semantic similarity among positive stimuli that causes false recognition but the higher inclusiveness of positive categories. Likewise, it is possible that the greater tendency of positive stimuli to co-occur in the same context creates false recognition (e.g., Lund & Burgess, 1996). The ways in which word stimuli can be related to one another are multifaceted (e.g. Gagné, 2000, see also Estes and Jones, 2009) and future research should try to disentangle which kind of relatedness (e.g. semantic similarity, co-occurrence, category inclusiveness) influences recognition memory. It is possible that multiple processes simultaneously affect recognition in the same direction and thereby contribute the apparent valence asymmetry. Such a scenario would be in line with our conceptualization of the density asymmetry as it allows for multiple processes to be at work; inter-item similarity and thus, density may serve as the uniting latent cause for these multiple processes. As positive information is naturally less diverse than negative information, positive words are semantically more similar, co-occur more frequently within the same context, have stronger associative relations, and are divided into fewer categories than negative words (see Koch et al., 2014). All of these factors might increase false recognition, but would also relate to stimulus density.

Our results also imply that researchers who are interested in examining valence effects in cognition have to make an important decision regarding the stimulus sample. The sample can either incorporate the naturally occurring density asymmetry for the sake of external validity (e.g. Brunswick, 1955), or stimuli can be selected to wipe out this asymmetry in order to observe "pure" effects of valence. Examples for the latter strategy are studies that use the Deese–Roediger–McDermott recognition paradigm (e.g. Budson et al., 2006; Roediger & McDermott, 1995). These studies use lists that typically entail several associated words from specific positive or negative domains (e.g. sex, man, violate) that converge on a critical lure (e.g., rape; see Budson et al., 2006). The stimuli are often matched regarding their attributes including strength of associative relation. In such a pre-selected stimulus sample we would not expect to find the typical density asymmetry. In fact, recent studies have shown that in recognition experiments using these DRM lists negative words elicit *more* false recognition than positive words which is contrary to the classical recognition asymmetry and to the results of the present experiments (Brainerd, Stein, Silveira, Rohenkohl, & Reyna, 2008; Brainerd, Holliday, Reyna, Yang, & Toglia, 2010). The authors argue that negative valence enhances the familiarity of the semantic content of critical distractors. Thus, it is possible that when the naturally occurring density asymmetry is controlled for, negative valence might *increase* perceived familiarity and thereby decrease memory accuracy. This would be contrary to what affect-based accounts and the warm-glow heuristic would predict (e.g., Ohira et al., 1998; Monin, 2003).

The present findings might explain why in some experiments negative information has a recognition advantage and in other experiments, that pre-select stimuli to counterbalance associative relations (e.g, DRM experiments), and thereby erase the density asymmetry, the reversed is true. However, as another recent study that carefully controlled for associative relatedness of new and old stimuli did not find any difference in false recognition between positive and negative DRM lists, it is an open question whether affective valence per se influences recognition (Dehon, Larøi, Van der Linden, 2010). Our research shows that controlling for stimulus density is crucial for any research aiming to examine “pure” valence effects.

3.7.2 Conclusions about valence, affect, and cognition.

Valence asymmetries in cognitive processes are commonly explained to be a result of the affective response of the organism. Affect-based accounts require the assumption that the mere confrontation with a stimulus (e.g., the word “war”) elicits an affective reaction strong enough to influence its processing. However, we want to emphasize that humans can process affectively charged stimuli in a relatively “cold” and non-affective way. Further, valence asymmetries in cognitive performance might not always be due to the affective reaction of the organism but instead arise from the natural structure of the information ecology. One structural characteristic of evaluative information is that negative information is more diverse than positive information. Consequently, positive information has a higher density than negative information in people’s mental representation. We suggest density as an important variable that creates valence asymmetries independent of affect.

Finally, we want to point out that there is a large and convincing body of research on affect-induced processing asymmetries, and that we do not question their general existence. We do however want to call for caution, as positive and negative information does not only vary in affective potential but also in “cold” properties like density. If these “cold” properties account for “hot” effects, it need not impact the functional outcomes—our memory is still more accurate for negative than for positive information—but it will substantially alter explanatory models, applications, and interventions.

Chapter 4 – General discussion

In the present dissertation, I propose a robust, general, and objective property of the information environment, namely that negative information is more diverse than positive information. I suggest this asymmetry, first described by Unkelbach and colleagues (2008) as “The Density Hypothesis”, results because positive attribute ranges occupy moderate values and are surrounded by two distinct negative ranges. This non-extremity of positive attributes strongly characterizes the environment humans live in and traverses physical, chemical, biological and psychological realities. It is mirrored by the strong prevalence of attributes that are normally distributed across human populations. As I described in chapter 1, the greater diversity of negative information is a strong and robust empirical phenomenon that applies to different stimulus classes such as exhaustive lists of substantives and verbs, representative samples of substantives and real life events, as well as to pictures from the IAPS database (Koch et al., 2016). Importantly, the lower diversity of positive information is also visible in a higher co-occurrence frequency, a more objective measure of stimulus similarity. As I showed in chapter 2, the density asymmetry also applies to people's mental representation of their social world. Despite the fact that people collect more knowledge about other people they like, they perceive liked others as more similar and describe them with more of the same traits compared to disliked others. In chapter 3, I investigated a downstream processing asymmetry namely that negative information enjoys a recognition advantage over positive information. I showed that participants were better at discriminating old from new words when the words were negative as negative words produced less false alarms compared to positive words. Importantly, the greater diversity of negative information inherent in the word sample fully accounted for this recognition asymmetry. In a second step, I showed that when inter-stimulus density is manipulated orthogonally to stimulus valence, the recognition parameters are a function of density and not of valence. Together, the findings presented in my dissertation provide converging evidence for the idea that stimulus diversity, or density, is more than a mere confounder of stimulus valence. In fact, it seems to constitute the driving force behind many observable processing asymmetries ranging from processing speed, and categorizations to memory performance.

4.1 Limitations and open questions

To evaluate the density hypothesis at its current state and to identify possible limitations I will distinguish between the empirical phenomenon and the theoretical explanation provided. Regarding the phenomenon, the empirical evidence is quite strong as the density asymmetry has now been demonstrated in more than a dozen studies (Alves et al., 2015; 2016a; Gräf & Unkelbach, 2016; Koch et al., 2016, Unkelbach et al., 2008), typically resulting in large effect sizes. However, there might still exist stimulus classes to which the density asymmetry does not apply and future research could extend to stimuli such as sounds, smells, tastes, body movements, or behaviors.

What remains somewhat unclear about the density asymmetry is to what contents of similarity it relates. While spatial similarity solutions provide overall similarity estimates, they do not specify in what respect stimuli are similar. Some researchers have argued that similarity is too broad of a construct to denote a meaningful independent variable for information processing (e.g. Murphy & Medin, 1985). Accordingly, the perception of similarity is extremely flexible and context dependent, and stating that stimulus A is similar to stimulus B is meaningless unless it is specified in what respect they are similar. It seems plausible that when participants judge the similarities of different word pairs, they flexibly switch content dimensions. While this reasoning puts the usefulness of similarity as a theoretical concept generally in question, other researchers have pointed to the high stability of spatial similarity solutions, the high correspondence between different similarity measures, and to the high across-subject agreement in ratings of similarity (e.g. Medin, Goldstone, & Gentner, 1993; Nosofsky, 1988; Smith, Shoben, & Rips, 1974). In addition, similarity reliably predicts various other processing parameters such as processing speed, categorization, generalizations, and memory performance (e.g. Brainerd, Reyna, Mojardin, 1999; Dyne, Humphreys, Bain & Pike, 1990). The results presented in my dissertation are well in line with this more optimistic view on similarity as a meaningful explanatory concept. They show a strong correspondence between different measures of similarity as well as its predictive power. Particularly the results presented in chapter 3 show that similarity as a stimulus property outperforms stimuli’s affective potential in predicting memory performance. Furthermore, the proposed higher similarity of positive stimuli reliably occurred across different stimulus classes and similarity measures. While this speaks to the robustness of the density asymmetry, it also supports the idea that similarity, regardless of the specific respect in which it is interpreted, constitutes a meaningful explanatory concept. I am therefore confident that defining inter-stimulus similarity as an abstract, content-independent stimulus property is well justified.

Regarding the theoretical explanation of the density asymmetry that I provided based on the non-extremity of positivity, more limitations remain. This is partly due to the general and distal nature of the non-extremity principle. While this principle is evident on many different observational levels and traverses physical, chemical, biological and psychological realities, it is difficult to empirically confirm its causal influence on the density asymmetry. As it is typically the case with overarching theoretical frameworks in Psychology, the confirmation process consists of successive accumulation of evidence that does not disconfirm the theoretical framework. Throughout this process, alternative explanations should be disconfirmed. In the present work, I mainly focused on affect-based alternative explanations according to which valence asymmetries are intrapsychic phenomena. I ruled out these affect-based explanation by showing that downstream processing asymmetries like processing speed and recognition memory were a function of density and not of valence, which should be the best proxy for stimuli’s affective potentials. I further showed that the density asymmetry is evident on “objective” stimulus parameters like frequency of co-occurrence, which makes it unlikely that the density asymmetry is a mere intrapsychic phenomenon.

However, there might remain other extrapsychic factors that contribute to positive information’s lower diversity that are not based on the non-extremity of positivity. One might arise from the stronger “contagiousness” of negativity (Rozin & Royzman, 2001) which also constitutes an ecological reality. Consider the example that one drop of poison can ruin a hundred liters of drinking water while one drop of drinking water does not make a hundred liters of poison drinkable. Similarly, a livable environment turns non-livable even when it has only one non-livable attribute (e.g. temperature too low), but many livable attributes (e.g. right amount of oxygen, and water). This suggests that negative attributes might have a naturally stronger impact than positive attributes. Similar to the non-extremity principle, this would also lead to conjunctive conditions of positive qualities (e.g. right amount of oxygen AND water) and to disjunctive conditions of negative qualities (e.g. wrong amount of oxygen OR water) which should also result in a greater diversity of negative qualities.

Another factor contributing to the density asymmetry might be the prevalence of positive attributes in the environment. A characteristic of most information ecologies is that people encounter positive attributes more often than negative attributes. There are different explanations for that, but the strongest force behind the frequency asymmetry is probably the fundamental human motivation to maximize positive and to minimize negative experiences, which I described in chapter 2 (Denrell, 2005; Fazio et al., 2004; Thorndike, 1898). As a result, people are more often exposed to positive compared to negative information (Augustine, Mehl, & Larsen, 2011; Boucher & Osgood, 1969; Dodds et al., 2015; Rozin, Berman, & Royzman, 2010). If positive stimuli are more likely to occur in the environment than negative stimuli, (i.e., $p_{pos} > p_{neg}$), frequency of co-occurrence, amplifies this ordinal order to the square. In other words, while it is more likely to encounter a positive stimulus than a negative stimulus, it is a lot more likely to simultaneously encounter two positive stimuli than to simultaneously encounter two negative stimuli (for a formal model see Alves, Koch, & Unkelbach, *in prep.*). If we then assume that perceived similarity of two stimuli is partly a function of their co-occurrence frequency (e.g., Griffiths, Steyvers, & Tenenbaum, 2007; Landauer & Dumais, 1997), a higher perceived similarity of positive stimuli follows as well.

These considerations suggest that multiple ecological forces may contribute to the density asymmetry. Positive information’s non-extremity, weaker impact and its higher frequency are all ecological realities that ultimately lead to a lower diversity of positive information. Future research should attempt to quantify the unique influences of these factors and identify their diverging predictions. For example, the frequency-based explanation suggests that in ecologies where negative instead of positive information is prevalent, negative stimuli should be perceived as more similar to one another, while the non-extremity and the impact explanations predict no such reversal. Regardless of the possible contribution of other ecological factors to positive information’s lower diversity, I suggest that the non-extremity principle accounts for the largest part. First, it is most basic as it manifests itself at the level of a single attribute dimension. Second, it is a widely observed phenomenon that traversed physical, chemical, biological and psychological attribute dimensions. Lastly, it is well in line with the ubiquitous Gaussian distribution of attributes.

4.2 Implications

As various implications of the current research are already discussed in chapters 2 and 3, I focus here on another general implication of the current research which I believe has great potential to spark new lines of research, and to renew our understanding of the comparative nature of human information processing.

From the non-extremity of positivity it not only follows that positive information is more similar than negative information but also that similarities are more positive than differences. This reasoning can best be illustrated from a feature perspective (Tversky, 1974). Again, it is assumed that entities (e.g. objects or persons) can be described as collections of absent and present features. Figure 10’s left half illustrates the feature vectors for two objects A and B. The filled rectangles symbolize present features, the unfilled rectangles symbolize absent features. Figure 10’s right half shows that if a given feature is simultaneously present in both objects, the feature is shared and constitutes a similarity. If the feature is only present in one of the two objects, the feature is unshared and constitutes a difference. The greater diversity of negative features that I propose in the current work, is modelled by assuming that there are more negative than positive features that an object can possibly possess (here: 20 negative and 10 positive features). If we randomly determine a given number of positive and negative features as present in each object (here: 6 positive and 6 negative features), shared features are more likely to be positive than unshared features. The simple reason is that objects or persons can greatly vary regarding their negative features while the variation in their positive features is more restricted.

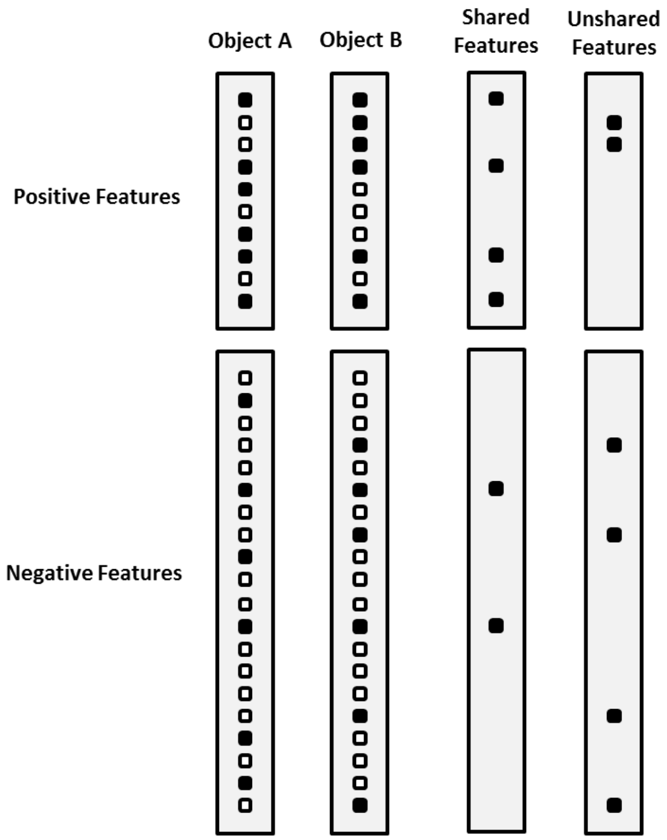

Figure 10. The two columns on the left depict the feature profiles of two objects over 10 positive and 20 negative features. The filled rectangles symbolize presence of a feature, the non-filled rectangles symbolize absence of a feature. Shared features are those that are present in both entities, while unshared features are those that are present in one entity only. The shared features are more likely to be positive than the unshared features.

Hence, from negative information's greater diversity another valence asymmetry follows, namely that objects' or persons' similarities are more positive than their differences. This notion may have a number of important consequences for human information processing in general and for comparison processes in particular, as similarities and differences are the building blocks of human cognitive processing. Similarity is the most basic organizing principle, by which individuals classify objects, form categories, and make generalizations and predictions (Quine, 1969; Holland, Holyoak, Nisbett, Thagard, 1986). Human information processing is always relative and therefore comparative, and similarities and differences are the basis for such comparison (e.g Mussweiler, 2003). In some situations people can be expected to search for or rely on similarities more than on differences, while other situations call for the identification of differences. For example, any integrative process such as abstraction (Trope & Liberman, 2010), inclusion (Bless & Fiedler, 2006), or alignment (Markman & Gentner, 1993) is based on objects' similarities. On the other hand, processes that call for differentiation such as choices or directed comparisons in general rely on objects' differences

(Houston, Sherman, & Baker, 1989; Tversky, 1972). For example, extensive research has shown that the evaluation of a comparison target is typically driven by its unique features while the shared features of target and standard are cancelled out (Tversky & Gati, 1978; see Hodges, 2005 for an overview). At the same time, a standard is typically evaluated based on its complete set of features regardless whether they are shared or not by the target (e.g. Agostinelli, Sherman, Fazio & Hearst, 1986; Houston et al., 1989; Srull & Gaelick, 1983; Tversky & Gati, 1978; Wänke, Schwarz, & Noelle-Neumann, 1995).

As the present model (cf. Figure 10) suggests unique, or unshared features are likely to be negative, a comparison target should have an evaluative disadvantage while comparison standards should enjoy an advantage. This rather abstract notion becomes more meaningful if we consider what usually determines whether an object is the target or the standard of comparison. Comparison standards are prototypical options (Hodges, 2005), which are more familiar to the perceiver (Karylowski, 1990), which do more frequently occur (Polk, Behensky, Gonzalez, & Smith, 2002), and which are encountered first (Houston, et al., 1989). We can therefore expect evaluative disadvantages for non-prototypical, unfamiliar, infrequent and new objects or persons, merely based on the probabilistic reasoning following from negative information’s greater diversity (c.f. Figure 10). That is, these objects or persons should be likely to be associated with and evaluated based on their unique features which are more likely to be negative. This may provide new explanations for a number of so-called “judgment biases” including intergroup bias, minority derogation, or self-serving biases in general. The idea is that people represent outgroups, minority groups, or unfamiliar persons in general, based on their unique and therefore negative attributes.

4.3 Meta-theoretical considerations

Psychological research faces the challenge to explain general behavioral tendencies that different individuals show in different situations. To master this challenge, Psychology has provided numerous theoretical concepts aimed at explaining and predicting behavior. These concepts come in a great variety of shapes. Some build on reductionist biological concepts, others borrow from an information processing computer metaphor, and yet others formulate abstract psychological concepts such as identity, attitude or motivation. It seems that the same behavioral phenomenon can be described on different levels of analysis, mirroring the co-existence of Psychology’s different sub-disciplines such as Neuroscience, Cognitive Psychology, Social Psychology or Personality Psychology. Different levels of analysis come with different problems. Explanatory concepts like attitudes or motivations that are very proximal to the phenomenon to be explained bear the risk of circularity and redundancy (Fiedler, 2014). Reductionist concepts like neuronal activity on the other hand are sometimes too distal from the behavioral phenomena to capture its complexity. The explanatory level is not the only problem that psychological theories face, another one arises from the common location of most explanatory factors as intrapsychic. Reductionist biological factors as well as concepts like affect, or attitudes are ultimately located within the individual. If the explanatory force is located within the same individual whose behavior is to be predicted, the relation between criterion and predictor is necessarily of correlational nature. When researchers observe intrapsychic forces they ultimately have to rely on correlational observations that do not allow causal inferences. Researchers usually try to overcome this shortcoming by experimentally manipulating the intrapsychic causal factor. However, here arises another problem from the fact that experimental manipulations are never intrapsychic themselves, they are always externally induced. The logic to experimentally manipulate intrapsychic factors by extrapsychic events is somewhat paradoxical and it brings us back to Brunswick's (1955) and Lewin's (1951) claim that causal explanations of human cognition and behavior should ultimately be located outside of the individual itself. In line with this argumentation, the present work aims at going beyond affect-based, intrapsychic explanations for valence asymmetries in information processing. Instead of referring to affect-induced processing differences that remain somewhat mystical, many of the observable processing asymmetries might be due to the ecological reality surrounding the information-processing individual. As the present work shows, this ecological reality is characterized by a greater diversity of negative information. I suggest that the resulting higher density of positive information constitutes as a strong extrapsychic explanation for various valence asymmetries in information processing.

5. References

Agostinelli, G., Sherman, S. J., Fazio, R. H., & Hearst, E. S. (1986). Detecting and identifying change: Additions versus deletions. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 445.

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. (2016a). „The Common Good" phenomenon: An amplification/attenuation model of person perception. Manuscript in preparation.

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

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

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

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*, 186-197.

Aristotle. (trans. 1999). Nicomachean ethics (W.D. Ross, Trans.). Kitchener, Ontario, Canada: Batoche Books.

Arndt, J., & Reder, L. M. (2002). Word frequency and receiver operating characteristic curves in recognition memory: Evidence for a dual-process interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 830-842.

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

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.

Badea, C., Brauer, M., & Rubin, M. (2012). The effects of winning and losing on perceived group variability. *Journal of Experimental Social Psychology*, *48*, 1094-1099.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation 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.

Berscheid, E. (1985). Interpersonal attraction. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., Vol. 2, pp. 413-484). New York: Random House.

Biktashev, V. N. (2014). A simple mathematical model of gradual Darwinian evolution: emergence of a Gaussian trait distribution in adaptation along a fitness gradient. *Journal of mathematical biology*, *68*, 1225-1248.

Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology*, *3*, 27-52.

Bless, H., & Fiedler, K. (2006). Mood and the regulation of information processing and behavior. In J.P. Forgas, K.D. Williams & W. van Hippel (Eds.), *Hearts and minds: Affective influences on social cognition and behavior* (pp. 65-84). NE York: Psychology Press.

Brainerd, C. J., Holliday, R. E., Reyna, V. F., Yang, Y., & Toglia, M. P. (2010). Developmental reversals in false memory: Effects of emotional valence and arousal. *Journal of Experimental Child Psychology*, *107*, 137-154.

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review*, *106*, 160-179.

Brainerd C, Stein L, Silveira R, Rohenkohl G, Reyna V. How does negative emotion cause false memories?. *Psychological Science,* 19, 919-925.

Brauer, M., & Bourhis, R. Y. (2006). Social power. *European Journal of Social Psychology*, *36*, 601-616.

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

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

Budson, A. E., Todman, R. W., Chong, H., Adams, E. H., Kensinger, E. A., Krangel, T. S., & Wright, C. I. (2006). False Recognition of Emotional Word Lists in Aging and Alzheimer Disease. *Cognitive and Behavioral Neurology*, *19*, 71-78.

Busey, T. A., & Tunnicliff, J. L. (1999). Accounts of blending, distinctiveness, and typicality in the false recognition of faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1210-1235.

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

Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of person as social entities. *Behavioural Science, 3,* 14–25.

Cantor, J., & Engle, R. W. (1993). Working-memory capacity as long-term memory activation: An individual-differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1101-1114.

Casasanto, D. (2008). Similarity and proximity: When does close in space mean close in mind? *Memory & Cognition*, 36, 1047–1056.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, *25*, 215-224.

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

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

Clark, M. S., & Isen, A. M. (1982). Toward understanding the relationship between feeling states and social behavior. *Cognitive Social Psychology*, *73*, 108.

Cohen J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: L. Erlbaum Associates.

Corneille, O. O., Monin, B. B., & Pleyers, G. G. (2005). Is positivity a cue or a response option? Warm glow vs evaluative matching in the familiarity for attractive and not-so-attractive face. *Journal of Experimental Social Psychology*, *41*, 431-437.

Craik, F. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268-294.

Dehon H, Larøi F, Van der Linden M. Affective valence influences participant's susceptibility to false memories and illusory recollection. *Emotion*, 10, 627-639.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452-478.

Denrell, J. (2005). Why most people disapprove of me: Experience Sampling in Impression Formation. *Psychological Review*, *112*, 951-978.

Dyne, A. M., Humphreys, M. S., Bain, J. D., & Pike, R. (1990). Associative interference effects in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 813-824.

Earles, J. L., Kersten, A. W., Curtayne, E. S., & Perle, J. G. (2008). That's the man who did it, or was it a woman? Actor similarity and binding errors in event memory. *Psychonomic Bulletin & Review*, *15*, 1185-1189.

Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, *37*, 90-93.

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

Estes, Z., Jones, L. L., & Golonka, S. (2012). Emotion affects similarity via social projection. *Social Cognition*, *30*, 584-609.

Fazio, R. H., Eiser, J., & Shook, N. J. (2004). Attitude Formation Through Exploration: Valence Asymmetries. *Journal of Personality and Social Psychology*, *87*, 293-311.

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

Feingold, G.A. 1914. The Influence of Environment on Identification of Persons and Things. *Journal of Criminal Law & Criminology* 5, 39-51.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117–140.

Fiedler, K. (1988). Emotional mood, cognitive style, and behavior regulation. In K. Fiedler & J. P. Forgas (Eds.), *Affect, cognition, and social behavior* (pp. 100–119). Toronto, Canada: Hogrefe Verlag.

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review*, *103*, 193-214.

Fiedler, K. (2001). Affective States Trigger Processes. *Theories of mood and cognition: A user's guidebook*, 85-98.

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

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.

Fiedler, K., Asbeck, J., & Nickel, S. (1991). Mood and constructive memory effects on social judgement. *Cognition and Emotion*, *5*, 363-378.

Fiske, A. P., Haslam, N., & Fiske, S. T. (1991). Confusing one person with another: What errors reveal about the elementary forms of social relations. *Journal of Personality and Social Psychology*, *60*, 656-674.

Flagg, P. W. (1976). Semantic integration in sentence memory?. *Journal of Verbal Learning & Verbal Behavior*, *15*, 491-504.

Forgas, J. P. (2001). *Feeling and thinking: The role of affect in social cognition*. Cambridge University Press.

Forgas, J. P., & Bower, G. H. (1987). Mood effects on person perception judgments. *Journal of Personality and Social Psychology*, 53, 53–60.

Forgas, J. P., Goldenberg, L., & Unkelbach, C. (2009). Can bad weather improve your memory? An unobtrusive field study of natural mood effects on real-life memory. *Journal of Experimental Social Psychology*, *45*, 254-257.

Gagné, C. L. (2000). Relation-based combinations versus property-based combinations: A test of the CARIN theory and the dual-process theory of conceptual combination. *Journal of Memory And Language*, *42*, 365-389.

Galton, F. (1869). *Hereditary genius*. Macmillan and Company.

Garb, J. L., & Stunkard, A. J. (1974). Taste aversions in man. *American Journal of Psychiatry*, *131*, 1204-1207.

Garcia-Marques, T., Mackie, D. M., Claypool, H. M., & Garcia-Marques, L. (2004). Positivity Can Cue Familiarity. *Personality and Social Psychology Bulletin*, *30*, 585-593.

Garner, W. R. (1974). *The processing of information and structure.* New York: Wiley.

Gawronski, B., & Quinn, K. A. (2013). Guilty by mere similarity: Assimilative effects of facial resemblance on automatic evaluation. *Journal of Experimental Social Psychology*, *49*, 120-125.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), Perceiving, acting, and knowing: Toward an ecological psychology (pp. 67-82). Hillsdale, NJ: Erlbaum

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1-67.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 5-16.

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, *100*, 546-567.

Glanzer, M., Koppenaal, L., & Nelson, R. (1972). Effects of relations between words on short-term storage and long-term storage. *Journal of Verbal Learning & Verbal Behavior*, *11*, 403-416.

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

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*, 116-139.

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*.

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.

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

Halberstadt, J., & Rhodes, G. (2003). It's not just average faces that are attractive: Computer-manipulated averageness makes birds, fish, and automobiles attractive. *Psychonomic Bulletin & Review*, *10*, 149-156.

Heathcote, A., Freeman, E., Etherington, J., Tonkin, J., & Bora, B. (2009). A dissociation between similarity effects in episodic face recognition. *Psychonomic Bulletin & Review*, *16*, 824-831.

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

Hirt, E. R., Melton, R., McDonald, H. E., & Harackiewicz, J. M. (1996). Processing goals, task interest, and the mood–performance relationship: A mediational analysis. *Journal of Personality and Social Psychology*, *71*, 245-261.

Hodges, S. D. (2005). A feature-based model of self-other comparisons. *The self in social judgment*, 131-153.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). Induction.

Holmes, J. B., Waters, H., & Rajaram, S. (1998). The phenomenology of false memories: Episodic content and confidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1026-1040.

Houston, D. A., Sherman, S. J., & Baker, S. M. (1989). The influence of unique features and direction of comparison of preferences. *Journal of Experimental Social Psychology*, *25*, 121-141.

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.

Humphreys, M. S., Murray, K. L., & Koh, J. Y. (2014). Reinstating higher order properties of a study list by retrieving a list item. *Memory & Cognition*, *42*, 570-582.

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., & Daubman, K. A. (1984). The influence of affect on categorization. *Journal of Personality and Social Psychology*, *47*, 1206-1217.

sen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality And Social Psychology*, *52*, 1122-1131.

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

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

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1534-1555.

Judd, C. M., & Park, B. (1988). Out-group homogeneity: Judgments of variability at the individual and group levels. *Journal of Personality and Social Psychology, 54,* 778-788.

Judd, C. M., Ryan, C. S., & Park, B. (1991). Accuracy in the judgment of in-group and out-group variability. *Journal of Personality and Social Psychology*, *61*, 366-379.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103, 54-69.

Karylowski, J. J. (1990). Social reference points and accessibility of trait-related information in self-other similarity judgments. *Journal of Personality and Social Psychology*, *58*, 975.

Kimura, M. (1965). A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proceedings of the National Academy of Sciences*, *54*, 731-736.

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

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*.

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

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language*, *37*, 555-583.

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

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, *85*, 445-463.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science*, 4, 195–208.

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

Lande, R. (1975). The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genetical research*, *26*, 221-235.

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., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology*, *98*, 668-682.

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

Leising, D., Ostrovski, O., & Zimmermann, J. (2013). 'Are we talking about the same person here?': Interrater agreement in judgments of personality varies dramatically with how much the perceivers like the targets. *Social Psychological and Personality Science*, *4*, 468-474.

Lepper, M. R. (1994). Hot" versus "cold" cognition: An Abelsonian voyage. In: R. C. Schank, & E. Langer (Eds.), *Beliefs, reasoning, and decision making: Psycho-logic in honor of Bob Abelson* (pp. 237-275). Hillsdale, New Jersey: Lawrence Erlbaum.

Lewandowsky, S., & Farrell, S. (2008). Phonological similarity in serial recall: Constraints on theories of memory. *Journal of Memory and Language*, *58*, 429-448.

Lewin, K. (1935). *A dynamic theory of personality*. New York, NY, US: McGraw-Hill.

Lickel, B., Hamilton, D. L., Wieczorkowska, G., Lewis, A., Sherman, S. J., & Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology*, *78*, 223-246.

Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology*, *57*, 165-188.

Linville, P. W., Salovey, P., & Fischer, G. W. (1986). Stereotyping and perceived distributions of social characteristics: An application to ingroup-outgroup perception. In J. F. Dovidio, S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 165-208). San Diego, CA, US: Academic Press.

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*,*16*, 149-157.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments & Computers*, *28*, 203-208.

Maguire, A. M., Humphreys, M. S., Dennis, S., & Lee, M. D. (2010). Global similarity accounts of embedded-category designs: Tests of the global matching models. *Journal of Memory and Language*, *63*(2), 131-148.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252-271.

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive psychology*, *25*, 431-467.

Mate, J., & Baqués, J. (2009). Visual similarity at encoding and retrieval in an item recognition task. *The Quarterly Journal of Experimental Psychology*, *62*, 1277-1284.

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

Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology*, *32*, 49-96.

Monin, B. (2003). The Warm Glow Heuristic: When Liking Leads to Familiarity. *Journal of Personality and Social Psychology*, *85*, 1035-1048.

Montefinese, M., Zannino, G., & Ambrosini, E. (2014). Semantic similarity between old and new items produces false alarms in recognition memory. *Psychological Research*, 2014.

Montoya, R. M., Horton, R., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*, 25, 889-922.

Murdock, B. (2003). The mirror effect and the spacing effect. *Psychonomic Bulletin & Review*, *10*, 570-588.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological review*, *92*, 289.

Mussweiler, T. (2001). Focus of comparison as a determinant of assimilation versus contrast in social comparison. *Personality and Social Psychology Bulletin*, *27*, 38–47.

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review*, *110*, 472-489.

Nesselroade, J. R., & Cattell, R. B. (1988). *Handbook of multivariate experimental psychology (2nd ed.)*. New York, NY, US: Plenum Press.

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

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 700-708.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 3-27.

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

Nosofsky, R. M., Clark, S. E., & Shin, H. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 282-304.

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.

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.

Ostrom, T. M., & Sedikides, C. (1992). Out-group homogeneity effects in natural and minimal groups. *Psychological Bulletin*, *112*, 536-552.

Park, B., & Hastie, R. (1987). Perception of variability in category development: Instance- versus abstraction-based stereotypes. *Journal of Personality and Social Psychology*, *53*, 621-635.

Park, B., & Rothbart, M. (1982). Perception of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members. *Journal of Personality and Social Psychology*, *42*, 1051-1068.

Peeters, G. (1971). The positive-negative asymmetry: On cognitive consistency and positivity bias. *European Journal of Social Psychology*, *1*, 455-474.

Pelchat, M. L., & Rozin, P. (1982). The special role of nausea in the acquisition of food dislikes by humans. *Appetite*, *3*, 341-351.

Phillips, S., Takeda, Y., & Kumada, T. (2006). An inter-item similarity model unifying feature and conjunction search. *Vision Research*, *46*, 3867-3880.

Polk, T. A., Behensky, C., Gonzalez, R., & Smith, E. E. (2002). Rating the similarity of simple perceptual stimuli: asymmetries induced by manipulating exposure frequency. *Cognition*, *82*, B75-B88.

Postman, L. (1951). The generalization gradient in recognition memory. *Journal of Experimental Psychology*, *42*, 231-235.

Potter, T., Corneille, O., Ruys, K. I., & Rhodes, G. (2007). 'Just another pretty face': A multidimensional scaling approach to face attractiveness and variability. *Psychonomic Bulletin & Review*, *14*, 368-372.

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

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

Queller, S., Mackie, D. M., & Stroessner, S. J. (1996). Ameliorating some negative effects of positive mood: Encouraging happy people to perceive intragroup variability. *Journal of Experimental Social Psychology*, *32*, 361-386.

Quine, W. V. O. (1969). *Ontological relativity and other essays* (No. 1). Columbia University Press.

Reitman, J. S., & Bower, G. H. (1973). Storage and later recognition of exemplars of concepts. *Cognitive Psychology*, *4*, 194-206.

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

Robinson-Riegler, G. L., & Winton, W. M. (1996). The role of conscious recollection in recognition of affective material: Evidence for positive-negative asymmetry. *Journal of General Psychology*, *123*, 93-104.

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

Rota, L. M., & Zellner, D. A. (2007). The categorization effect in hedonic contrast: Experts differ from novices. *Psychonomic Bulletin & Review*, *14*, 179-183.

Rubin, M., & Badea, C. (2012). They're all the same! ... But for several different reasons: A review of the multicausal nature of perceived group variability. *Current Directions in Psychological Science*, *21*, 367-372.

Schopenhauer, A. (1844). *Die Welt als Wille und Vorstellung. 2. Welcher Ergänzungen zu den vier Büchern des ersten Bandes enthält* (Vol. 2). Brockhaus

Schrauf, R. W., & Sanchez, J. (2004). The preponderance of negative emotion words across generations and across cultures. *Journal of Multilingual and Multicultural Development, 25,* 266–284.

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

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

Simonton, D. K. (2003). Francis Galton's Hereditary Genius: Its place in the history and psychology of Science. In R. J. Sternberg (Ed.), *The anatomy of impact: What makes the great works of psychology great* (pp. 3-18). American Psychological Association: Washington, D.C.

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

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

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

Smith, D. G., & Duncan, M. J. (2004). Testing Theories of Recognition Memory by Predicting Performance Across Paradigms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 615-625.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological review*, *81*, 214.

Solomon, R. L., & Wynne, L. C. (1954). Traumatic avoidance learning: the principles of anxiety conservation and partial irreversibility. *Psychological review*, *61*, 353.

Srull, T. K., & Gaelick, L. (1983). General principles and individual differences in the self as a habitual reference point: An examination of self-other judgments of similarity. *Social Cognition*, *2*, 108.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers*, *31*, 137-149.

Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory and Language*, *33*, 203-217.

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1379-1396.

Stroessner, S. J., Mackie, D. M., & Michalsen, V. (2005). Positive Mood and the Perception of Variability Within and Between Groups. *Group Processes & Intergroup Relations*, *8*, 5-25.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, *33*1-39.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal Of Social Psychology*, *1*, 149-178.

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

Thompson, G.L., (2008). Eliminating aggregation bias in experimental research: Random Coefficient Analysis as an alternative to performing a ‘by-subjects’ and/or ‘by-tems’ ANOVA. *Tutorials in Quantitative Methods for Psychology*, *4*, 21-34.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, *2*, i-109.

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

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

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological review*, *117*, 440.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*, 281.

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

Tversky, A., & Gati, I. (1978). Studies of similarity. *Cognition and Categorization*, *1*, 79-98.

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

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.

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., Guastella, A. J., & Forgas, J. P. (2008b). Oxytocin selectively facilitates recognition of positive sex and relationship words. *Psychological Science*, *19*, 1092-1094.

Verde, M. F. (2004). Associative interference in recognition memory: A dual-process account. *Memory & Cognition*, *32*, 1273-1283.

Vokey, J. R., & Read, J. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition*, *20*, 291-302.

Wänke, M., Schwarz, N., & Noelle-Neumann, E. (1995). Asking Comparative Questions The impact of the Direction of Comparison. *Public Opinion Quarterly*, *59*, 347-372.

Ward, R., Duncan, J., & Shapiro, K. (1997). Effects of similarity, difficulty, and nontarget presentation on the time course of visual attention. *Perception & Psychophysics*, *59*, 593-600.

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

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

Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes are attractive because they are easy on the mind. *Psychological science*,*17*, 799-806.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114*, 152–176.

Wixted, J. T., Ghadisha, H., & Vera, R. (1997). Recall latency following pure- and mixed-strength lists: A direct test of the relative strength model of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 523-538

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1341-1354.

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and motivation in face recognition: A critical review of theories of the cross-race effect. *Personality and Social Psychology Review*, *16*, 116-142.

Zajonc, R. B. (1968). Attitudinal Effects of Mere Exposure. *Journal of Personality and Social Psychology*, *9*, 1-27.

Zaki, S. R., & Nosofsky, R. M. (2001). Exemplar accounts of blending and distinctiveness effects in perceptual old–new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1022-1041.

Zárate, M. A., & Sanders, J. D. (1999). Face categorization, graded priming, and the mediating influences of similarity. *Social Cognition*, *17*, 367-389.

6. Footnotes

[1] Experiment 7 was conducted following the comments of two anonymous reviewers.

[2] We calculated the across-participants trait overlap within the eight different target persons to reduce the amount of necessary comparisons. Note that we did not compare each participant's first liked target with all other participants' second, third, and fourth liked targets and so on. This would have resulted in more than 150000 target comparisons.

[3] Note again that trait overlap across participants was calculated within the two different target persons as it was done in Experiment 5.

[4] An additional assumption that has to be made in order for this asymmetry to occur is that participants do not respond "new" more often than "old" (response criterion <= 1).

[5] A pretest showed that old-new discriminations for the 40 word stimuli from Unkelbach and colleagues (2008) produce ceiling effects (very few false alarms). We therefore familiarized participants with all 40 stimuli in the beginning to make the subsequent recognition task more difficult and to increase false alarm rates.

[6] As the calculation of d' and C is valid only on the participant level, item level analysis was conducted on false alarm and hit rated only.

7. Appendix

List of 20 Most Extreme Positive and 20 Most Extreme Negative Word Stimuli (Bargh et al., 1992; Fazio et al., 1986; Klauer and Musch, 1999; Unkelbach, 2008).

Negative		Positive	
Stimulus	German translation	Stimulus	German translation
War	Krieg	Cake	Kuchen
Hitler	Hitler	Kitten	Kätzchen
Bombs	Bomben	Chocolate	Schokolade
Alcoholism	Alkoholismus	Ice Cream	Eiscreme
Disease	Krankheit	Butterfly	Schmetterling
Toothache	Zahnschmerz	Baby	Baby
Hate	Hass	Pizza	Pizza
Funeral	Beerdigung	Food	Essen
Virus	Virus	Hawaii	Hawaii
Crime	Verbrechen	Birthday	Geburtstag
Death	Tod	Movies	Kino
Hell	Hölle	Strawberry	Erdbeere
Divorce	Scheidung	Gift	Geschenk
Cancer	Krebs	Flowers	Blumen
Guns	Gewehre	Party	Party
Recession	Rezession	Music	Musik
Garbage	Müll	Summer	Sommer
Litter	Abfall	Holiday	Urlaub
Cockroach	Kakerlake	Sunshine	Sonnenschein
Taxes	Steuern	Friend	Freund